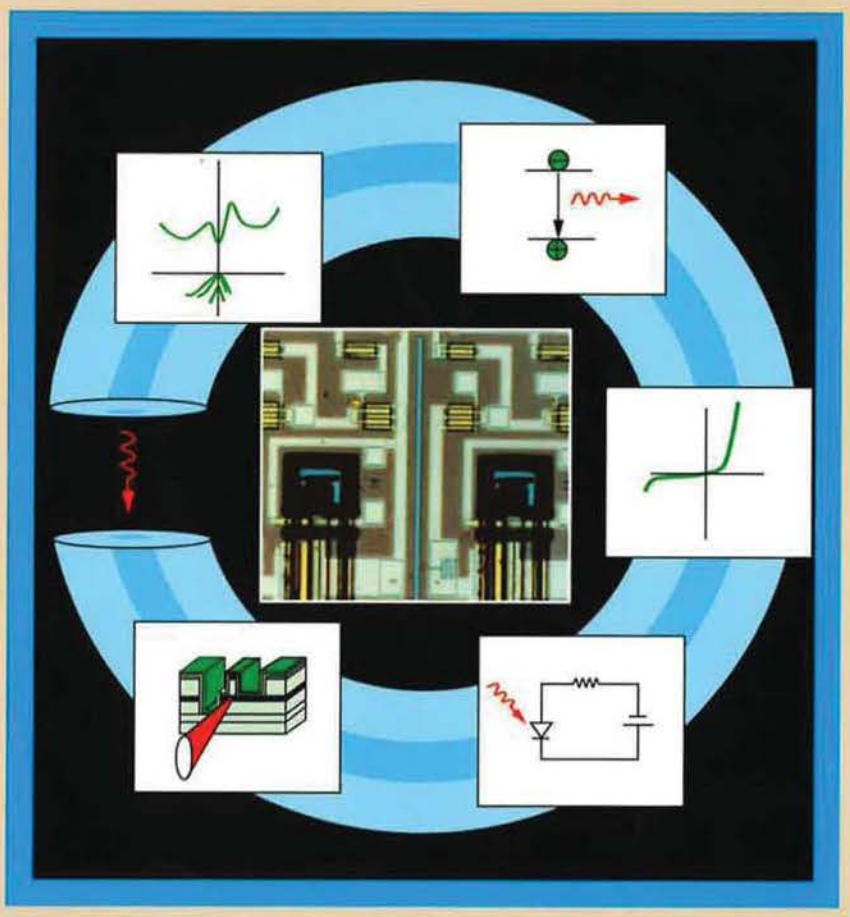
Semiconductor Optoelectronic Devices



PALLAB BHATTACHARYA

Vizio EX1016 Page 0001

LIBRARY OF CONGRESS
0 015 993 914 5

Semiconductor Optoelectronic Devices

PALLAB BHATTACHARYA

This book provides a comprehensive view of compound semiconductors and their applications in photonic devices. The material is well presented and can be used as either a textbook or as a reference source.
—Professor Ben Streetman, The University of Texas at Austin

In the past decade, there has been remarkable innovation in optoelectronic devices, components, and circuits. They have moved out of the laboratory and into a wide variety of systems and applications. To help readers take advantage of these advances, author **Pallab Bhattacharya** offers an in-depth, accessible introduction to the principles of these devices with a unique tutorial/reference approach.

The first real, comprehensive text/reference book on the subject, **Semiconductor Optoelectronic Devices:**

- Describes the basic properties of semiconductors and heterojunctions, and epitaxial techniques used to grow them.
- Carefully presents the basic optical processes of absorption and recombination in bulk and quantum well structures.
- Analyzes and describes light-emitting diodes, lasers, photodetectors, solar cells, and light-modulators.
- Reviews the emerging, important topic of optoelectronics integrated circuits (OEIC).

The principles of the devices are presented with appropriate analyses and derivations, measurement techniques, and recent experimental results. The author introduces new concepts such as pseudomorphic materials, quantum well, distributed-feedback and surface-emitting lasers, modulated barrier photodiodes, coherent and wavelength selective detection, and quantum well modulation devices.

About the Author

Pallab Bhattacharya received his Ph.D. degree in Electronic and Electrical Engineering from the University of Sheffield, U.K., in 1978.

Dr. Bhattacharya is Professor of Electrical Engineering and Computer Science at the University of Michigan at Ann Arbor. He was invited Professor at the Ecole Polytechnique Federale de Lausanne, Switzerland, during 1981-1982. He is Associate Editor of the IEEE Transactions on Electron Devices and serves on the advisory board of the Electrical and Communications Systems Division at the National Science Foundation.

His research has focused on liquid phase and molecular beam epitaxy, electronic and optoelectronic properties of semiconductors, optoelectronic devices, and optoelectronic integrated circuits.

PRENTICE HALL
Englewood Cliffs, NJ 07632

ISBN 0-13-805748-6
90000
9 780138 057480

Library of Congress Cataloging-in-Publication Data

Bhattacharya, Pallab.
Semiconductor optoelectronic devices / Pallab Bhattacharya.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-805748-6
1. Optoelectronic devices. 2. Semiconductors. I. Title.
TK8320.B52 1994
621.3815'2--dc20 93-38501 93-15773
CIP

TK8320
.B52
1994
copy 2

Publisher: Alan Apt
Production Editor: Mona Pompili
Cover Designer: Pallab Bhattacharya
Copy Editor: Brenda Melissaratos
Prepress Buyer: Linda Behrens
Manufacturing Buyer: Dave Dickey
Supplements Editor: Alice Dworkin
Editorial Assistant: Shirley McGuire
Cover Illustration: Optoelectronic integrated circuit shown on cover is part of a 4-channel transmitter array (courtesy of O. Wada, Fujitsu Limited, Atusgi, Japan).

© 1994 by Prentice-Hall, Inc.
A Paramount Communications Company
Englewood Cliffs, New Jersey 07632

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs ot determine their effectiveness. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-805748-6

PRENTICE-HALL INTERNATIONAL (UK) LIMITED, *London*
PRENTICE-HALL OF AUSTRALIA PTY. LIMITED, *Sydney*
PRENTICE-HALL CANADA, INC., *Toronto*
PRENTICE-HALL HISPANOAMERICANA, S.A., *Mexico*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*
SIMON & SCHUSTER ASIA PTE. LTD., *Singapore*
EDITORA PRENTICE-HALL DO BRASIL, LTDA., *Rio de Janeiro*

Vizio EX1016 Page 0003

Semiconductor Optoelectronic Devices

PALLAB BHATTACHARYA
Department of Electrical Engineering
and Computer Science
University of Michigan, Ann Arbor

Prentice Hall, Englewood Cliffs, New Jersey 07632

Vizio EX1016 Page 0004

LIBRARY OF CONGRESS
NOV 09 1993
COPYRIGHT OFFICE
4

Dedicated to the
cherished memory
of my father.

Preface

SCOPE

Since the demonstration of the first light-emitting diode (LED) and junction laser, optoelectronics has made remarkable progress. The developments in this field have been driven by the needs of lightwave communication systems, alternate energy sources, and optical or optoelectronic counterparts of electronic switching and logic elements. Optical devices and components and optical fibers are selectively replacing electronic devices and circuits, offering unique advantages. In fact, optoelectronic devices and circuits have unobtrusively and efficiently made their way into our daily lives. In the light of this enormous progress in the field, it is hoped that this book will reflect these dramatic changes in the field and serve two main purposes: (1) to formally introduce senior-level undergraduate and graduate students to optoelectronics, thereby helping them to guide their studies and career developments, and (2) to provide, in an accessible textbook format, a good and well-focused reference/tutorial book for practicing engineers and physicists.

PRESENTATION

The text has been developed at two levels, to benefit both the seniors and graduate students. The book is intended to be self-sufficient and extensive reference work should not be necessary. A background of a first course in semiconductors is assumed.

The first four chapters lay the foundations for the optoelectronic devices. The first chapter describes compound semiconductor materials and their epitaxy. Much of the present-day device concepts would not be realized without sophisticated and matured epitaxial techniques. Semiconductor statistics and carrier transport properties are described in Chapter 2. The basic optical processes of absorption and recombination in bulk and quantum well structures are analyzed and described in Chapter 3. Here, detailed quantum mechanical calculations are excluded, since these are found in at least half-a-dozen texts. However, appropriate references are provided as foot-

Vizio EX1016 Page 0006

notes. Chapter 4 describes junction theory, including metal-semiconductor junctions and heterojunctions. The case of high-level injection, which becomes important for laser operation, is emphasized in this chapter.

The devices themselves are described in Chapters 5–11 in the following order: light-emitting diodes, lasers, photodetectors, solar cells, and light-modulators. Lasers and photodetectors, which are perhaps the more important and common optoelectronic devices, are each described in two chapters.

The principles of the devices are presented with appropriate analyses and derivations. Measurement techniques and recent experimental results are also included to give the reader a feel for real parameter values.

The organization of the device chapters should provide the instructor the flexibility to present material to both undergraduate and graduate students. In discussing the different devices, I have introduced new concepts, within the scope of the text. For example, pseudomorphic materials, quantum wells, distributed-feedback and surface-emitting lasers, modulated barrier photodiodes, coherent and wavelength selective detection, and quantum well modulation devices are all described and analyzed in various levels of detail. Finally, the text is concluded in Chapter 12 with a review of optoelectronic integrated circuits (OEICs), an emerging and important subject.

HOW TO USE THIS BOOK

This book is flexible and can be adapted to your local curriculum and course needs. An example of a one-term senior undergraduate course that could be taught from the book would cover chapters 1, 2, and 4 as review; chapters 3, 5, 6, part of 7, 8, and part of 11 should be treated as essential; and chapter 10 could be treated as optional. The rest of chapters 7 and 11, and chapters 9 and 12 I consider advanced material that you can teach in a graduate level course or material that you can choose from selectively to tailor the course to your desired emphasis and objectives.

READING LIST AND PROBLEMS

Suggested texts for more extensive reading and key articles from journals and periodicals are listed at the end of each chapter and as footnotes. These will help the more inquisitive students to go beyond the confines of the text and course and to enhance their knowledge and understanding. Also included are a set of problems at the end of each chapter, in addition to a few worked-out examples. The purpose of these problems is twofold: (1) to enhance the understanding of the different devices and underlying concepts and (2) to get a feel for practical values of different device and material parameters and their units.

UNITS

The rationalized MKS system of units has been mostly followed, with convenient changes. For example cm is more often used as the unit of length, and the electron

Vizio EX1016 Page 0007

volt (eV) is used in place of joule (J) as the unit of energy. The cgs system is sometimes used to keep in line with common use.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge all those who have provided help, inspiration, and encouragement as I proceeded with this arduous task. I would like to thank Professor Ben Streetman for providing invaluable suggestions and warning me of the many pitfalls before I got started. Discussions with Dr. Niloy Dutta are gratefully acknowledged. I am indebted to my colleagues, Professor George Haddad for his encouragement and Professor Jasprit Singh for his many valuable comments and suggestions. The text has greatly benefited from the help and comments of many of my present and former graduate students. Special thanks are due to Larry Davis, Augusto Gutierrez-Aitken, Yeeloy Lam, Sanjay Sethi, and Hsiang-Chi Sun, and Drs. Yaochung Chen, Subrata Goswami, Shantanu Gupta, Weiqi Li, and Doyle Nichols. I would like to thank my professional colleagues who have generously contributed data and photographs. They are J. Goldman (RIBER SA) and Drs. G. A. Antypas (Crystacomm Inc.), S. N. G. Chu (AT&T Bell Laboratories), I. Hayashi (Optoelectronics Technology Research Laboratory), J. L. Jewell (Photonics Research Incorporated), J. Loehr (Wright Laboratories, WPAFB), D. A. B. Miller (AT&T Bell Laboratories), D. Pooladdej (Laser Diode, Inc.), R. Sahai (Rockwell International), J. Singh (University of Michigan), S. Swirhun (Bandgap Technology Corporation), W. T. Tsang (AT&T Bell Laboratories), O. Wada (Fujitsu Limited), I. Weinberg (NASA, Lewis Research Center), and E. Woelk (AIXTRON GmbH). This book would not have seen the light of day without the hard work of three persons: Christina Baydl, who transformed my hieroglyphics into a readable form, and Mary Ann Pruder and Mina Hale, who did all the artwork. The help of Alan Apt and Mona Pompili is greatly appreciated. I will remain indebted to all of you.

Last, and by no means the least, I am indebted to my mother for her support and to my family for their understanding, support, and willingness to sacrifice many evenings, weekends, and holidays.

Pallab Bhattacharya
Ann Arbor, Michigan

Vizio EX1016 Page 0008

Contents

Vizio EX1016 Page 0009

Vizio EX1016 Page 0010

Vizio EX1016 Page 0011

Vizio EX1016 Page 0012

Vizio EX1016 Page 0013

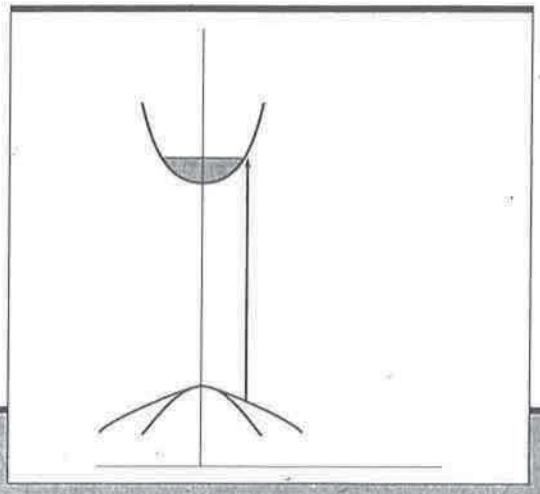
Vizio EX1016 Page 0014

Vizio EX1016 Page 0016

Vizio EX1016 Page 0017

3

Optical Processes in Semiconductors



Chapter Contents

Vizio EX1016 Page 0018

3.1 ELECTRON-HOLE PAIR FORMATION AND RECOMBINATION

The operation of almost all optoelectronic devices is based on the creation or annihilation of electron-hole pairs. Pair formation essentially involves raising an electron in energy from the valence band to the conduction band, thereby leaving a hole behind in the valence band. In principle, any energetic particle incident on a semiconductor, which can impart an energy at least equal to the bandgap energy to a valence band electron, will create pairs. With respect to the bonding in the lattice, this process is equivalent to breaking a covalent bond. The simplest way to create electron-hole pairs is to irradiate the semiconductor. Photons with sufficient energy are absorbed, and these impart their energy to the valence band electrons and raise them to the conduction band. This process is, therefore, also called *absorption*. The reverse process, that of electron and hole recombination, is associated with the pair giving up its excess energy. Recombination may be *radiative* or *nonradiative*. In a nonradiative transition, the excess energy due to recombination is usually imparted to phonons and dissipated as heat. In a radiative transition, the excess energy is dissipated as photons, usually having energy equal to the bandgap (i.e., $\hbar\omega = \mathcal{E}_g$). This is the *luminescent* process, which is classified according to the method by which the electron-hole pairs are created. *Photoluminescence* involves the radiative recombination of electron-hole pairs created by injection of photons. *Cathodoluminescence* is the process of radiative recombination of electron-hole pairs created by electron bombardment. *Electroluminescence* is the process of radiative recombination following injection with a p-n junction or similar device.

In a semiconductor in equilibrium (i.e., without any incident photons or injection of electrons), the carrier densities can be calculated from an equilibrium Fermi level by using Fermi-Dirac or Boltzmann statistics outlined in Sec. 2.5.3. When excess carriers are created by one of the techniques described above, nonequilibrium conditions are generated and the concept of a Fermi level is no longer valid. One can, however, define nonequilibrium distribution functions for electrons and holes as

$$f_n(\mathcal{E}) = \frac{1}{1 + \exp\left(\frac{\mathcal{E}-\mathcal{E}_{fn}}{k_B T}\right)} \tag{3.1}$$

$$f_p(\mathcal{E}) = \frac{1}{1 + \exp\left(\frac{\mathcal{E}-\mathcal{E}_{fp}}{k_B T}\right)} \tag{3.2}$$

These distribution functions define $\mathcal{E}_{fn}$ and $\mathcal{E}_{fp}$, the *quasi-Fermi levels* for electrons and holes, respectively. In some texts they are referred to as IMREFs (Fermi spelled backward). When the excitation source creating excess carriers is removed, $\mathcal{E}_{fn} = \mathcal{E}_{fp} = \mathcal{E}_F$. The difference $(\mathcal{E}_{fn} - \mathcal{E}_{fp})$ is a measure of the deviation from equilibrium. As with equilibrium statistics, we obtain for the nondegenerate case

$$f_n(\mathcal{E}) \cong \exp\left(\frac{\mathcal{E}_{fn} - \mathcal{E}}{k_B T}\right) \tag{3.3}$$

Vizio EX1016 Page 0019

$$f_p(\mathcal{E}) \cong \exp\left(\frac{\mathcal{E} - \mathcal{E}_{fp}}{k_B T}\right) \tag{3.4}$$

and the nonequilibrium carrier concentrations are given by

$$n = N_C \exp\left(\frac{\mathcal{E}_{fn} - \mathcal{E}_C}{k_B T}\right) \tag{3.5}$$

$$p = N_V \exp\left(\frac{\mathcal{E}_V - \mathcal{E}_{fp}}{k_B T}\right) \tag{3.6}$$

The concept of quasi-Fermi levels is extremely useful, since it provides a means to take into account changes of carrier concentration as a function of position in a semiconductor. As we shall see in Chapter 4, in a p-n junction under forward bias a large density of excess carriers exist in the depletion region and close to it on either side. The concentration of these carriers can be determined from the appropriate quasi-Fermi levels. A junction laser is operated under such forward bias injection conditions to create population inversion. To consider a simple example, assume that an n-type semiconductor with an equilibrium electron density n_o (= N_D, the donor density) is uniformly irradiated with intrinsic photoexcitation (above-bandgap light) so as to produce Δn electron-hole pairs with a generation rate G. The nonequilibrium electron and hole concentrations are given by

$$n = \Delta n + n_o \tag{3.7}$$

$$p = \Delta n + n_i^2/n_o \tag{3.8}$$

Using Eqs. 3.3–3.8, Fig. 3.1 illustrates the change in the energy position of the quasi Fermi levels in GaAs as the generation rate changes.

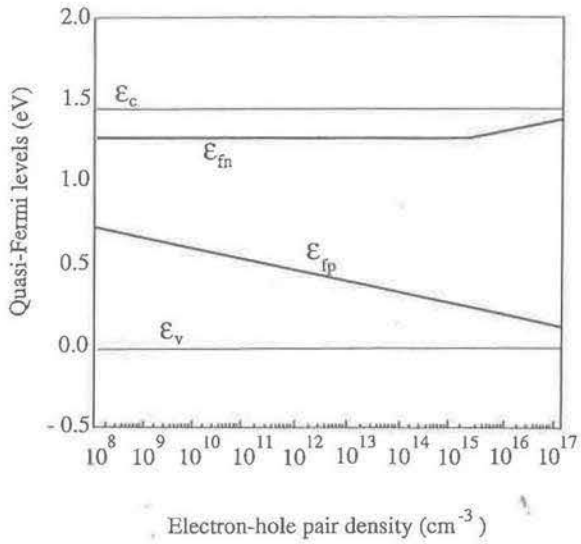


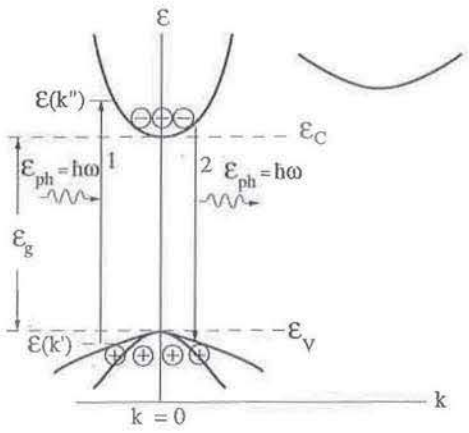
Figure 3.1 Energy position of the electron and hole quasi-Fermi levels as a function of pair generation rate in GaAs at room temperature. It is assumed that the sample is n-type with $N_D = 10^{15}$ cm^{-3} (from M. Shur, *Physics of Semiconductor Devices*, ©1990. Reprinted by permission of Prentice-Hall, Englewood Cliffs, New Jersey).

Vizio EX1016 Page 0020

The excess carriers created in a semiconductor must eventually recombine. In fact, under steady-state conditions the recombination rate must be equal to the generation rate:

$$G = R \tag{3.9}$$

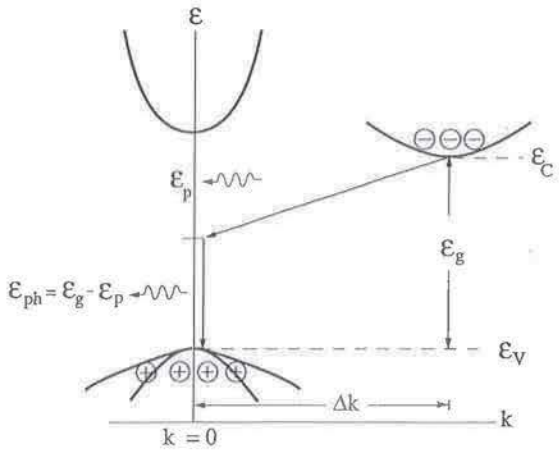
Generation and recombination processes involve transition of carriers across the energy bandgap and are therefore different for direct and indirect bandgap semiconductors, as illustrated in Fig. 3.2. In a direct bandgap semiconductor, as shown in Fig. 3.2(a), the valence band maximum and the conduction band minimum occur at the zone center ($\mathbf{k} = 0$) and an upward or downward transition of electrons does not require a change in momentum or the involvement of a phonon. Therefore, in direct bandgap semiconductors such as GaAs, an electron raised to the conduction band, say, by photon absorption, will dwell there for a very short time and recombine again with a valence band hole to emit light of energy equal to the bandgap. Thus, the probability of *radiative recombination* is very high in direct bandgap semiconductors. The processes are quite different in an indirect bandgap semiconductor. Considering the band diagrams shown in Fig. 3.2(b) and (c), since the conduction band minima are not at $\mathbf{k} = 0$, upward or downward transition of carriers require a change in momentum, or the involvement of a phonon. Thus, an electron dwelling in the conduction band minimum, at $\mathbf{k} \neq 0$, cannot recombine with a hole at $\mathbf{k} = 0$ until a phonon with the right energy and momentum is available. Both phonon emission or absorption processes can assist the downward transition. In order for the right phonon collision to occur, the dwell time of the electron in the conduction band increases. Since no crystal is perfect, there are impurities and defects in the lattice that manifest themselves as traps and recombination centers. It is most likely that the electron and hole will recombine nonradiatively through such a defect center, and the excess energy is dissipated into the lattice as heat. The competing nonradiative processes reduce the probability of radiative recombination in indirect bandgap materials such as Si, Ge, or



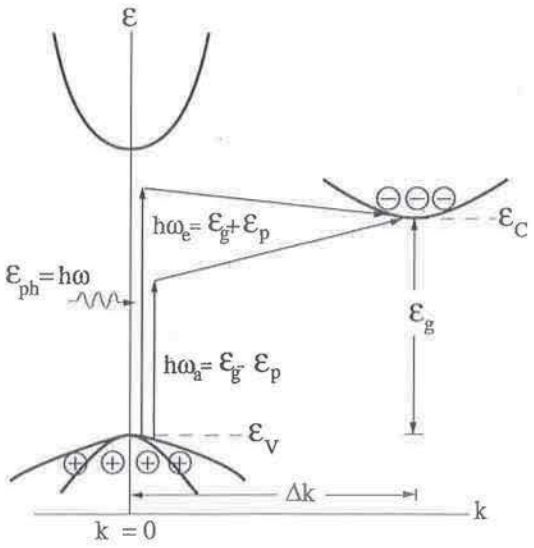
(a)

Figure 3.2(a) Illustration of band-to-band of absorption and recombination processes in (a) direct bandgap semiconductor and (b) and (c) indirect bandgap semiconductor.

Vizio EX1016 Page 0021



(b)



(c)

Figure 3.2 *(continued)*

GaP. These semiconductors are therefore, in general, not suitable for the realization of light sources such as light-emitting diodes and lasers.

3.1.1 Radiative and Nonradiative Recombination

For continuous carrier generation by optical excitation or injection, a quasi equilibrium or steady state is produced. Electrons and holes are created and annihilated in pairs and, depending on the injection level, a steady-state excess density $\Delta n = \Delta p$ is established in the crystal. This equality is also necessary for the maintenance of overall charge neutrality. When the excitation source is removed, the density of excess carriers returns to the equilibrium values, n_o and p_o. The decay of excess carriers usually follows an exponential law with respect to time $\sim \exp(-t/\tau)$, where τ is defined as

Vizio EX1016 Page 0022

the lifetime of excess carriers. The lifetime is determined by a combination of intrinsic and extrinsic parameters, and the performance characteristics of most optoelectronic devices depend on it. In the discussion that follows, we will be concerned mainly with *bulk* recombination processes. It is important to remember that, depending on the semiconductor sample and its surface, there can be a very strong *surface* recombination component which depends on the density of surface states.

In general, the excess carriers decay by radiative and/or nonradiative recombination, in which the excess energy is dissipated by photons and phonons. The former is of importance for the operation of luminescent devices. Nonradiative recombination usually takes place via surface or bulk defects and traps (Fig. 3.3), as discussed in Chapter 2, and reduces the radiative efficiency of the material. Therefore the total lifetime τ can be expressed as

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} \tag{3.10}$$

where τ_r and τ_{nr} are the radiative and nonradiative lifetimes, respectively. Also, the total recombination rate R_{total} is given by

$$R_{\text{total}} = R_r + R_{nr} = R_{sp} \tag{3.11}$$

where R_r and R_{nr} are radiative and nonradiative recombination rates per unit volume, respectively, and R_{sp} is called the *spontaneous* recombination rate, to distinguish

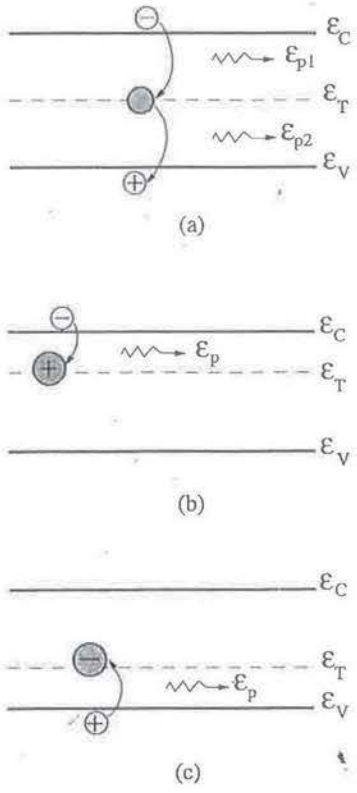


Figure 3.3 Nonradiative recombination at (a) recombination center, (b) electron trap, and (c) hole trap. The excess carrier energy in all cases is dissipated by single or multiple phonons.

Vizio EX1016 Page 0023

R_{total} from the *stimulated* recombination rate to be defined later in Chapter 6. The internal quantum efficiency or radiative recombination efficiency is defined as

$$\eta_r = \frac{R_r}{R_r + R_{nr}} \tag{3.12}$$

For an exponential decay process, $\tau_r = \Delta n/R_r$ and $\tau_{nr} = \Delta n/R_{nr}$ where Δn is the excess electron concentration. Therefore,

$$\eta_r = \frac{1}{1 + \tau_r/\tau_{nr}} \tag{3.13}$$

To achieve high internal quantum efficiency, the ratio τ_r/τ_{nr} should be as small as possible, or τ_{nr} should be as large as possible. The value of τ_{nr} is controlled by the properties of defects, which produce levels in the bandgap of a semiconductor. The excess energy of carriers recombining at these levels is dissipated by phonons. Another nonradiative process is *Auger recombination*, to be discussed in Sec. 3.8. It also follows from Eq. 3.9 that under steady-state conditions $\Delta n = G\tau_r$.

3.1.2 Band-to-Band Recombination

The simplest carrier decay process is spontaneous band-to-band recombination, whose rate, without momentum conservation, is given by

$$R_{sp} = B_r np \tag{3.14}$$

where B_r is defined as the coefficient for band-to-band recombination in units of $cm^3.s^{-1}$. B_r is related to the transition probability P to be discussed in the next section. In terms of the equilibrium and excess carrier densities,

$$R_{sp} = B_r(n_o + \Delta n)(p_o + \Delta p) \tag{3.15}$$

where $\Delta n = \Delta p$. The spontaneous radiative recombination rate for excess carriers can be expressed as

$$R_{sp}^{ex} = \frac{\Delta n}{\tau_r} \tag{3.16}$$

and therefore

$$R_{sp} = R_{sp}^{o} + R_{sp}^{ex} \tag{3.17}$$

where

$$R_{sp}^{o} = B_r n_o p_o \tag{3.18}$$

is the spontaneous recombination rate in thermal equilibrium. From Eqs. 3.15 and 3.18,

$$R_{sp} = B_r\left[n_o p_o + \Delta n(n_o + p_o) + (\Delta n^2)\right] \tag{3.19}$$

$$R_{sp}^{ex} = B_r \Delta n\,[n_o + p_o + \Delta n] \tag{3.20}$$

and

$$\tau_r = \frac{1}{B_r(n_o + p_o + \Delta n)} \tag{3.21}$$

Vizio EX1016 Page 0024

When $\Delta n \gg n_o, p_o$, which is relevant to laser operation,

$$\tau_r \cong \frac{1}{B_r \Delta n} \tag{3.22}$$

This is the *bimolecular* recombination regime, when the lifetime changes with Δn. At low injection levels, such that $\Delta n < n_o, p_o$

$$\tau_r \cong \frac{1}{B_r(n_o + p_o)} \tag{3.23}$$

which remains constant, being determined by the background carrier concentrations. For an intrinsic semiconductor under low-level injection, since $n_o = p_o = n_i$,

$$\tau_r = \frac{1}{2B_r n_i} \tag{3.24}$$

Eq. 3.22 is valid for $\Delta n \leq 10^{18}$ cm^{-3}. For higher values of Δn,

$$\tau_r \cong \tau_o \tag{3.25}$$

which is usually constant for any material. For example for GaAs, $\tau_o \cong 0.5$ ns.

The value of the recombination coefficient depends on the bandgap and whether the semiconductor has a direct or an indirect bandgap. Direct bandgap semiconductors usually have values of B_r ranging from 10^{-11} to 10^{-9} cm^3.s^{-1} and indirect bandgap semiconductors have values of B_r ranging from 10^{-15} to 10^{-13} cm^3.s^{-1}.

EXAMPLE 3.1

Objective. To calculate τ_r in GaAs having $n_o = 10^{14}$ cm^{-3} under high- and low-level injections for $B_r = 7 \times 10^{-10}$ cm^3/s.

At a high injection level of 10^{18} cm^{-3}, $\tau_r = (7 \times 10^{-10} \times 10^{18})^{-1} s = 1.43$ns.

At a low injection level of 10^{16} cm^{-3}, $\tau_r = 143$ ns for the same value of B_r. This value of τ_r is almost an order of magnitude larger than that measured in pure GaAs samples. Therefore, the value of B_r used here is only valid for the high-level injection case and should be larger for the case of low-level injection.

3.2 ABSORPTION IN SEMICONDUCTORS

3.2.1 Matrix Elements and Oscillator Strength for Band-to-Band Transitions

The operation of optical devices that we will describe and discuss in this text depends on the upward and downward transitions of carriers between energy bands. These transitions result in absorption or emission of light, which is electromagnetic energy. The measurement of absorption and emission spectra in semiconductors constitutes an important aspect of materials characterization. They provide information not only on the bandgap, but the measurements also provide information on direct and indirect

Vizio EX1016 Page 0025

transitions, the distribution of states, and the energy position of defect and impuri levels. The absorption spectrum spans a wide energy (or wavelength) range, extendir from the near-bandgap energies to the low-energy transitions involving free carrie and lattice vibrations. In the context of this text the more important ones are th near-bandgap transitions.

The process of photon absorption results in the transition of an electron from lower energy state to a higher energy state, the simplest form of which may be a dire transition from the valence to the conduction band. The different possible transitior are outlined in this chapter. In what follows, the process of band-to-band transitior in semiconductors, in which photons are absorbed, is analyzed.

The energy-momentum diagrams of a direct and an indirect semiconductor wer shown in Figs. 3.2(a) and (b), respectively. Considering the case of an electro raised from the top of the valence band to the bottom of the conduction band du to absorption of a photon in a direct transition, there is no change in momentun Strictly, there is a small change in $\mathbf{k}$ due to the finite momentum of the photor which is equal to h/λ. For most III–V semiconductors $\lambda \sim 1\mu$m, and the resultan momentum change is very small. An indirect transition due to the absorption of a photon is illustrated in Fig. 3.2(c). Since a large change in momentum is involve in this case, the transition can occur only by the emission or absorption of a phonon The process can be described by the equation

$$\mathcal{E}_{ph} \pm \mathcal{E}_p = \mathcal{E}_g \tag{3.26}$$

and the change in momentum is given by

$$\Delta\mathbf{k} = \mathbf{k}_p \tag{3.27}$$

where $\mathbf{k}_p$ is the wavevector of the phonon and $\mathcal{E}_{ph}$ and $\mathcal{E}_p$ are the photon and phonor energies, respectively. Therefore, an optical or acoustic phonon with the right energy and momentum must be involved in an indirect transition.

The wavelength dependence of direct transitions, for the case of absorption, is illustrated in Fig. 3.2(a), where we consider a transition away from the zone center. The top of the valence band is taken as the zero of energy. The transition occurs in energy from $\mathcal{E}(\mathbf{k}')$ to $\mathcal{E}(\mathbf{k}'')$ where

$$\mathcal{E}(\mathbf{k}') = -\frac{\hbar^2\mathbf{k}'^2}{2m_h^*} \tag{3.28}$$

and

$$\mathcal{E}(\mathbf{k}'') = \mathcal{E}_g + \frac{\hbar^2\mathbf{k}''^2}{2m_e^*}. \tag{3.29}$$

Here $\mathcal{E}_g$ is the direct bandgap at $\mathbf{k} = 0$. The energy of the absorbed photon is $\mathcal{E}_{ph} = \mathcal{E}(\mathbf{k}'') - \mathcal{E}(\mathbf{k}')$, which is the requirement for energy conservation. Remember that such a transition will take place only if the level at $\mathcal{E}(\mathbf{k}')$ is filled and that at $\mathcal{E}(\mathbf{k}'')$ is empty. Also, for momentum conservation $\mathbf{k}'$ must be nearly equal to $\mathbf{k}''$. This is called the *k-selection rule.*

Vizio EX1016 Page 0026

The matrix element and probability of an optical transition from $\mathcal{E}(\mathbf{k}')$ to $\mathcal{E}(\mathbf{k}'')$ can be calculated by considering first-order time-dependent perturbation theory. The time-independent form of the Schrödinger equation is

$$\mathbf{H}_0\Psi = \mathcal{E}(\mathbf{k}')\Psi \tag{3.30}$$

where $\mathbf{H}_0$ is the Hamiltonian of the unperturbed system. In the case of a perturbation $\mathbf{H}_1$, which in our context is light or electromagnetic radiation, causing a carrier transition from a state at $\mathcal{E}(\mathbf{k}')$to a state at $\mathcal{E}(\mathbf{k}'')$, the time-dependent Schrödinger equation can be expressed as

$$(\mathbf{H}_0 + \mathbf{H}_1)\Psi = j\hbar\frac{d\Psi}{dt} \tag{3.31}$$

with

$$\Psi = \sum_m A_m(t)\Psi_m e^{-j\mathcal{E}_m(\mathbf{k}')t/\hbar} \tag{3.32}$$

It may be noted that $|A_m(t)|^2$ is the transition probability. The calculation of the matrix element of an optical transition has been described in detail in a few texts and is not repeated here. The matrix element for direct transitions, where the condition

$$\mathbf{k}'' - \mathbf{k}' = \mathbf{k} \cong 0 \tag{3.33}$$

is satisfied, is given by

$$\begin{aligned} H_{k''k'} &= \int \Psi^*_{k''}\mathbf{H}_1\Psi_{k'}d\mathbf{r} \\ &= \frac{jq\hbar A}{2m_0}\int u^*_C(\mathbf{r},\mathbf{k}'')[\mathbf{a}_0\cdot\nabla u_V(\mathbf{r},\mathbf{k}') + j(\mathbf{a}_0\cdot\mathbf{k}')u_V(\mathbf{r},\mathbf{k}')]d\mathbf{r} \end{aligned} \tag{3.34}$$

where u_V and u_C are the Bloch functions corresponding to the valence and conduction bands, respectively, $\mathbf{A}$ is the magnetic vector potential of the electromagnetic wave and $\mathbf{a}_0$ is a polarization unit vector. The first term represents the matrix element for *allowed* direct transitions and is usually much larger than the second term, which represents *forbidden* transitions. If $\mathbf{k}' = \mathbf{k}''$, the matrix element of the forbidden transition is zero. However, because of the small change in momentum due to the small but finite momentum of the photon, the matrix element of the forbidden transition has a finite value. It can be shown that the transition probability per unit volume per unit time for an allowed direct transition is given by†

$$P(\hbar\omega) = \frac{q^2|\mathbf{A}|^2(2m_r^*)^{3/2}p_{CV}^2}{4\pi m_0^2\hbar^4}(\hbar\omega - \mathcal{E}_g)^{1/2} \tag{3.35}$$

where p_{CV} is the matrix element of the momentum operator (or the momentum matrix element) and m_r^* is the reduced mass given by

$$m_r^* = \frac{m_e^*m_h^*}{m_e^* + m_h^*} \tag{3.36}$$

†R. H. Bube, *Electronic Properties of Crystalline Solids*, Academic Press, New York, 1974.

Vizio EX1016 Page 0027

Equation 3.35 contains the *joint density of states* to be discussed in Chapter 5. When the **k**-selection rule is obeyed, $|p_{CV}|^2 = 0$ unless $\mathbf{k}' = \mathbf{k}''$. Equation 3.35 is an important relationship and shows that the transition probability of a direct allowed transition varies as $(\hbar\omega - \mathcal{E}_g)^{1/2}$. The transition probability includes the summation over all filled valence band states and empty conduction band states and over all $\mathbf{k}'$ and $\mathbf{k}''$ values that satisfy energy and momentum conservation. In the form expressed in Eq. 3.35, it is assumed that the semiconductor is at 0°K when the valence band is completely filled and the conduction band is empty.

Since the absorption of a photon of energy $\hbar\omega$ is involved in a direct transition, it is important to calculate the absorption coefficient α. Assume that a monochromatic photon flux $\Im_i$, given by

$$\Im_i = \frac{|\mathbf{S}|}{\hbar\omega} \qquad (\text{photons/cm}^2.\text{s}) \tag{3.37}$$

is incident on the crystal. Here $|\mathbf{S}|$ is the radiation energy crossing unit area in unit time, or the *Poynting vector*. The transmitted intensity $\Im_d$ is then

$$\Im_d = \frac{|\mathbf{S}|}{\hbar\omega} - P(\hbar\omega)d \tag{3.38}$$

where d is the thickness of the sample. The second term represents the number of photons absorbed per unit time per unit area, normal to the incident light in a thickness d. Equation 3.38 can be written as

$$\begin{aligned} \Im_d &= \frac{|\mathbf{S}|}{\hbar\omega} e^{-\alpha d} \\ &\cong \frac{|\mathbf{S}|}{\hbar\omega}(1 - \alpha d) \end{aligned} \tag{3.39}$$

for small αd. Thus,

$$\alpha(\hbar\omega) = \frac{P\hbar\omega}{|\mathbf{S}|} \tag{3.40}$$

The average value of the Poynting vector over a period of the electromagnetic wave can be expressed as

$$|\mathbf{S}| = \frac{1}{2} n_r \epsilon_0 c \omega^2 |\mathbf{A}^2| \tag{3.41}$$

where n_r is the refractive index of the crystal. Substitution of Eqs. 3.35 and 3.41 into Eq. 3.40 leads to

$$\alpha(\hbar\omega) = C_1 n_r^{-1} \left(\frac{2m_r^*}{m_0}\right)^{3/2} \frac{f_{CV}}{\hbar\omega} (\hbar\omega - \mathcal{E}_g)^{1/2} \tag{3.42}$$

where

$$C_1 = \frac{q^2 m_0^{1/2}}{4\pi\hbar^2 \epsilon_0 c} \tag{3.43}$$

Vizio EX1016 Page 0028

and

$$f_{CV} = \frac{2p_{CV}^2}{m_0} \tag{3.44}$$

Expressing $\hbar\omega$ and $\mathcal{E}_g$ in eV,

$$\alpha(\hbar\omega) = 2.64 \times 10^5 n_r^{-1} \left(\frac{2m_r^*}{m_0}\right)^{3/2} \frac{f_{CV}}{\hbar\omega}(\hbar\omega - \mathcal{E}_g)^{1/2} \quad (cm^{-1}) \tag{3.45}$$

f_{CV} is called the *oscillator strength* for the transition. It has a value approximately equal to 20 eV in most semiconductors. Therefore, for GaAs ($f_{CV} = 23$ eV), we get from Eq. 3.45,

$$\alpha(\hbar\omega) = 5.6 \times 10^4 \frac{(\hbar\omega - \mathcal{E}_g)^{1/2}}{\hbar\omega} \quad (cm^{-1}) \tag{3.46}$$

The value of α expressed in Eqs. 3.42 or 3.45 corresponds to a fixed photon energy $\hbar\omega$ when the semiconductor is at 0°K and the values of the Fermi-Dirac distribution functions in the conduction and valence bands are zero and unity, respectively. To express the temperature dependence of α, one must include the Fermi functions in the summation over all energies used to calculate the probability function $P(\hbar\omega)$. As a result, the absorption coefficient expressed by Eq. 3.42 or 3.45 must include a factor $[f_p(\mathcal{E}(\mathbf{k}')) - f_n(\mathcal{E}(\mathbf{k}''))]$, where f_n and f_p are the Fermi functions in the respective bands. It is also assumed that the semiconductor is very pure. Impurity atoms will induce scattering, which will relax the momentum conservation requirements. In addition, impurity levels will give rise to bandtail states that will result in a finite value of α for photon energies $\hbar\omega < \mathcal{E}_g$. The absorption coefficient α expressed in Eq. 3.46 corresponds to a measured value for photon energies $\hbar\omega > \mathcal{E}_g$ under the ideal conditions described above.

Similarly, starting with the matrix element for the direct forbidden transition in Eq. 3.34, it can be shown that the transition probability is given by

$$P(\hbar\omega) = \frac{q^2|\mathbf{A}|^2}{12\pi m_0^2\hbar^4}(2m_r^*)^{5/2} f'_{CV}(\hbar\omega - \mathcal{E}_g)^{3/2} \tag{3.47}$$

It is important to note that the probability is proportional to $(\hbar\omega - \mathcal{E}_g)^{3/2}$. The absorption coefficient is given by

$$\alpha(\hbar\omega) = C_2 n_r^{-1} \left(\frac{2m_r^*}{m_0}\right)^{5/2} \frac{f'_{CV}}{\hbar\omega}(\hbar w - \mathcal{E}_g)^{3/2} \tag{3.48}$$

where

$$C_2 = \frac{q^2 m_0^{1/2}}{6\pi\hbar^2\epsilon_0 c} \tag{3.49}$$

f'_{CV} is the oscillator strength for the forbidden transition, and its value is much less than unity. Again, if $\hbar\omega$ and $\mathcal{E}_g$ are expressed in eV,

$$\alpha(\hbar\omega) = 1.76 \times 10^5 \frac{n_r^{-1}}{\hbar\omega} \left(\frac{2m_r^*}{m_0}\right)^{5/2} f'_{CV}(\hbar\omega - \mathcal{E}_g)^{3/2} \quad (cm^{-1}) \tag{3.50}$$

Vizio EX1016 Page 0029

From experimental results it is evident that the direct optical transition corresponding to the absorption of a photon with $\hbar\omega > \mathcal{E}_g$ is dominantly observed in most direct-bandgap semiconductors.

EXAMPLE 3.2

Objective. To calculate α for the allowed transitions in GaAs at a photon energy $\hbar\omega = 1.52$ eV.

Assuming $\mathcal{E}_g = 1.5$ eV at 0°K, from Eq. 3.46 $\alpha = 5.2 \times 10^3$ cm^{-1}.

The following general comments may be made regarding the relations given above. For very small values of $(\hbar\omega - \mathcal{E}_g)$ and in a pure semiconductor, *excitons* are generally formed due to the Coulomb interaction between electrons and holes. The expressions for the matrix element, transition probability, and absorption coefficient given above are only for band-to-band transitions and do not consider exciton-related processes. Second, the equations are valid insofar as the parabolic band approximation is true, and are therefore generally not valid for higher-lying regions of the direct band. For large values of photon energy, contributions from the satellite valleys can become important. In most III–V semiconductors, the valence band is degenerate at the zone center. Strictly, the effect of such degenerate bands will contribute to the absorption. For a heavily doped semiconductor, the absorption edge moves to higher energies, as illustrated in Fig. 3.4. In addition, bandtail states are formed. It is important to note that the matrix element for transitions between bandtail states is different from that involving free-electron and hole states. Consequently, the **k**-selection rule does not apply.

With variation of temperature, the variation of the absorption coefficient follows the variation of the bandgap with temperature given by Eq. 2.15. Finally, it should

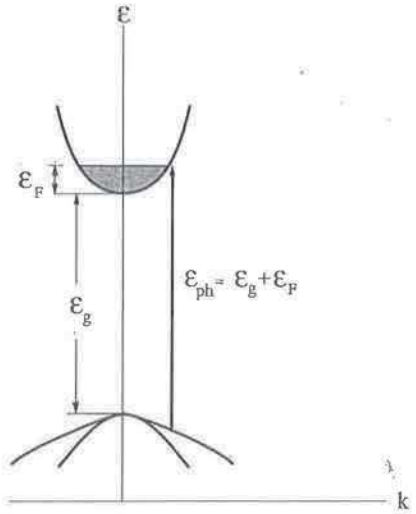


Figure 3.4 Simplified illustration of absorption of photons of energy larger than the bandgap in a degenerately doped n-type semiconductor.

Vizio EX1016 Page 0030

be remembered that the process of radiation is complementary to absorption and is governed by similar equations for transition probability and oscillator strength.

3.2.2 Indirect Intrinsic Transitions

The momentum or wavevector change required in an indirect transition may be provided by single or multiple phonons, although the probability of the latter to occur is very small. As seen in Chapter 2, there are optical and acoustic phonons. Each of these has transverse and longitudinal modes of vibrations, with characteristic energy and momentum. The indirect transition process is illustrated in Fig. 3.2(c). Conservation of momentum requires

$$\mathbf{k}'' \pm \mathbf{k}_p = \mathbf{k}' + \mathbf{k}_{ph} \tag{3.51}$$

where $\mathbf{k}''$ and $\mathbf{k}'$ are the electron wavevectors for the final and initial states, $\mathbf{k}_p$ is the wavevector of the phonon, and $\mathbf{k}_{ph}$ is the wavevector of the absorbed photon. Since the latter is small, the conservation of momentum for an indirect transition can be expressed as

$$\mathbf{k}'' - \mathbf{k}' = \mp \mathbf{k}_p \tag{3.52}$$

Similarly, the conservation of energy for the two cases of phonon emission and absorption can be expressed as

$$\hbar\omega_e = \mathcal{E}_C - \mathcal{E}_V + \mathcal{E}_p \tag{3.53}$$

$$\hbar\omega_a = \mathcal{E}_C - \mathcal{E}_V - \mathcal{E}_p \tag{3.54}$$

where the left-hand side represents the energy of the photon absorbed. Note that in the first case of phonon emission, the energy of the absorbed photon could be equal to the direct gap at or very near $\mathbf{k} = 0$. From this energy state the electron finally reaches the indirect valley by phonon scattering. The intermediate energy state of the electron is termed a *virtual* state, in which the carrier resides until a phonon of the right energy and momentum is available for the scattering process. Indirect transition probabilities involving virtual states can be calculated using a second-order time-dependent perturbation theory. However, there is a process to slightly counterbalance the low transition probability, which is often overlooked. From Eqs. 3.53 and 3.54 it is evident that the initial and final states of the electron in the valence and conduction bands, respectively, can have an energy range given by $\hbar(\omega_{ph} \pm \omega_p)$, where ω_p and ω_{ph} correspond to the angular frequencies of the phonon and photon, respectively. The total probability is obtained by a summation over these energy states, as long as each particular transition conserves energy between initial and final states.

For a transition with phonon absorption,

$$\alpha_a(\hbar\omega) \propto \frac{(\hbar\omega - \mathcal{E}_g + \mathcal{E}_p)^2}{e^{\mathcal{E}_p/k_BT} - 1} \tag{3.55}$$

for a photon energy $\hbar\omega > (\mathcal{E}_g - \mathcal{E}_p)$. Similarly, for a transition with phonon emission the absorption coefficient is given by

$$\alpha_e(\hbar\omega) \propto \frac{(\hbar\omega - \mathcal{E}_g - \mathcal{E}_p)^2}{1 - e^{-\mathcal{E}_p/k_BT}} \tag{3.56}$$

Vizio EX1016 Page 0031

for $\hbar\omega > (\mathcal{E}_g + \mathcal{E}_p)$. Since for $\hbar\omega > (\mathcal{E}_g + \mathcal{E}_p)$ both phonon emission and absorption are possible, under these conditions

$$\alpha(\hbar\omega) = \alpha_a(\hbar\omega) + \alpha_e(\hbar\omega) \tag{3.57}$$

The temperature dependence of the absorption coefficient is illustrated in Fig. 3.5. At very low temperatures, the density of phonons available for absorption becomes small and therefore α_a is small. With increase of temperature, α_a increases. The shift of the curves to lower energies with increase of temperature reflects the temperature dependence of $\mathcal{E}_g$. In fact, the plots of $\sqrt{\alpha_e}$ and $\sqrt{\alpha_a}$ extrapolate to the energy axis at $(\mathcal{E}_g + \mathcal{E}_p)$ and $(\mathcal{E}_g - \mathcal{E}_p)$, respectively. This is a convenient technique to experimentally determine the bandgap.

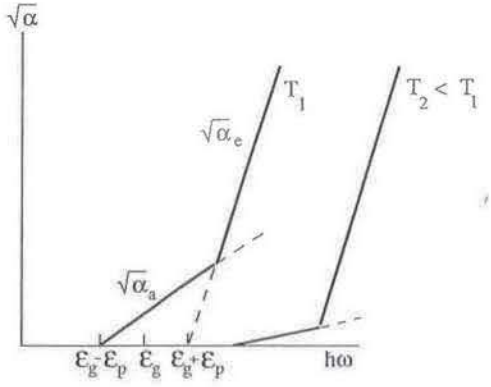


Figure 3.5 Energy-dependent absorption coefficient due to phonon emission and absorption as a function of temperature.

3.2.3 Exciton Absorption

In very pure semiconductors, where the screening effect of free carriers is almost absent, electrons and holes produced by the absorption of a photon of near-bandgap energy pair to form an *exciton*. This is the free exciton. The binding energy of the exciton, $\mathcal{E}_{ex}$, is calculated by drawing analogy with the Bohr atom for an impurity center, and is quantized. It is therefore expressed as

$$\begin{aligned}\mathcal{E}_{ex}^{l} &= \frac{-m_r^* q^4}{2(4\pi\epsilon_r\epsilon_o\hbar)^2} \cdot \frac{1}{l^2}, \quad l = 1, 2, 3, \cdots \\ &= \frac{-13.6}{l^2}\frac{m_r^*}{m_0}\left(\frac{1}{\epsilon_r}\right)^2 (eV).\end{aligned} \tag{3.58}$$

Here m_r^* is the reduced effective mass of the exciton given by Eq. 3.36 and l is an integer.

The optical excitation and formation of excitons usually manifest themselves as a series of sharp resonances (peaks) at the low energy side of the band edge in the absorption spectra of direct bandgap semiconductors. The total energy of the exciton is given by

$$\mathcal{E}_{ex} = \frac{\hbar^2 \mathbf{k}_{ex}^2}{2(m_e^* + m_h^*)} - \mathcal{E}_{ex}^{l} \tag{3.59}$$

Vizio EX1016 Page 0032

where the first term on the right is the kinetic energy of the exciton. The kinetic energy contributes to a slight broadening of the exciton levels. For a direct transition conservation of momentum requires that $\mathbf{k}_{ex} \cong 0$. This is because the electron and hole must move in the same direction. Usually a sharp line transition is observed for direct excitonic transitions, which broadens with increase of temperature. Data on the excitonic absorption in pure GaAs are shown in Fig. 3.6.

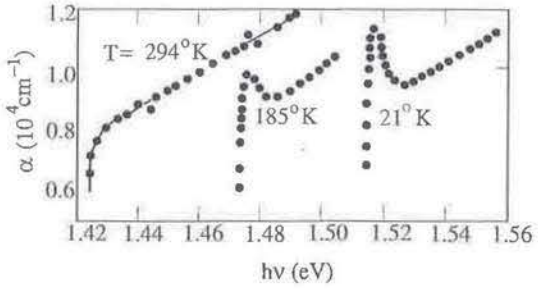


Figure 3.6 Absorption coefficient measured near the band edge of GaAs at T = 294, 185, and 21°K. The two absorption peaks at $h\nu$ slightly below the respective band gap at 185 K and 21 K are due to bound excitons (from M. D. Sturge, *Physical Review*, **127**, 768, 1962).

In indirect bandgap semiconductors excitons may also be formed with the absorption or emission of a phonon. In this case the center of gravity of the exciton may have a finite momentum $\hbar\mathbf{k}_{ex}$, conserved by an interacting phonon. Again, transverse and longitudinal acoustic and optical phonons may participate. An increase in absorption coefficient is obtained near the band edge due to exciton absorption, given by

$$\hbar\omega = \mathcal{E}_g \mp \mathcal{E}_p - \mathcal{E}_{ex} \tag{3.60}$$

where the two signs of the second term on the right hand side correspond to the cases of phonon absorption or emission. Exciton-related transitions are seen in the absorption spectra of an indirect bandgap semiconductor as a large number of steps near the absorption edge. Note that steps are observed instead of peaks, as in direct bandgap semiconductors, because the interacting phonons allow the carrier transition between states with equal $d\mathcal{E}/dk$ in the valence and conduction bands at energies greater than in direct band gap semiconductors where usually excitons are formed at the zone center ($d\mathcal{E}/dk = 0$).

Excitons are formed in very pure semiconductors at low temperatures. In fact, excitons were not observed in semiconductors until epitaxial techniques enabled the growth of very pure crystals. In such crystals, the very few unintentional impurities that are present—donors and acceptors—are neutral. If an electric field is applied, it can ionize these impurities, and the additional charge modifies the bandedge potential. This is seen in the experimental absorption spectra as a change in the slope of the absorption edge. In addition, the ionized carriers screen the Coulomb interaction between the electrons and holes, thereby inhibiting or preventing the formation of excitons. This is observed in the experimental absorption spectra as a disappearance of the excitonic resonances, peaks or steps, as the case may be.

3.2.4 Donor-Acceptor and Impurity-Band Absorption

Intentionally or unintentionally, both donors and acceptors are simultaneously present in a semiconductor, and any semiconductor is usually always compensated to some

Vizio EX1016 Page 0033

degree. Depending on the temperature and the state of occupancy of the impurity levels, it is possible to raise an electron from the acceptor to the donor level by absorbing a photon. This process is shown in Fig. 3.7. The energy of the photon absorbed is given by

$$\hbar\omega = \mathcal{E}_g - \mathcal{E}_D - \mathcal{E}_A + \frac{q^2}{\epsilon_0 \epsilon_r r} \tag{3.61}$$

where the last term on the right-hand side accounts for the Coulomb interaction between the donor and acceptor atoms in substitutional sites, which results in a lowering of their binding energies. This can be understood as follows. Assume that at very low temperatures the donor and acceptor atoms are neutral. If they are brought closer together, the additional orbiting electron of the donor becomes "shared" by the acceptor, as in a covalent bond, and both become more ionized, resulting in a lowering of their binding energy. Also, it is important to remember that since the donor and acceptor atoms are located at discrete substitutional sites in the lattice, r varies in finite increments, being the smallest for nearest neighbors. Therefore, for the ground state of the impurities, the energies $\mathcal{E}_D$ and $\mathcal{E}_A$ correspond to the most distant pairs and $\hbar\omega \cong \mathcal{E}_g - \mathcal{E}_D - \mathcal{E}_A$. For fully ionized impurities, such as for nearest neighbors, the excited states may lie within the respective band and it is possible that $(q^2/\epsilon_0\epsilon_r r) > \mathcal{E}_D + \mathcal{E}_A$. At low temperatures the absorption resonances modify the bandedge absorption, with the lowest energy transitions for the most distant pairs and higher-energy transitions for nearer pairs. However, because the resonances occur so close to the absorption edge, they are not always very clearly defined. The pair transitions are more clearly identified in emission experiments.

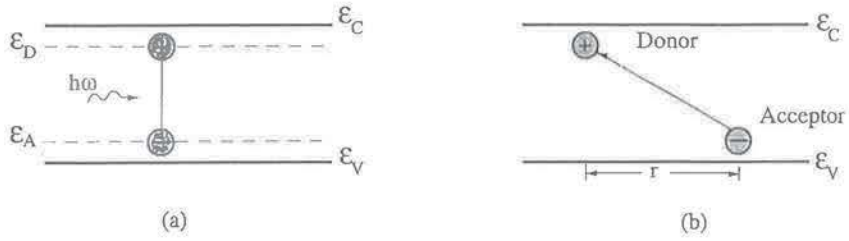


Figure 3.7 Illustration of photon absorption due to a donor-acceptor transition. The separation between the impurity centers, r, is shown in (b).

High-energy (near-bandgap) transitions can occur between ionized impurity levels and the opposite bandedge, as illustrated in Fig. 3.8. The photon energy absorbed is $\hbar\omega \cong \mathcal{E}_g - \mathcal{E}_b$, where $\mathcal{E}_b$ is the binding energy of the donor or acceptor level. It should be noted that the impurity levels need to be ionized. Since the transition occurs between a discrete impurity level and a band of energies, the transitions are observed as shoulders on the low-energy side of the absorption edge. In the emission spectra, these transition are observed as peaks. As in a band-to-band transition, phonons need to be involved in impurity-band transitions in an indirect bandgap semiconductor for momentum conservation.

Vizio EX1016 Page 0034

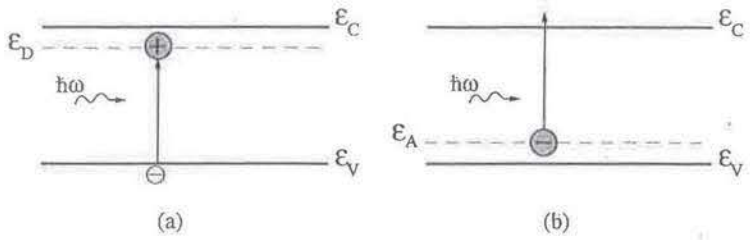


Figure 3.8 Illustration of (a) donor-band and (b) acceptor-band absorption transitions.

The absorption spectrum is largely altered if the doping level is increased and gradually taken to the point of degeneracy. For example, in a degenerately doped n-type semiconductor, the Fermi level $\mathcal{E}_{fn}$ is above the conduction bandedge. If the semiconductor is direct bandgap, as shown in Fig. 3.4, then, for the conservation of momentum, the transition resulting from the absorption of a photon will involve states in the conduction band that are at or higher than $\mathcal{E}_g + \mathcal{E}_{fn}$. This shift of the absorption to higher energies due to doping-induced band-filling is called the *Burstein-Moss shift*. An indirect bandgap semiconductor will be similarly affected, except that phonons need not be involved in the transition. Momentum is conserved by impurity scattering.

Degeneracy in semiconductors not only pushes the Fermi level into the band, but also results in a shrinkage of the bandgap. This effect is more commonly known as *bandtailing*, which results in an exponentially increasing absorption edge with photon energy as shown is Fig. 3.9 for GaAs.

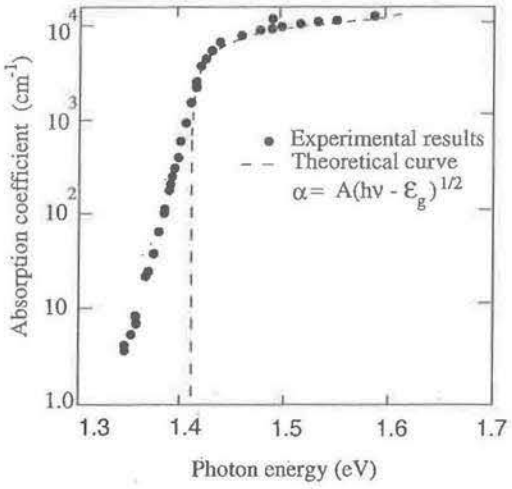


Figure 3.9 Absorption edge of GaAs at room temperature (from T. S. Moss, *Journal of Applied Physics*, **32**, 2136, 1961).

3.2.5 Low-Energy (Long-Wavelength) Absorption

Several types of transitions involving shallow impurity levels, bandedges, split bands, and free carriers give rise to resonances at very small energies in the absorption

Vizio EX1016 Page 0035

spectra. These are observed as steps or peaks in the long-wavelength region of the absorption spectra. The different processes are briefly described below.

3.2.5.1 Impurity-Band Transitions. We have seen impurity band transitions that have energies close to the bandgap. These higher-energy impurity-band transitions usually require that the impurity levels are ionized (or empty). At low temperatures, when these shallow impurity levels are usually filled with their respective carriers, these carriers can be excited to the respective bandedge by a photon (Fig. 3.10). For this absorption process the energy of the photon must be at least equal to the ionization energy of the impurity. This energy usually corresponds to the far infrared region of the optical spectrum. Peaks corresponding to such impurity-band transitions have been observed in many semiconductors.

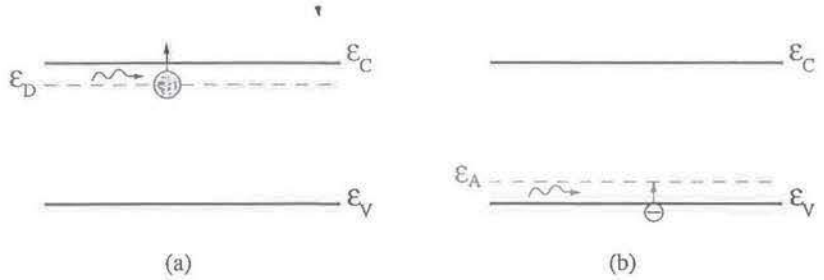


Figure 3.10 Low-energy (a) donor-band and (b) acceptor-band absorption transition.

3.2.5.2 Intraband Transitions. At the zone center the valence band structure of most semiconductors consists of the light-hole (LH), the heavy-hole (HH) bands, and the split-off (SO) band. The three subbands are separated by spin-orbit interaction. In a p-type semiconductor the valence band is filled with holes and the occupancy of the different bands depend on the degree of doping and the position of the Fermi level. Absorption of photons with the right energy can result in transitions from LH to HH, SO to HH, and SO to LH bands, depending on the doping and temperature of the sample. These transitions have been observed experimentally. They are normally not observed in n-type semiconductors.

3.2.5.3 Free-Carrier Absorption. As the name suggests, this mechanism involves the absorption of a photon by the interaction of a free carrier within a band, which is consequently raised to a higher energy. The transition of the carrier to a higher energy within the same valley must conserve momentum. This momentum change is provided by optical or acoustic phonons, optical phononse, or by impurity scattering. Free-carrier absorption usually manifests in the long-wavelength region of the spectrum as a monotonic increase in absorption with a wavelength dependence of the form λ^p, where p ranges from 1.5 to 3.5. The value of p depends on the nature of the momentum-conserving scattering (i.e., the involvement of acoustic phonons,

Vizio EX1016 Page 0036

optical phonons, or ionized impurities). The absorption coefficient due to free-carrier absorption can be expressed as

$$\alpha = \frac{Nq^2\lambda^2}{4\pi^2 m^* n_r c^3 \epsilon_0}\left\langle\frac{1}{\tau}\right\rangle \tag{3.62}$$

where N is the free-carrier concentration, n_r is the refractive index of the semiconductor, and $\left\langle\frac{1}{\tau}\right\rangle$ is the average value of the inverse of the relaxation time of the scattering process.

As a concluding note to this section, the absorption coefficients for different elemental and III–V compound semiconductors at room temperature are shown in Fig. 3.11.

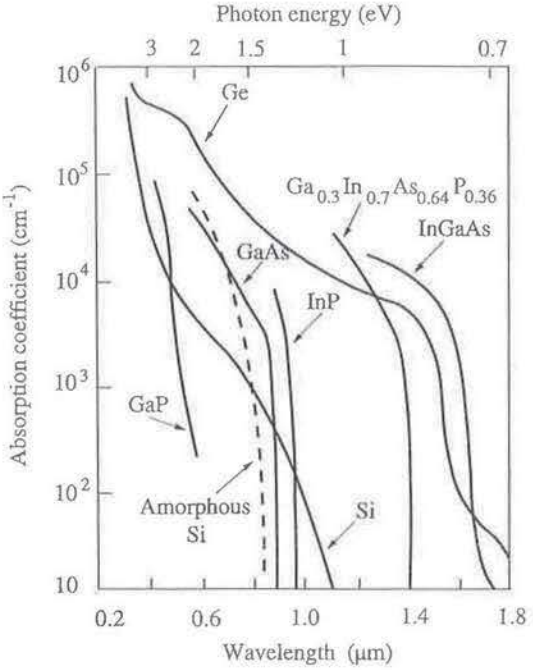


Figure 3.11 Near-bandgap absorption spectra of different semiconductors (from M. Shur, *Physics of Semiconductor Devices*, ©1990. Reprinted by permission of Prentice Hall, Englewood Cliffs, New Jersey).

3.3 EFFECT OF ELECTRIC FIELD ON ABSORPTION: FRANZ-KELDYSH AND STARK EFFECTS

The change in absorption in a semiconductor in the presence of a strong electric field is the *Franz-Keldysh effect,* which results in the absorption of photons with energies less than the bandgap of the semiconductor. The energy bands of a semiconductor in the presence of an electric field E and with an incident photon of energy $\hbar\omega < \mathcal{E}_g$ are shown in Figs. 3.12(a) and (b). It is important to note that at the classical turning points marked A and B, the electron wavefunctions change from oscillatory to decaying behavior. Thus, the electron in the energy gap is described by an exponentially decaying function $u_k e^{jkx}$, where $\mathbf{k}$ is imaginary. With increase of electric field, the distance AB decreases and the overlap of the wavefunctions within the gap increases.

Vizio EX1016 Page 0037

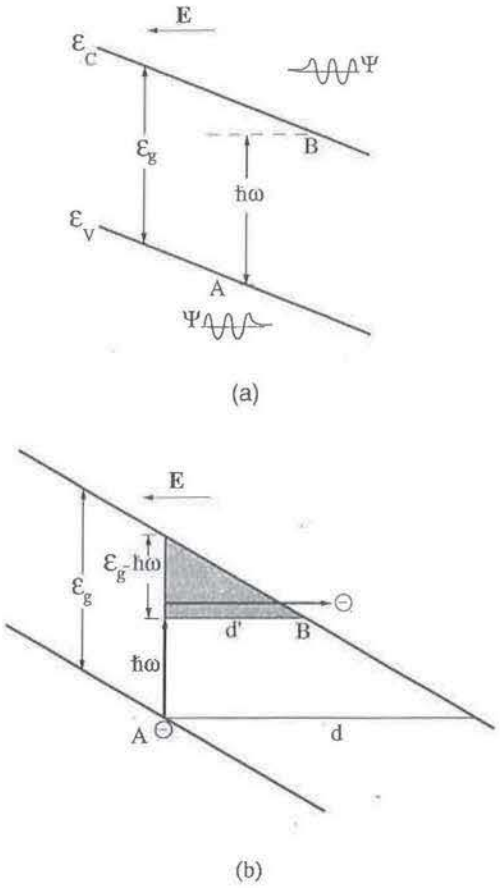


Figure 3.12 (a) Bending of bands due to an applied electric field and (b) absorption of photon with $\hbar\omega < \mathcal{E}_g$ due to carrier tunneling (Franz-Keldysh effect).

In the absence of a photon, the valence electron has to tunnel through a triangular barrier of height $\mathcal{E}_g$ and thickness d, given by $d = \mathcal{E}_g/qE$. With the assistance of an absorbed photon of energy $\hbar\omega < \mathcal{E}_g$, it is evident that the tunneling barrier thickness is reduced to $d' = (\mathcal{E}_g - \hbar\omega)/qE$, the overlap of the wavefunctions increases further, and the valence electron can easily tunnel to the conduction band. The net result is that a photon with $\hbar\omega < \mathcal{E}_g$ is absorbed. One has to keep in mind, of course, the conservation of momentum in these transitions and in this case the transverse component of momentum is conserved. The Franz-Keldysh effect is therefore, in essence, photon assisted tunneling.

It can be shown that the electric-field dependent absorption coefficient is given by[†]

$$\alpha = K(E')^{1/2}(8\beta)^{-1}\exp\left(-\frac{4}{3}\beta^{3/2}\right) \tag{3.63}$$

Here $E' = \left(\frac{q^2E^2\hbar^2}{2m_r^*}\right)^{1/3}$, $\beta = \frac{\mathcal{E}_g - \hbar\omega}{E'}$, and K is a material-dependent parameter that has

[†]S. Wang, *Fundamentals of Semiconductor Theory and Device Physics*, Prentice Hall, Englewood Cliffs, NJ, 1989.

Vizio EX1016 Page 0038

a value of 5×10^4 cm^{-1} (eV)$^{-1/2}$ in GaAs. Although not derived here, the various terms in Eq. 3.63 can be examined qualitatively. The exponential term is the transmission coefficient (or tunneling probability) of an electron through a triangular barrier of height ($\mathcal{E}_g - \hbar\omega$) and can be obtained from the well-known Wentzel-Kramers-Brillouin (WKB) approximation. The other factors are related to the upward transition of an electron due to photon absorption. Substituting appropriate values for the different parameters, it is seen that in GaAs $\alpha = 4cm^{-1}$ at a photon energy of $\mathcal{E}_g - 20$ meV with electric field $E \sim 10^4$ V/cm. This value of absorption coefficient is much smaller than the values of α at the band edge at zero field. Therefore, the Franz-Keldysh effect will be small unless $E \geq 10^5$ V/cm.

The *Stark effect* refers to the change in atomic energy upon the application of an electric field. The electric field affects the higher-order, or outer, orbits of the precessing electrons so that the center of gravity of the elliptical orbit and the focus are displaced with respect to each other and linearly aligned in the direction of the electric field. As a result, there is a splitting of the energy of the outer $2s$ or $2p$ states, and the energy shift is simply given by $\Delta\mathcal{E} = qdE$, where d is the eccentricity of the orbit. This is the *linear Stark effect*. The effect of the electric field on ground state orbits also leads to an energy shift of the state, and this is the *quadratic* or *second-order Stark effect.*

3.4 ABSORPTION IN QUANTUM WELLS AND THE QUANTUM-CONFINED STARK EFFECT

In a bulk semiconductor the exciton binding energy is given by Eq. 3.58. For example in GaAs, upon substitution of the effective mass and dielectric constant, a ground-state binding energy of 4.4 meV is obtained. This is comparable to the thermal broadening of $\sim$ 4 meV produced by optical phonon scattering and inhomogeneous broadening, to be discussed in Chapter 6. In other words, the exciton dissociates in a very short time (a few hundred femtoseconds) and is hardly detected in the absorption spectra at room temperature, except in very pure samples.

The situation is drastically altered in a quantum well. In a single-quantum well (SQW) or multiple-quantum well (MQW) with thick barriers ($\geq$ 100 Å), electrons and holes are confined in the region defined by the well width, and the overlap of their wavefunctions is increased. This results in an increase in the oscillator strength of the interband transitions between the discrete electron and hole energy bound states, which are produced by the size quantization. Consequently, strong resonances corresponding to the heavy-hole and light-hole transitions are seen near the bandedge of the well material even at room temperature. Shown in Fig. 3.13 are the measured and calculated absorption spectra of an $In_{0.53}Ga_{0.47}As$ (100 Å)/$In_{0.52}Al_{0.48}As$ (100 Å) MQW lattice matched to InP, in which the resonances for l = 2 and 3, are also clearly seen. It may be noted that distinct resonances are seen for heavy-hole and light-hole transitions. This is because a splitting between the HH and LH bands occurs at the zone center due to the difference in the energy eigenvalues resulting from different hole masses.

Vizio EX1016 Page 0039

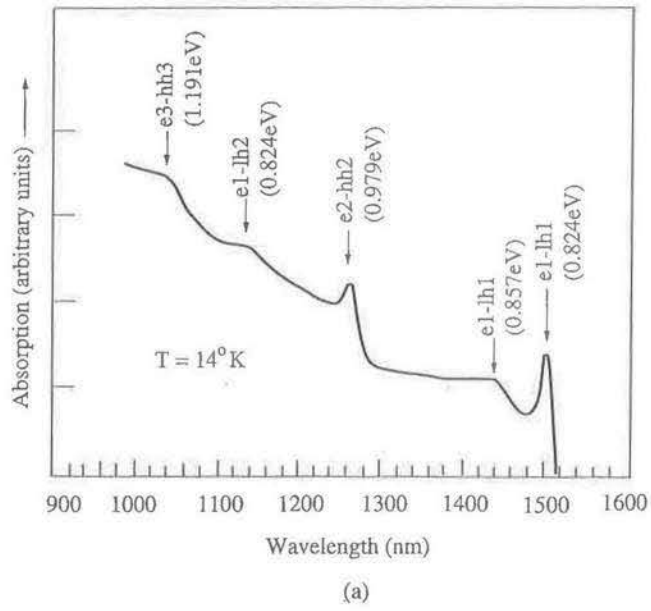


Figure 3.13 (a) Absorption spectrum of 40-period lattice-matched InGaAs/InAlAs MQW (measured by the author and co-workers) at 14°K; and (b) the calculated transitions based on a finite square well model at room temperature (from S. Gupta et al., *Journal of Applied Physics*, **69**, 3219, 1991).

In a two-dimensional quantum well the exciton is compressed like a pancake. However, since typical well dimensions are $\sim$ 100 Å and the exciton diameter is $\geq$ 300 Å, there is some penetration of the exciton wavefunction into the barrier material. This is depicted in Fig. 3.14. In the limit, for small well widths, the situation becomes similar to a three-dimensional solid. It has been shown that for a purely two-dimensional exciton, its binding energy is four times the bulk value, given by Eq. 3.58. However, because of the extension of its wavefunction into the barrier, the binding energy in practical SQW and MQW structures ranges between $2\mathcal{E}_{ex}$ and $3\mathcal{E}_{ex}$. Since this binding energy is much larger than the thermal broadening, the exciton

Vizio EX1016 Page 0040

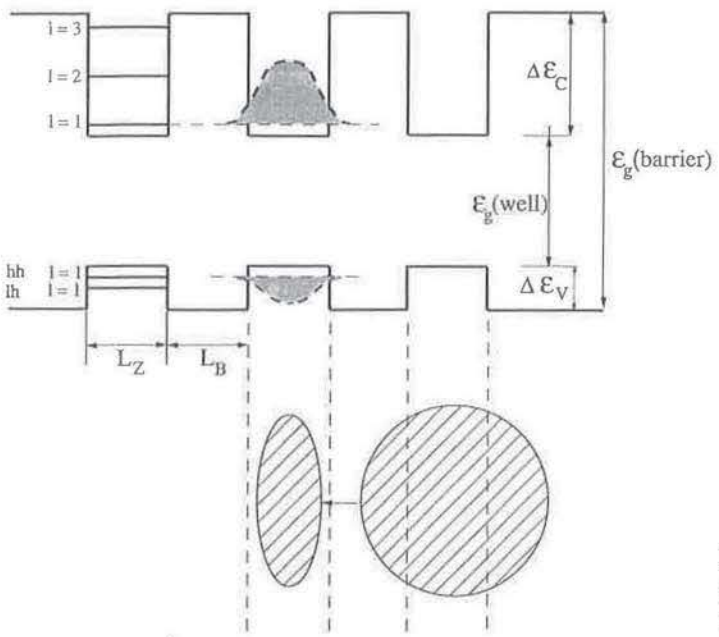


Figure 3.14 A typical multiquantum well and the compression of the bulk exciton in the well region.

resonances are clearly seen in the absorption spectra even at room temperature. The coefficient for intersubband absorption in a quantum well can be expressed in cgs units as†

$$\alpha_{2D}(\hbar\omega) = \frac{4\pi^2 q^2 \hbar}{n_r c m_0^2 \hbar\omega} p_{CV}^2 a_p \frac{N_{2D}(\hbar\omega - \mathcal{E}_g)}{L_z}$$

$$= 1.77 \times 10^{-28} \frac{a_p}{n_r} \left(\frac{f_{CV}}{\hbar\omega}\right) \frac{N_{2D}}{L_z} \tag{3.64}$$

where L_z is the width of the well, $N_{2D} = m_r^*/\pi\hbar^2$ is the 2-D density of states, a_p is a factor due to polarization dependence of the matrix elements (in GaAs $a_p = 1/2$). This equation is valid for a single pair of conduction and valence subbands. Note also that the value of α is constant and will remain so till the next subband energies are reached. This is evident in Fig. 3.13. It may seem from Eq. 3.64 that the absorption coefficient can be increased by decreasing L_z. However, as L_z is decreased, the electron and hole wavefunctions spread outside the well and the overlap decreases. Therefore, for every material system there is an optimum well size. For example, for GaAs this $L_z \sim 50$Å and for $In_{0.53}Ga_{0.47}As$ $L_z \sim 80$Å.

The ground-state wavefunctions of the electron and hole subband with no applied transverse field (in the direction perpendicular to the layers) are shown in Fig. 3.15(a). With the application of an electric field several things happen. The electron and hole wavefunctions are separated and pushed toward opposite sides of the well, as

†J. Singh, *Physics of Semiconductors and Their Heterostructures*, McGraw-Hill, New York, 1993.

Vizio EX1016 Page 0041

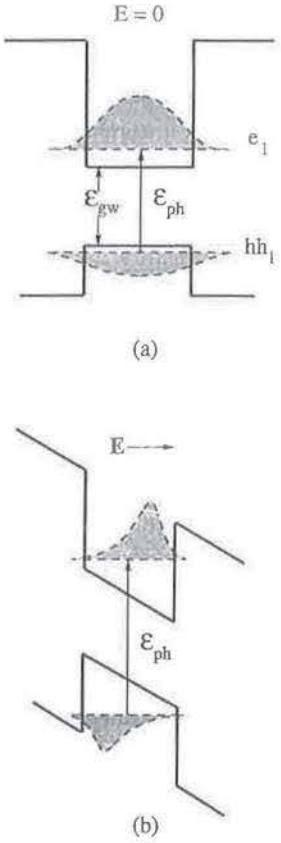


Figure 3.15 Absorption in a quantum well in the (a) absence and (b) presence of a transverse electric field.

shown in Fig. 3.15(b). The reduced overlap results in a corresponding reduction in absorption and in luminescence. The probability of carriers tunneling out of the wells also increases, resulting in a decrease in carrier lifetimes and a broadening of the absorption spectra. The transition energy is given by

$$\mathcal{E}_{ph} = \mathcal{E}_e + \mathcal{E}_h + \mathcal{E}_{gw} - \mathcal{E}_{ex} \tag{3.65}$$

where $\mathcal{E}_e$ and $\mathcal{E}_h$ are the electron and hole subband energies. With the application of moderate electric fields ($10^4 - 10^5$ V/cm), there is little change in $\mathcal{E}_{ex}$ and a very small change in $\mathcal{E}_{gw}$ due to the Stark effect in the well material. However, due to the modification of the envelope functions, there is a reduction in $\mathcal{E}_e$ and $\mathcal{E}_h$, the subband energies. This results in a shift of the absorption spectrum to lower energies, including the heavy- and light-hole resonances. The shifts are much larger than the Stark shift in bulk materials and is $\sim$ 20 meV for $E = 10^5$ V/cm in a 100 Å $GaAs/Al_{0.3}Ga_{0.7}As$ quantum well. Experimental data are shown in Fig. 3.16 and the phenomenon is known as the *quantum confined Stark effect* (QCSE). As we shall see in Chapter 11, the effect can be used for the design and realization of very efficient light modulators.

Vizio EX1016 Page 0042

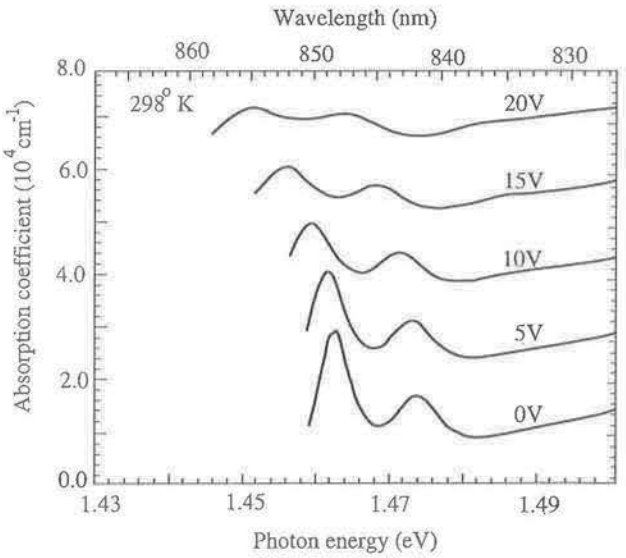


Figure 3.16 Room temperature optical absorption coefficients measured by the author and co-workers in a 100-Å $GaAs/Al_{0.3}Ga_{0.7}As$ MQW structure at various bias levels (from J. Singh et al., *Journal of Lightwave Technology*, **6**, 818, ©1988 IEEE).

3.5 THE KRAMER-KRÖNIG RELATIONS

The complex refractive index of a semiconductor is given by

$$\mathbf{n}_c = \mathbf{n}_r + j\mathbf{k}_a \tag{3.66}$$

where $\mathbf{n}_r$ and $\mathbf{k}_a$ are the real and imaginary parts of $\mathbf{n}_c$. It may be noted that $\mathbf{k}_a$ is an attenuation or damping factor and is a measure of the loss in power of a wave propagating through the semiconductor. Therefore, $\mathbf{k}_a$ is directly proportional to the absorption coefficient α of the material. If a plane wave described by

$$\begin{aligned}\mathbf{E} &= \mathbf{E}_0 \exp\left[j\omega\left(\frac{x}{\vartheta} - t\right)\right] \\ &= \mathbf{E}_0 \exp\left[j\omega\left(\frac{n_r x}{c} - t\right)\right] \exp\left(-\frac{\omega \mathbf{k}_a x}{c}\right)\end{aligned} \tag{3.67}$$

is propagating in the x-direction in a semiconductor with a velocity $\vartheta = \frac{c}{n_c}$, then it can be shown (Problem 3.7) that the absorption coefficient is given by

$$\alpha = \frac{2\omega \mathbf{k}_a}{c} = \frac{4\pi \nu \mathbf{k}_a}{c} \tag{3.68}$$

In a material whose conductivity $\sigma \to 0$, the refractive index is related to the dielectric constant by the relation

$$n_r \cong \sqrt{\epsilon_r} \tag{3.69}$$

where ϵ_r is the static dielectric constant ϵ_s and

$$k_a \cong 0 \tag{3.70}$$

It is also useful to know that the refractive index is inversely related to the bandgap of a semiconductor.

The complex dielectric constant of a material is given by

$$\epsilon_r(\omega) = \epsilon_r'(\omega) + j\epsilon_r''(\omega) \tag{3.71}$$

Vizio EX1016 Page 0043

In time-invariant form, the electric field **E** and the electric flux density **D** are related by

$$\begin{aligned}\mathbf{D} &= \epsilon_0\left(1+\chi^e\right)\mathbf{E} \\ &= \epsilon_r\epsilon_o\mathbf{E}\end{aligned} \tag{3.72}$$

where χ^e is the electric *susceptibility*. The temporal response of **D** to a change or switching of **E** must include the change of polarization with time, and can be expressed by a causality relation of the type

$$\mathbf{D}(t) = \epsilon_o\epsilon'_\infty\delta(t)\mathbf{E}(t) + \int_{-\infty}^{t}\epsilon_o f(t-t')\mathbf{E}(t')dt' \tag{3.73}$$

in which the integral represents the response of the system at time t to the applied field **E** at a previous time t'. Note that $\epsilon(t)$ has a δ-function singularity at $t = 0$. ϵ'_∞ is the high-frequency dielectric constant, and $\epsilon''_\infty = 0$. Now the following Fourier transforms can be written for **D** and **E**:

$$\mathbf{D}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{D}(\omega)e^{-j\omega t}d\omega \tag{3.74}$$

and

$$\mathbf{E}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{E}(\omega)e^{-j\omega t}d\omega \tag{3.75}$$

Substituting these in Eq. 3.73 one gets

$$\int_{-\infty}^{\infty}\left[\mathbf{D}(\omega) - \epsilon_o(\epsilon'_\infty + f(\omega))\mathbf{E}(\omega)\right]e^{-j\omega t}d\omega = 0 \tag{3.76}$$

with

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}f(t)e^{j\omega t}dt \tag{3.77}$$

Equation 3.76 must be valid for all values of t. Therefore, the relation

$$\mathbf{D}(\omega) = \epsilon_o\left(\epsilon'_\infty + f(\omega)\right)\mathbf{E}(\omega) \tag{3.78}$$

is valid between the Fourier components, so that

$$\begin{aligned}\epsilon_r(\omega) &= \epsilon'_\infty + \frac{1}{2\pi}\int_{-\infty}^{\infty}f(t)e^{j\omega t}dt \\ &= \epsilon'_\infty + f(\omega)\end{aligned} \tag{3.79}$$

From Eq. 3.79, by application of the Cauchy theorem to the function $\left[\epsilon_r(\omega) - \epsilon'_\infty\right]/(\omega' - \omega)$, the following relations can be derived (Problem 3.14):

$$\epsilon'_r(\omega) = \epsilon'_\infty + \frac{2}{\pi}P\int_0^\infty\frac{\omega'\epsilon''_r(\omega')d\omega'}{\omega'^2 - \omega^2} \tag{3.80}$$

$$\epsilon''_r(\omega) = -\frac{2\omega}{\pi}P\int_0^\infty\frac{\left[\epsilon'_r(\omega') - \epsilon'_\infty\right]d\omega'}{\omega'^2 - \omega^2} \tag{3.81}$$

Vizio EX1016 Page 0044

where P is the principal value of the Cauchy integrals. These integrals are known as the Kramer-Krönig relations. A more relevant form is the relation between refractive index and absorption coefficient. By analogy with Eq. 3.80,

$$n_r(\mathcal{E}) - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\mathcal{E}' k_a(\mathcal{E}')}{\mathcal{E}'^2 - \mathcal{E}^2} d\mathcal{E}' \tag{3.82}$$

and, by virtue of the relation in Eq. 3.68,

$$n_r(\mathcal{E}) - 1 = \frac{ch}{2\pi^2} P \int_0^\infty \frac{\alpha(\mathcal{E}')}{\mathcal{E}'^2 - \mathcal{E}^2} d\mathcal{E}' \tag{3.83}$$

which enables the determination of the refractive index from the absorption spectrum. The dielectric constant and refractive index of some important binary III–V compounds are given in Table 3.1.

TABLE 3.1 DIELECTRIC CONSTANT AND REFRACTIVE INDEX IN SOME BINARY III-V COMPOUNDS.

Material	Static Dielectric Constant (ϵ_s)	High-Frequency Dielectric Constant (ϵ_∞)	Refractive Index (n_r) at Bandgap Energy
AlAs	10.06	8.5	3.17
GaP	11.11	9.11	3.45
GaAs	13.18	10.89	3.66
InP	12.56	9.61	3.45
InAs	15.15	12.3	3.52

3.6 RADIATION IN SEMICONDUCTORS

3.6.1 Relation between Absorption and Emission Spectra

In the last few sections of this chapter we have studied the various absorption processes. In these a photon is absorbed in the semiconductor, as a result of which an electron is usually raised from a lower-energy filled state to a higher-energy empty state, and in most cases the energy difference between the two states is equal to the energy of the absorbed photon. If the higher-energy level to which the electron is raised to is not the equilibrium state, then it will make a downward transition to the lower-energy empty state and emit electromagnetic radiation in the process. The energy of the radiation is very close to the energy difference between the higher and lower energy states. These are radiative transitions.

In principle, the reverse of all the absorption processes we have considered can occur to produce radiation. However, there is an important difference between absorption and emission spectra. While the absorption process can couple a broad energy range of filled and empty states (with momentum conservation) to produce a broad absorption spectrum, the emission process usually couples a narrow band of nonequilibrium filled states with a narrow band of empty states, to give a narrow emission spectrum. Therefore, a shoulder in the absorption spectrum can very well correspond

Vizio EX1016 Page 0045

to a narrow emission peak. Also, it is essential for the semiconductor to have a nonequilibrium population in the higher-energy states to produce a spontaneous emission spectrum. Depending on how the nonequilibrium state is produced, we defined different types of luminescence in Sec. 3.1.

The absorption and spontaneous emission spectra are related by the principle of detailed balance as calculated by van Roosbroeck and Shockley. At thermodynamic equilibrium, the rate of spontaneous photon emission $R_{sp}(\nu)$ at frequency ν in an interval $d\nu$ is given by

$$R_{sp}(\nu)d\nu = P_{abs}(\nu)\varphi(\nu)d\nu \tag{3.84}$$

where $P_{abs}(\nu)$ is the probability of absorbing a photon of energy $h\nu$ per unit time, and $\varphi(\nu)d\nu$ is the radiation density of frequency ν in an interval $d\nu$. This is obtained from Planck's radiation law (Appendix 4) as:

$$\varphi(\nu)d\nu = \frac{8\pi\nu^3 n_r^3}{c^3}\frac{1}{\exp\left(\frac{h\nu}{k_B T}\right) - 1}d\nu \tag{3.85}$$

The absorption probability $P(\nu)$ can be calculated in the following way. If the absorption coefficient of the photon is $\alpha(\nu)$ and it travels with a velocity $v = c/n_r$ in the material with refractive index n_r, then the mean lifetime of the photon is given by $\tau(\nu) = 1/\alpha(\nu)v$ and the absorption probability is given by

$$P(\nu) = \frac{1}{\tau(\nu)} = \alpha(\nu)v = \alpha(\nu)\frac{c}{n_r} \tag{3.86}$$

Substituting Eqs. 3.85 and 3.86 into Eq. 3.84, we get

$$R_{sp}(\nu)d\nu = \frac{\alpha(\nu)8\pi\nu^3 n_r^2}{c^2\left[\exp(h\nu/k_B T) - 1\right]}d\nu \tag{3.87}$$

which expresses the desired relation between absorption and emission spectra. Substitution of Eq. 3.68 leads to

$$R_{sp}(\nu)d\nu = \frac{32\pi^2 k_a(\nu)n_r^2\nu^4}{c^3\left[\exp\left(h\nu/k_B T\right) - 1\right]}d\nu \tag{3.88}$$

The total emission rate per unit volume is obtained by integrating Eq. 3.87 over all frequencies, or energies, as

$$R_{sp} = \frac{8\pi n_r^2(k_B T)^4}{c^2 h^4}\int_0^\infty \frac{\alpha(\nu)u^3}{e^u - 1}du \tag{3.89}$$

where $u = h\nu/k_B T$. Although derived for thermodynamic equilibirum, Eqs. 3.87 and 3.89 express the fundamental relation between absorption and emission spectra for any means of excitation. This formulation is valid for any transition between a higher-energy and a lower-energy state. The relation between emission and absorption spectra is schematically illustrated in Fig. 3.17.

Vizio EX1016 Page 0046

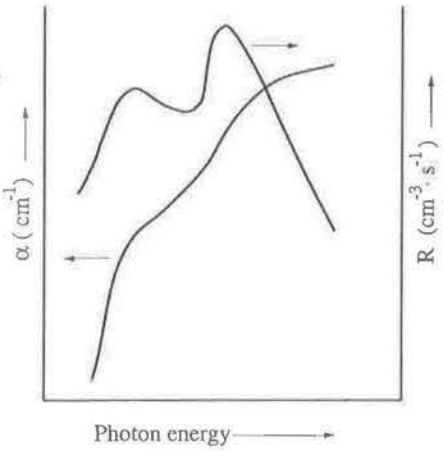


Figure 3.17 Schematic illustration of relation between emission and absorption spectra.

3.6.2 Stokes Shift in Optical Transitions

The Stokes shift is a difference in transition energy of the emission and absorption spectra resulting from defects in the material and, in general, partial nonradiative decay. The process can be understood with respect to the configuration coordinate diagram shown in Fig. 3.18. The lower and upper curves represent the energies of the lower and upper states of an optical transition as a function of distance that could be the ground and excited states of an impurity atom, a host lattice atom, or a deep-level trap. The point labeled 1 in the lower curve is the minimum energy, or equilibrium, position in the ground state. Due to photon absorption of energy $\hbar\omega_1$, an electron may be raised to the point 2 in the upper state ($\Delta \mathbf{k} = 0$), which is not the minimum energy configuration. The displacement Δx of the excited state may be caused by the defect or impurity potential. Therefore, the system relaxes to the state 3, which is at the lowest energy. This phenomenon is usually termed *lattice relaxation* and a phonon is involved. After living a mean lifetime the excited carrier returns to the ground state

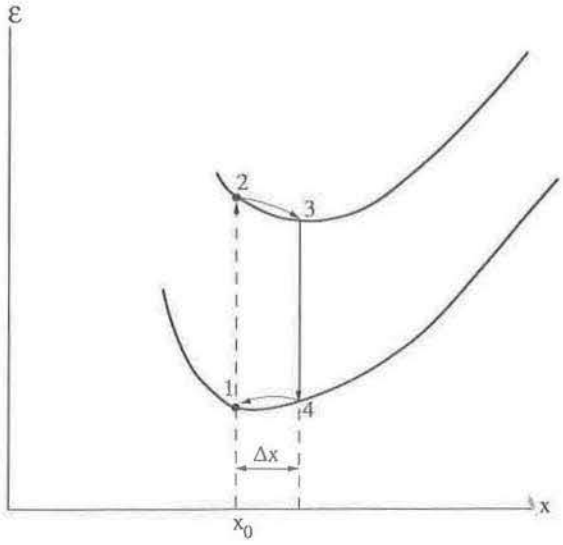


Figure 3.18 Configuration coordinate diagram illustrating the Stokes shift in a semiconductor.

Vizio EX1016 Page 0047

by radiative recombination at the point 4, which is again not the minimum energy configuration. The energy of the emitted photon is $\hbar\omega_2$. Therefore, the system relaxes again with phonon participation to the point 1. Both optical and/or acoustic phonons may be involved but usually optical phonons produce the largest change in energy per unit displacement. The energy difference of the photon absorbed and emitted, $(\hbar\omega_1 - \hbar\omega_2)$ is called the Stokes shift or Frank-Condon shift. This degradation of optical energy arises directly from imperfections in the material or interfaces, such as in a heterostructure or quantum well.

3.6.3 Near-Bandgap Radiative Transitions

3.6.3.1 Exciton Recombination. We have seen in Sec. 3.2.3 that electrons and holes produced by the absorption of a photon of near-bandgap energy can pair to form an exciton. Recombination of the electron-hole pair results in a narrow and sharp peak in the emission spectra. The energy of the emitted photon is

$$\hbar\omega = \mathcal{E}_g - \mathcal{E}_{ex} \tag{3.90}$$

where $\mathcal{E}_{ex}$ is quantized. In other words, in very pure crystals emission lines corresponding to the ground state and higher-order states may be seen. The process is shown in Fig. 3.19(a). In indirect bandgap semiconductors, a phonon needs to be involved, in the transition for momentum conservation, as shown in Fig. 3.19(b). Therefore, the probability of exciton recombination transitions is very low in indirect bandgap materials.

In semiconductors with impurities—donors and/or acceptors—present, the free exciton couples with the impurity atoms to produce *bound excitons*. Bound excitons produce sharp peaks at photon energies lower than that of the free exciton. Also, the linewidth of the bound exciton resonances are much smaller than that of the free excitons—almost by a factor of 10. In most semiconductors, free and bound exciton resonances are seen simultaneously in the emission spectra. Figure 3.20

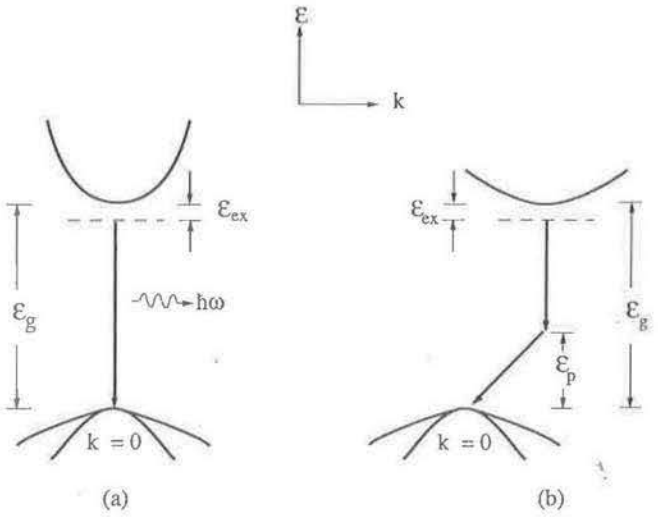


Figure 3.19 Exciton recombination in (a) direct bandgap and (b) indirect bandgap semiconductor.

Vizio EX1016 Page 0048

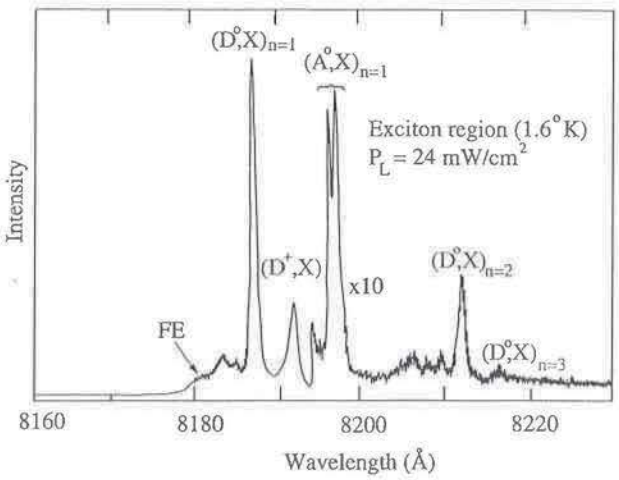


Figure 3.20 Exciton-related recombinations seen in the photoluminescence spectrum of a high-purity GaAs sample grown by the hydride VPE process. FE denotes free exciton and $D^\circ - X$ and $A^\circ - X$ are neutral donor-bound and acceptor-bound excitonic transitions, respectively. $D^+ - X$ denotes an ionized donor-bound excitonic transition (from B. J. Skromme et al., *Journal of Electronic Materials*, **12** (2), 433, ©1983 IEEE).

shows the low-temperature photoluminence spectrum of high-purity GaAs grown by VPE. In addition to the free-exciton peak, a host of peaks attributed to excitons bound to impurities and electrically active defects are seen in the energy range 1.4–1.51 eV.

3.6.3.2 Band-to-Band Recombination. If the temperature of the sample is high enough so that $k_B T > \mathcal{E}_{ex}$, or if there are sufficient number of free carriers in the semiconductor producing local fields to dissociate the exciton, then most photogenerated carriers exist as separate electrons and holes in the bands. Most of these free carriers live a mean lifetime and then recombine radiatively. In direct gap semiconductors the process is complementary to the absorption process and electrons recombine with holes with momentum conservation. The energy position of the emission peak depends on the temperature and intensity of excitation. At low temperature and low excitation intensity the recombination is characterized by a single peak with the peak energy or the low energy cut-off at $\hbar\omega = \mathcal{E}_g$. As the temperature or excitation energy is increased, electrons and holes are filled at higher energies in the respective bands and these recombine to produce photons of higher energy. In the emission spectrum this is seen as a temperature or intensity-dependent tail on the high-energy side. Similarly, as the doping of the sample is increased, the whole curve may move to lower energies due to bandtailing effects. Some experimental curves obtained for GaAs are shown in Fig. 3.21. One may wonder at this point about radiative direct transitions involving the light-hole band that is at a higher (hole) energy for $\mathbf{k} \neq 0$. Since the hole mass is considerably lower in this band, the density of states and transition probability is much lower, and therefore transitions involving light holes are normally not seen in emission spectra of bulk semiconductors.

In indirect bandgap semiconductors radiative transitions from the conduction band to the valence band take place with the help of phonons to conserve momentum. Usually the process of phonon emission is most likely and the probability of phonon emission remains high even at very low temperatures.

Vizio EX1016 Page 0049

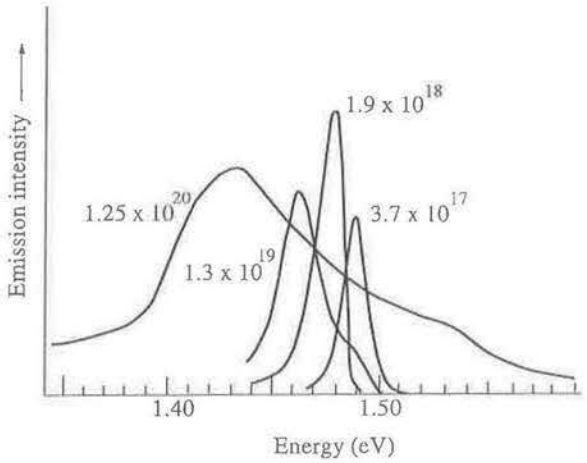


Figure 3.21 Cathodoluminescence spectra of Zn-doped GaAs at 4.2°K (from J. I. Pankove, Proceedings of the International Conference on the Physics of Semiconductors, Kyoto, 1966, *Journal of the Physical Society of Japan*, **21**, Supplement, 1966).

3.6.3.3 Donor-Acceptor and Impurity-Band Transitions. Intentional and unintentional donor and acceptor levels in semiconductors give rise to radiative transitions. In this section we will restrict the discussion to *shallow*, hydrogenic donors and acceptors, for which the energy of the emitted photons is close to the fundamental bandgap. It may be remembered that in GaAs typical donor energies are between 4 and 8 meV and typical acceptor energies are between 25 and 40 meV. As in the case of absorption, the energy of the emitted photon is given by Eq. 3.61. In general, the donor-acceptor (D-A) transition gives rise to a broad peak in the emission spectrum.

The other important near-bandgap transitions are band-to-impurity transitions, which are complementary to the absorption process. In semiconductors where the donor and acceptor binding energies are nearly equal (due to equality of effective masses), it is not easy to distinguish between the two types of impurity-band transitions (donor-band and acceptor-band). In this case the conductivity type of the material has to be known and the temperature of the sample is an important factor. For indirect impurity-related transitions a phonon emission process is involved, and in this case the emitted photon energy is given by $\hbar\omega = \mathcal{E}_g - \mathcal{E}_i - \mathcal{E}_p$ where $\mathcal{E}_i = \mathcal{E}_D$ or $\mathcal{E}_A$.

It is important to know the difference in transition probability between the D-A and impurity-band transitions, provided there are carriers available for both to occur. These transition probabilities can be determined from quantum mechanical calculations, as outlined for the case of photon absorption in Sec. 3.2. The important parameter is the carrier lifetime for the relevant process, which is crucial for the operation of injection lasers. From these calculations one gets the impurity-band carrier lifetimes of the order of several nanoseconds, while that for the band-to-band transitions varies in the range of several hundred picoseconds to one nanosecond. What it amounts to roughly is that if there are electrons in the conduction band and donor level, and there are holes in the valence band and acceptor level, the probability of the band-to-band transition is approximately four times that of the impurity-band transitions.

Donor-acceptor or impurity-band transitions can be selectively observed in the luminescence spectra by causing selective occupation of the bandedge or impurity

Vizio EX1016 Page 0050

levels. This can be done by altering the temperature of the sample or by changing the excitation intensity. Consider a GaAs sample that has both donor ($\mathcal{E}_D \simeq 5$ meV) and acceptor ($\mathcal{E}_A = 30$ meV) impurities. At very low temperatures ($\sim 4°$K) both donor and acceptor levels are occupied with electrons and holes, respectively, and the prominent peak seen in the luminescence spectrum is due to a D-A transition. If the temperature is raised slightly, to 20°K, some of the donor atoms are ionized and the electrons from these levels are raised to the conduction band. The acceptor level will still remain filled with holes. A shoulder develops to the high-energy side of the D-A peak in the luminescence spectrum, which corresponds to band-to-acceptor te (B-A) transitions. As the temperature is raised, the B-A transition becomes more prominent and the D-A transition is quenched. In fact, the energy separation between the two peaks corresponds to the donor ionization energy. The D-A and B-A peaks are schematically shown in Fig. 3.22. Selective occupation of donor levels and the conduction bandedge can also be achieved at a fixed (low) temperature by varying the excitation intensity. At low excitation (photon density) levels the D-A transition is the prominent one. As the excitation intensity is increased, the conduction band is filled with electrons and the B-A transition becomes more prominent.

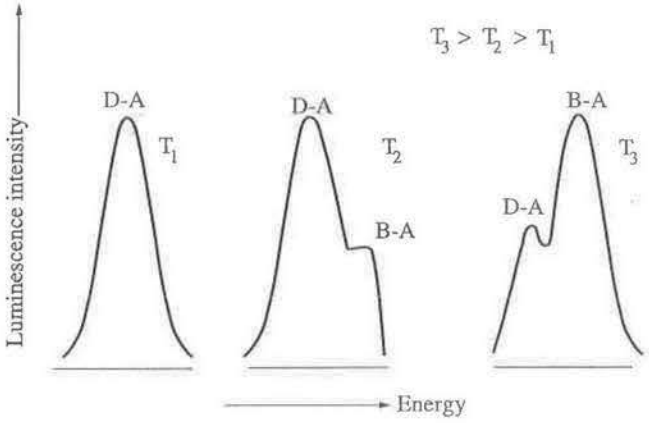


Figure 3.22 Schematic illustration of the evolution of band-acceptor and donor-acceptor transitions in the PL spectra of a semiconductor at varying temperatures.

As a concluding note, it should be mentioned that impurity-band transitions, involving the impurity level *closest* to the bandedge, can give rise to low-energy transitions. The transition energy corresponds to the impurity binding energy. Peaks believed to be due these transitions have been observed in the far-infrared region of the emission spectra of semiconductors. However, considering the transition probabilities it is debatable whether such transitions are radiative or nonradiative. In other words, instead of emitting a photon, single or multiple phonons may be emitted. In the latter case the excess energy is dissipated as heat in the lattice.

3.7 DEEP-LEVEL TRANSITIONS

Deep levels in the forbidden energy gap of semiconductors essentially act as carrier recombination or trapping centers and adversely affect device performance. Native

Vizio EX1016 Page 0051

defects in the lattice, such as vacancies, interstitial or substitutional impurities, or impurity-vacancy complexes can give rise to deep levels in semiconductors. It is generally believed that the excess energy of carriers recombining at these levels is carried away by single or multiple phonons. They were therefore also called "killer centers." Some deep levels due to impurities and defects are, however, radiative. Examples are Cu (0.18 and 0.41 eV) and Mn (0.10 eV) in GaAs, and a native defect in the same semiconductor called the EL2 level, which has an energy of ~ 0.8 eV below the conduction bandedge. Similar deep radiative transitions have been observed in $Al_xGa_{1-x}As$. It is important to realize that although a radiative transition is obtained at an energy equal to the ionization energy of the deep level, such transitions constitute a degradation of the radiative efficiency at near-bandgap energies. Also, it may be noted that the Franck-Condon shift of deep levels increases as the ionization energy of the level increases. This is because the increased localization of the carrier trapped in the potential of the impurity leads to a stronger interaction with the surrounding ions and therefore a larger lattice relaxation.

3.8 AUGER RECOMBINATION

Recombination via deep levels provides a nonradiative shunt path for the dissipation of radiative energy corresponding to the bandgap. Another important nonradiative process, which becomes important in heavily doped semiconductors, is *Auger recombination*. The Auger process is a three-carrier nonradiative recombination process, in which the excess energy released by the recombination of an electron-hole pair is transferred as kinetic energy of a third free carrier, which is raised in energy deep into the respective band. This carrier finally thermalizes back to the bottom of the band. The various possible Auger processes in semiconductors containing excess electrons and holes, are shown in Fig. 3.23. The carrier-concentration-dependent radiative recombination rate leading to spontaneous emission can be expressed as

$$R(n) = An + Bn^2 + Cn^3 \tag{3.91}$$

In this equation the first term on the right accounts for Shockley-Read-Hall recombination at defects and traps and $A = s\vartheta_{th}N_T$. The second term accounts for spontaneous

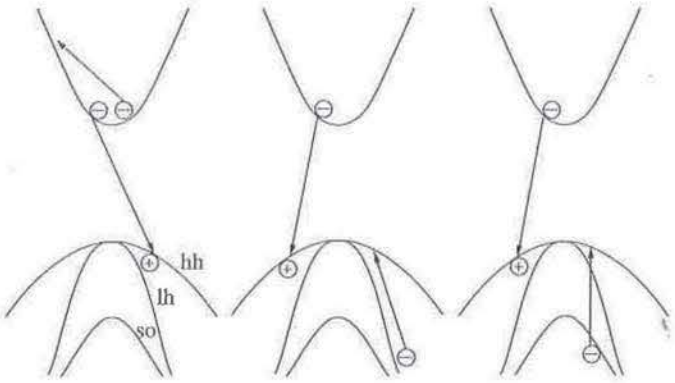


Figure 3.23 Different possible band-to-band Auger recombination processes in a direct bandgap semiconductor. Similar Auger transitions are also possible for impurity-band or D-A recombinations.

Vizio EX1016 Page 0052

radiative recombination, and the third term accounts for Auger recombination, which plays a significant role in the operation of junction lasers. It may be noted that the second and third terms on the right-hand side of Eq. 3.91 are valid for a Boltzmann distribution.

The efficiency of Auger recombination depends on the concentration of excess carriers, which is proportional to $(k_BT/\mathcal{E}_g)^{3/2}\exp(-\mathcal{E}_g/k_BT)$. Therefore, the rate of the Auger process increases with increasing temperature and with decreasing bandgap. Auger processes are more dominant in narrow bandgap semiconductors, and play an important role in limiting the performance of junction lasers made with these materials. It may also be noted that Auger recombination is, in a sense, the reverse of the impact ionization process to be discussed in Chapter 8.

3.9 LUMINESCENCE FROM QUANTUM WELLS

If excess carriers are produced in a quantum well by any of the excitation techniques, they recombine to produce a narrow emission peak. The energy position of this peak is given by

$$\hbar\omega \cong \mathcal{E}_{gw} + \mathcal{E}_{e1} + \mathcal{E}_{hh1} \tag{3.92}$$

where $\mathcal{E}_{gw}$ is the bandgap of the well material and $\mathcal{E}_{e1}$ and $\mathcal{E}_{hh1}$ are the electron and hole subband energies, respectively. The density of states of the heavy-hole band is much larger than that of the light-hole band, and therefore the heavy holes are primarily involved in the recombination process. Remember that we are considering nonresonant excitation, in which an excess electron-hole population is generally created in the barriers, from which they quickly diffuse and thermalize to the lowest bound states in the well region. If resonant excitation is used, then radiative transitions corresponding to the light-hole bands and higher-order bands can be seen in the luminescence spectrum.

For very large well sizes ($\geq$ 200 Å), the luminescence spectrum begins to resemble that of the bulk well material. As the well size is reduced, quantization effects become important. At $L_z = 100$ Å for GaAs/AlGaAs quantum wells, for example, most of the wavefunctions corresponding to the bound states are still confined to the well. For smaller well sizes, a substantial part of the wavefunctions is in the barriers. Therefore, for well sizes of 100 Å or less, the luminescent properties of the quantum well are determined by the quality of the well material, the perfection of the interface, and the quality of the barrier material. In fact, the width of the luminescence spectrum can be analyzed to understand the optical quality of the quantum well.

3.10 MEASUREMENT OF ABSORPTION AND LUMINESCENCE SPECTRA

The measurement of absorption coefficient of a semiconductor is schematically shown in Fig. 3.24. Light from a variable wavelength light source, or monochromator, is incident normally on the sample. The light is partly absorbed and the intensity of the transmitted light is measured by a photodetector. The absorption coefficient $\alpha(\hbar\omega)$ is

Vizio EX1016 Page 0053

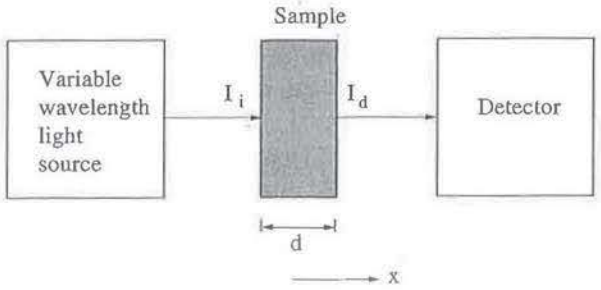


Figure 3.24 Schematic for optical absorption (transmission) experiment.

related to the relative decrease of light intensity, $\Im$, as it passes through the absorption medium, according to

$$\alpha = -\frac{1}{\Im}\frac{d\Im}{dx} \tag{3.93}$$

or

$$-\frac{d\Im(x)}{dx} = \alpha\Im(x) \tag{3.94}$$

The solution to this equation is

$$\Im(x) = \Im_i e^{-\alpha x} \tag{3.95}$$

where $\Im_i$ is the intensity of the incident light. If the sample thickness is d, then the intensity of the transmitted light falling on the detector is

$$\Im_d = \Im_i e^{-\alpha d} \tag{3.96}$$

The value of α will depend on the photon energy or wavelength. For a semiconductor of bandgap $\mathcal{E}_g$, photons of energy $\hbar\omega < \mathcal{E}_g$ pass unattenuated through the sample and $\alpha \cong 0$. As shallow impurity levels and the bandgap energy are approached, α begins to increase and rises abruptly at bandgap energy. The threshold of the absorption profile gives the value of $\mathcal{E}_g$.

Photoluminescence is the optical radiation emitted by a material (in addition to the thermal equilibrium blackbody radiation) resulting from its nonequilibrium state caused by an external light excitation. Three distinct processes take place to result in the light emission from the system: (1) absorption of exciting light and thus creation of electron-hole pairs, (2) partial radiative recombination of these electron-hole pairs, and (3) escape of this radiation from the system. Since the electron-hole pairs are generated by the absorbed exciting light, their highest concentration is near the illuminated surface of the sample; the resulting carrier distribution is both inhomogeneous and nonequilibrium. To regain homogeneity and equilibrium, the excess carriers will diffuse away from the surface while recombining radiatively and nonradiatively. The emitted radiation is subject to self-absorption; thus, it will not propagate far from the emission region. Most of the processes are therefore occurring in a region within an absorption length of the illuminated surface, and the emitted light escapes mostly through the nearby illuminated surface.

Vizio EX1016 Page 0054

A typical photoluminescence measurement system consists of an excitation source, a variable-temperature cryostat and sample holder assembly, a high-resolution scanning spectrometer, and a detection system; a schematic diagram is shown in Fig. 3.25. The excitation source for most III–V semiconductors can be an argon-ion laser capable of delivering 0–15 mW of power at 5145 Å. A narrow band interference filter is used between the laser and the sample because of the rather intense radiation from the argon plasma. The laser beam is typically focused to a 100–500 μm diameter spot on the sample surface. By varying the laser power and using neutral density filters, it is possible to vary the power density on the sample in the range $10–10^3$ W/cm^2. In order to get the maximum amount of information from the radiative relaxation process, it is necessary to cool the sample to cryogenic temperatures. This is accomplished by mounting the sample with cryogenic thermally conductive grease on a gold-coated copper cold finger of a liquid He cryostat equipped with optical windows. A Lambertian distribution of the luminescence from the sample is collected by a suitable lens and is focused on the slit of a high-resolution scanning spectrometer with a resolution

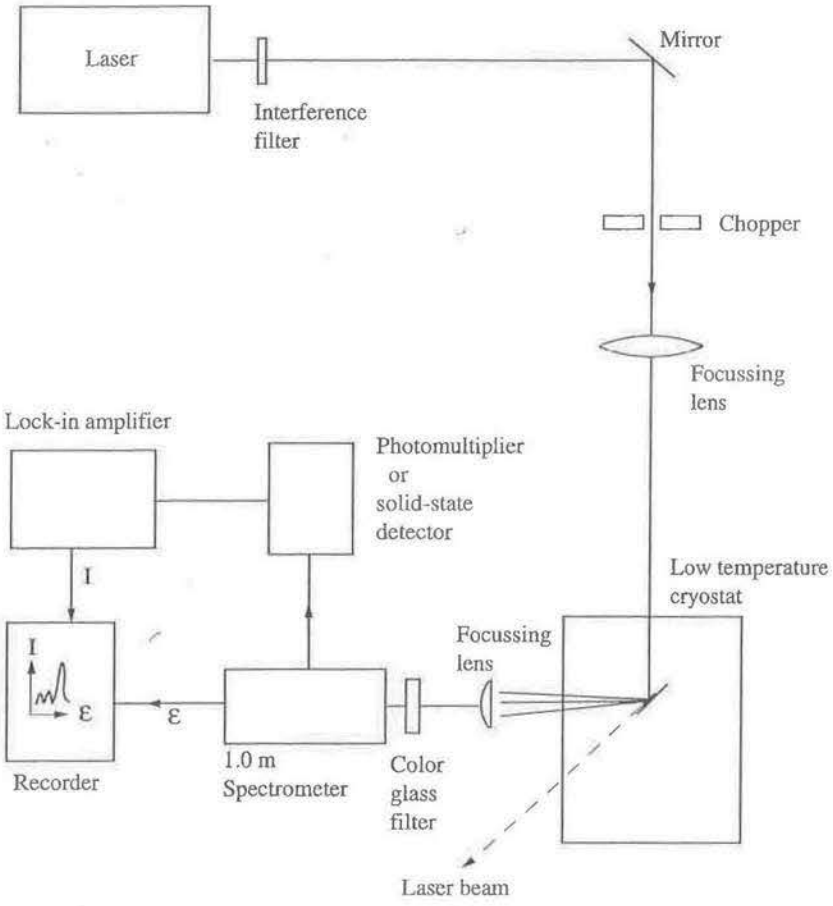


Figure 3.25 Photoluminescence measurement system.

Vizio EX1016 Page 0055

$\sim$ 1–2 Å. The laser beam is incident at an angle on the sample so that the reflected beam does not go into the collecting lens. For GaAs and GaAs-based materials such as AlGaAs a liquid nitrogen-cooled photomultiplier tube is used to detect the luminescence signal. For longer wavelength (smaller bandgap) materials, Ge and CdS detectors are generally used.

3.11 TIME-RESOLVED PHOTOLUMINESCENCE

With intrinsic photoexcitation electron-hole pairs are generally created at energies much higher than the bandgap and the luminescence arises in the semiconductor after the following processes have occurred: (1) a rapid loss (in approximately 5–20 ps) of energy of the excess carriers by emitting phonons, (2) thermalization of the carriers, and (3) radiative and nonradiative recombination of the carriers. The recombination process is characterized by the rate equation

$$\frac{dn}{dt} = G - \frac{n}{\tau} \tag{3.97}$$

where n is the total number of excess carriers in the well and G is the generation rate. The time constant τ is related to the radiative and nonradiative time constants τ_r and τ_{nr} by Eq. 3.10. Recombination times in bulk semiconductors can vary between 500 ps and a few nanoseconds. In quantum wells, a faster recombination of the electron-hole pairs is observed, compared to bulk semiconductors, due to the enhanced overlap of electron and hole wavefunctions. Typical recombination times, depending on well width, are in the range 300–800 ps. The time resolved photoluminescence technique is used to measure these times. Tunneling of carriers out of a quantum well through the barrier will affect these recombination times. The tunneling rates will depend on transverse electric fields, due to the change the latter causes to the shape of the barrier, and these tunneling rates can be very high. Tunneling times as low as 100 ps is observed for barriers $\sim$ 50–100 Å. Time-resolved photoluminescence spectroscopy is useful to investigate all such dynamical processes in the temporal range of 10 ps up to a few nanoseconds.

A typical time-resolved luminescence measurement system is similar to that for steady-state luminescence, except in the excitation source and detection system. Excitation is provided by 1–5 picosecond pulses from a tunable dye laser, Ti: sapphire laser, or tunable color center laser depending on the material under test. These lasers are usually pumped by a mode-locked Nd:YAG laser, which has a repetition rate of 76 MHz. Typical output power characteristics of this laser are 10 W average power, 100 ps pulse width. The output is frequency doubled in a KDP crystal, giving an average power of 0.9–1.0 W. The dye laser cavity is folded into a Z configuration, with cavity length equal to that of the Nd:YAG pump laser, so that its repetition rate is also 76 MHz. The wavelength is tuned with a birefringent filter, which consists of three flat and parallel crystalline quartz plates placed inside the dye laser cavity at Brewster angle. The birefringent filter has low loss for linearly polarized light at

Vizio EX1016 Page 0056

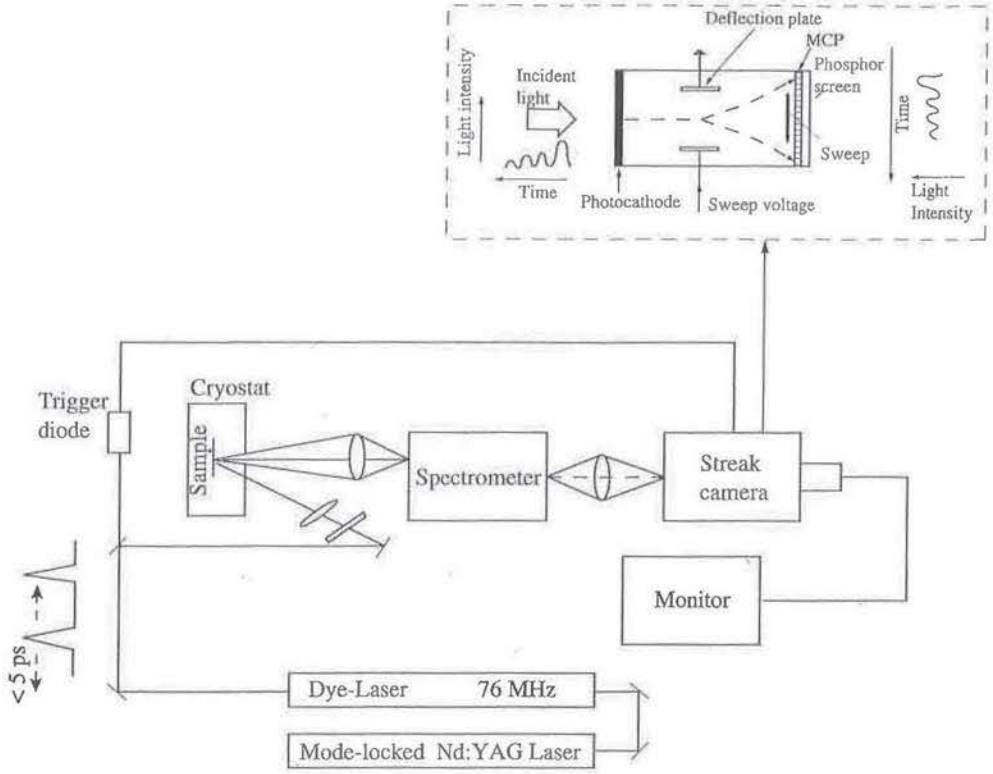


Figure 3.26 Arrangement for picosecond luminescence measurements with a synchronously pumped dye laser and a synchroscan streak camera.

a particular wavelength. The laser output wavelength is tuned. The output power of the dye laser or color center laser varies in the range 10–100 nW.

The luminescence from the sample is collected by a suitable lens and is focused on the slit of a high-resolution auto-scan spectrometer. The analyzed spectrum is detected by a streak camera with a typical temporal resolution of 10 ps. Figure 3.26 shows the operating principle of the streak tube, which forms the heart of the streak camera. The light pulse to be measured is projected on to the slit of the streak camera. The slit image of the incident light is focused on the photocathode of the streak tube via a relay lens where the photons are converted into electrons. The electrons are accelerated by the strong electrostatic field between the photocathode and the mesh-electrode, and conducted into the deflection field beyond which they are swept at high speed in a direction perpendicular to the slit length by applying a deflection voltage synchronized with the arrival of the electrons to the deflection field. The electrons are then multiplied in a microchannel plate (MCP) by a factor of approximately 3×10^3. Electrons exiting the MCP then bombard the phosphor screen of the steak tube and are converted to the optical image (called "streak image"). As a result of this structure and the sweeping system used, the time at which electrons were released from the photocathode surface can be determined by their

Vizio EX1016 Page 0057

deflected angle (vertical position on the phosphor screen). Therefore, the time axis of the incident light corresponds to the vertical axis on the phosphor screen, and the intensity of the incident light can be determined by the density of the streak image. Since it is necessary that the timing of the high-speed deflection is synchronized to the arrival time of electrons at the deflection field, part of the incident light is usually focused onto a PIN photodiode to generate a trigger signal for the sweeping. Typical time-resolved luminescence data obtained from measurements on GaAs-AlGaAs multiquantum wells are shown in Fig. 3.27. The excitation energy is above the conduction bandedge of the AlGaAs barrier. The excess carrier therefore quickly reach the well. This contributes to the observed rise time. The decay time reflects radiative and nonradiative recombination in the well and some tunneling into the barriers. The different dynamical processes for electrons are illustrated in Fig. 3.28.

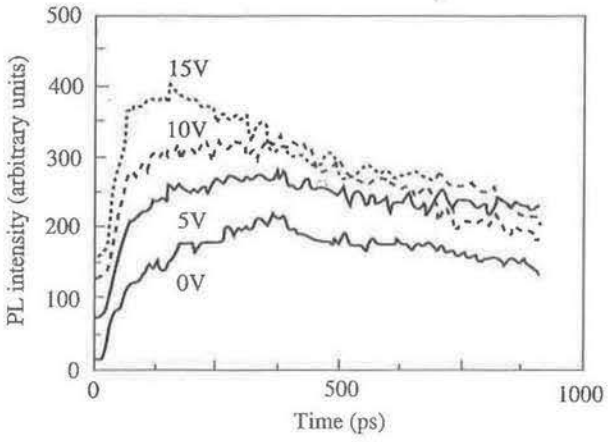


Figure 3.27 Temporal profile of the low temperature luminescence (T = 10°K) intensity from a 100 Å GaAs quantum well with 50 Å $Al_{0.3}Ga_{0.7}$ As barrier at different applied biases due to photoexcitation with 10 ps optical pulses of energy higher than the AlGaAs bandgap. The decay times are proportional to carrier tunnel rates from the quantum wells (N. Debbar, *Investigation of Optical Properties of Quantum Wells for Optoelectronic Device Applications*, Ph.D. Thesis, University of Michigan, 1989).

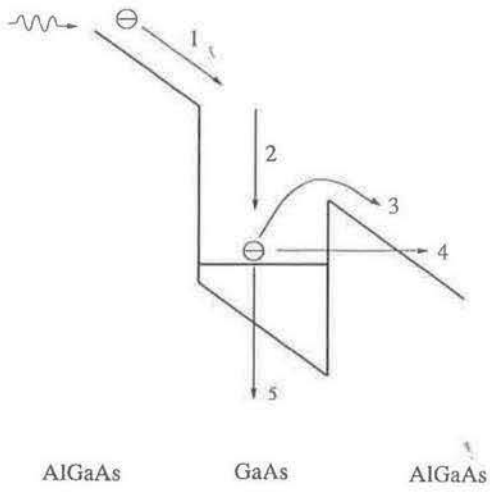


Figure 3.28 Carrier dynamics in a quantum well upon photoexcitation at an energy above the AlGaAs bandgap. The different processes are (1) drift and diffusion, (2) thermalization in the well by phonon scattering, (3) thermionic emission, (4) tunneling, and (5) radiative and nonradiative recombination. Photoexcited holes undergo the same processes.

Vizio EX1016 Page 0058

PROBLEMS

3.1 An n-type sample of GaAs has 10^{14} cm^{-3} electrons in the conduction band in equilibrium. Intrinsic photoexcitation is used to create 10^{20} cm^{-3}.s^{-1} electron-hole pairs in the sample and the radiative lifetime is 1 ns. Assuming that there is no nonradiative recombination in the sample, find the energy position of the hole quasi-Fermi level in the energy bandgap under photoexcitation conditions. Comment on the position of the electron quasi-Fermi level. (Remember that under steady state conditions, Eq. 3.9 is valid.)

3.2 Absorption data taken from the literature for Ge at 300°K and 77°K are given in Fig. P3.2. Analyze the 300°K data to obtain the value of the direct bandgap, the indirect bandgap, and the phonon energy participating in the indirect transitions. Which category of phonons participate in these transitions?

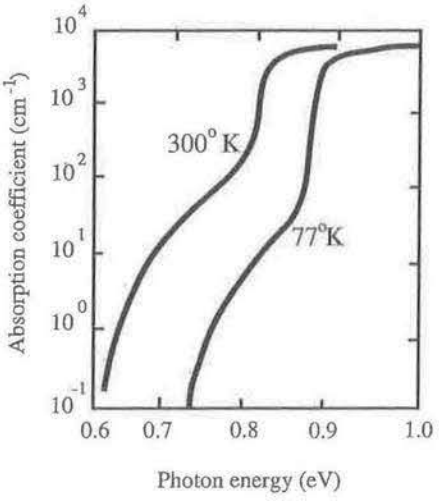


Figure P3.2

3.3 Alternate multiple layers of $n(5 \times 10^{18}$ cm$^{-3})$ and $p(5 \times 10^{18}$ cm$^{-3})$-type GaAs, each 500 Å thick, are grown by molecular beam epitaxy. What does the band diagram of this composite material look like? What are its special properties? What will the absorption and luminescence spectra of the modulated structure look like? Comment on its radiative lifetime.

3.4 Photons with an energy $\hbar\omega$ slightly less than the bandgap $\mathcal{E}_g$ are incident on a semiconductor material. Can an absorption edge spectrum still be obtained? If so, mention the mechanisms by which this might be possible in the case of a direct bandgap semiconductor and an indirect bandgap semiconductor.

3.5 Intrinsic light is incident at one end of a 10-cm-long semiconductor sample. If 15% of the light incident on the sample is absorbed per centimeter,

Vizio EX1016 Page 0059

calculate the fraction of the incident light that is transmitted through the sample and the absorption coefficient of the material.

3.6 Discuss, with figures, the process of electroabsorption in bulk semiconductors, multiquantum wells, and superlattices. Remember that in the last case the quantum wells are coupled (thin barriers).

3.7 Derive Eq. 3.68.

3.8 From the data of Fig. 3.11, calculate the bandedge emission spectrum for GaAs at 300°K.

3.9 As we shall see in Chapter 11, the refractive index n_r is uniquely defined for transverse electric (TE) and transverse magnetic (TM) mode propagations. In fact, $n_r^{TE} - n_r^{TM} = \Delta n_r$ is directly proportional to the electro-optic coefficients of the crystal. By approximating the heavy-hole and light-hole excitonic resonances in the absorption spectrum of a semiconductor as δ-functions, show that Δn_r is directly related to the energy $(\mathcal{E}_{lh} - \mathcal{E}_{hh})$.

3.10 Discuss and illustrate the changes in the luminescence spectrum of n-type GaAs ($N_D = 10^{15}$ cm^{-3}) at 4.2°K when

(a) the donor density is gradually increased

(b) when the compensation in the sample is gradually increased

3.11 Calculate the value of α for direct allowed transitions in InP at 2°K. Does this value agree with the data of Fig. 3.11 after taking the temperature dependence of bandgap in account?

3.12 Derive the relation between the real and imaginary parts of the complex refractive index and the conductivity σ.
[Hint: You need to start with Maxwell's equations.]

3.13 Describe, with diagrams, the following processes:

(a) Auger recombination

(b) Shockley-Read-Hall recombination

(c) radiative recombination

How would you minimize the rates of each of these processes?

3.14 Derive Eqs. 3.80 and 3.81.

READING LIST

AGRAWAL, G. P., and DUTTA, N. K. *Long Wavelength Semiconductor Lasers.* Van Nostrand Reinhold, New York, 1986.

KRESSEL, H., and BUTLER, J. K. *Semiconductor Lasers and Heterojunction LEDs.* Academic Press, New York, 1977.

Vizio EX1016 Page 0060

PANKOVE, J. I. *Optical Processes in Semiconductors.* Prentice Hall, Englewood Cliffs, NJ, 1971.

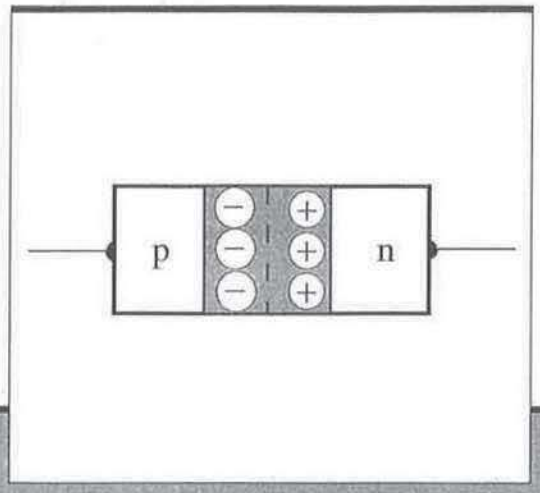
SINGH, J. *Physics of Semiconductors and Their Heterostructures.* McGraw-Hill, New York, 1993.

WANG, S. *Fundamentals of Semiconductor Theory and Device Physics.* Prentice Hall, Englewood Cliffs, NJ, 1989.

Vizio EX1016 Page 0061

4

Junction Theory



Chapter Contents

Vizio EX1016 Page 0063

4.1 INTRODUCTION

At the present time, p-n junctions have wide applications in the design and realization of a host of electronic devices. This is because these junction diodes can, like vacuum tubes, perform functions such as rectification and switching. When incorporated in a bipolar transistor, amplification can be obtained. It will be seen in the following chapters that p-n junctions are also crucial for the operation of most optoelectronic devices. For example, in a photodiode, photogenerated carriers are separated and collected by a reverse bias. In a solar cell, the photogenerated carriers create a short-circuit current or an open-circuit voltage. In lasers and light-emitting diodes a forward bias is used to inject and create a nonequilibrium population of minority carriers, which then recombine radiatively with majority carriers. This chapter will therefore outline the basic principles of p-n junctions in their simplest form as anisotype homojunctions. The latter usually signifies a junction formed by the same semiconductor doped n- and p-type in different regions. Two other kinds of junctions have assumed roles of great importance in present-day devices. These are metal-semiconductor junctions and heterojunctions. A heterojunction is formed with two dissimilar semiconductors (unequal bandgaps), which can be either of the same doping type (isotype) or of the opposite doping type (anisotype). The basic principles of these junction will also be outlined in this chapter.

4.2 P-N JUNCTIONS

The simplest semiconductor junction diode is formed by "bringing together" a p-type and an n-type semiconductor. The subject of p-n junctions has been extensively treated and widely discussed in a number of texts. Therefore, only the aspects relevant to optoelectronic devices will be described here. These are junction formation, electrostatic properties, and current voltage characteristics. Also of importance are practical graded junctions and the case of high-level injection.

4.2.1 Junction Formation

The three common techniques of junction formation at the present time are epitaxial growth, diffusion and ion implantation. The principles and techniques of epitaxial growth have been outlined in Chapter 1. As an example, to form a GaAs p-n junction by liquid phase epitaxy (LPE), an appropriately doped substrate is moved successively over melts that have dissolved dopant species, usually in their elemental form. Typical n- and p-type dopants for GaAs and AlGaAs are Te and Zn. To have n-type GaAs the Ga-rich growth melt is nearly saturated with As and has an appropriate amount of Te, whose incorporation may be described by

$$Te(l) + V_{As} \rightleftharpoons Te(s)^{+} + e^{-} \qquad (4.1)$$

In this reaction Te reacts with an As vacancy in the liquid phase to give an ionized substitutional donor on an As site, $Te(s)^{+}$, and a free electron. The amount of Te to be dissolved in the melt to obtain the required doping density is determined by the solid solubility isotherm, which is independently determined by measuring dopant concen-

Vizio EX1016 Page 0064

trations in the liquid phase (at a constant temperature) and in the solid crystal. The distribution coefficients of Te in GaAs as a function of temperature are shown in Fig. 4.1.

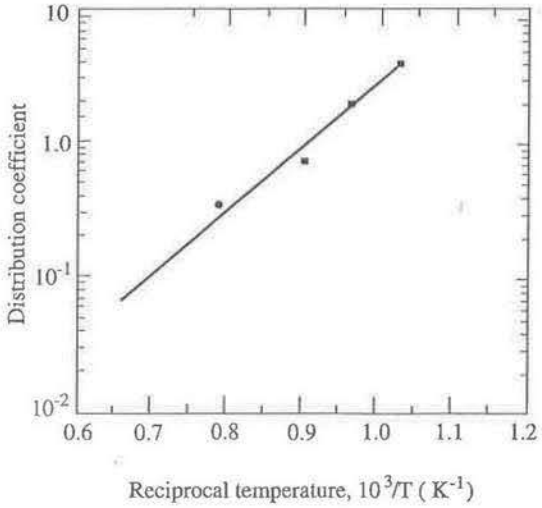


Figure 4.1 The distribution coefficient of Te in GaAs as a function of temperature (from H. C. Casey and M. B. Panish, *Heterostructure Lasers, Part B: Materials and Operating Characteristics*, copyright © 1978. Reprinted by permission of Academic Press, New York). The data points are adapted from H. C. Casey, Jr. et al., *J. Phys. Chem. Solids*, Pergamon Press, **32**, 571–580, 1971 and C. S. Kang and P. E. Greene, *Gallium Arsenide: 1968 Symp. Proc.*, 18, Inst. of Physics and Phys. Soc., London, 1969.

The incorporation of dopants during molecular beam epitaxy (MBE) is achieved by co-evaporating the dopant species, usually in elemental form, with the main flux species. In the case of GaAs, for example, a p-n junction is realized from effusing fluxes of Ga, As, Si(or Sn) (n-type), and Be (p-type). The appropriate dopant cells can be shuttered on or off to form the junction. The effusion cell temperature and dopant flux are decided by the equilibrium vapor pressure curve of the dopant species. Figure 4.2 shows the measured p-type concentration due to Be doping in GaAs as a

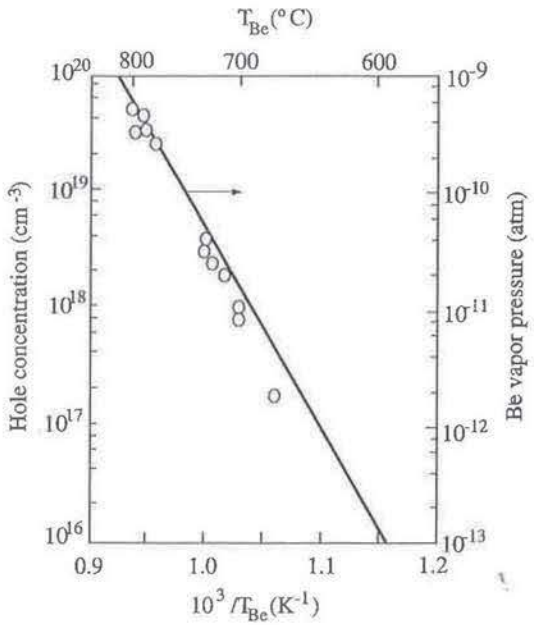


Figure 4.2 Hole concentration versus Be oven temperature for GaAs layers grown by MBE. The solid line represents the Be equilibrium vapor pressure curve (from M. Ilegems, *Journal of Applied Physics*, **48**, 1278, 1977).

Vizio EX1016 Page 0065

function of Be cell temperature. The data agree fairly well with the Be vapor pressure curve, also shown in the same figure, indicating a unity incorporation coefficient. Most dopants exhibit similar behavior. Similarly, in organo-metallic vapor phase epitaxy (OMVPE), the dopant source, usually in the form of a liquid, is vaporized and introduced to the reaction mixture by hydrogen carrier gas. n- and p-type materials are obtained by switching the dopant flows on and off by mass flow controllers.

A mention should be made regarding the abruptness of junctions obtained by the three epitaxial techniques. In liquid phase epitaxy there is an inherent formation of a compensated region around the metallurgical interface due to melt dissolution, creating a doping profile shown in Fig. 4.3(a). Very abrupt junctions, as shown in Fig. 4.3(b), can be obtained by MBE, since the fluxes are terminated by mechanical shutters. Similarly, abrupt junctions can be obtained by OMVPE with mass flow controllers and concurrent pumping to avoid memory effects in the gas lines.

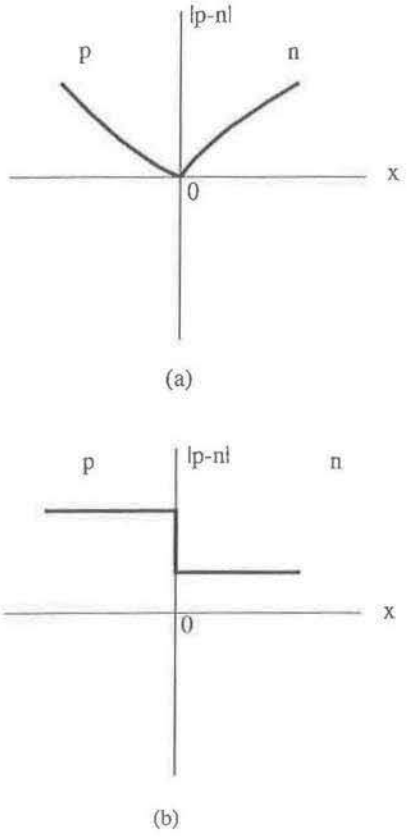


Figure 4.3 (a) Nonabrupt doping profile created near the junction by melt dissolution and (b) abrupt doping profile created by MBE or OMVPE.

Diffusion of impurities is a common technique of doping in semiconductors, more so in Si than in III–V semiconductors. Compound semiconductor solar cells and the heavily doped contact regions of lasers are commonly formed by the diffusion process. Diffusion is a thermally activated random motion process, characterized by a diffusion coefficient, D. The values of D for common impurities in GaAs are shown in Fig. 4.4. The process of diffusion doping involves the substitution of a host lattice atom by

Vizio EX1016 Page 0066

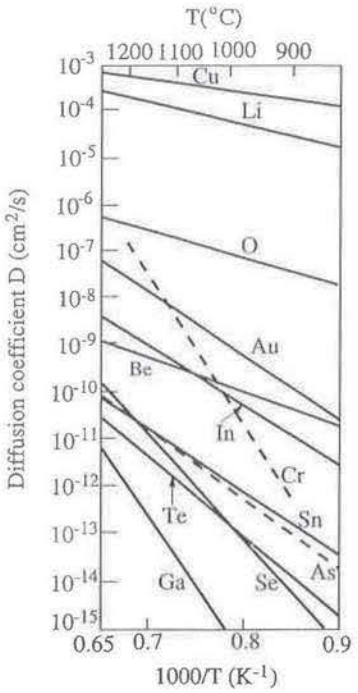


Figure 4.4 Diffusion constant of several impurity species in GaAs (from M. Zambuto, *Semiconductor Devices*, McGraw-Hill, 1989). The data points are adapted from D. L. Kendall, *Diffusion, Semiconductors and Semimetals*, **4**, Academic Press, New York, 163, 1968.

the impurity atom. In other words, the lattice atom has to move from the lattice site to make room for the diffusing impurity atom, which is until then in a nonlattice or *interstitial* site. This substitution typically involves an energy barrier of 1 eV and therefore temperatures around 1000°C are necessary to complete the diffusion process in a time of 1 hour or so.

There are two types of diffusion processes, determined by the amount of source (impurity) atoms. In essence there are correspondingly two types of boundary conditions and two types of solutions to the diffusion equation (Fick's second law):

$$\frac{\partial N}{\partial t} = D\frac{\partial^2 N}{\partial x^2} \tag{4.2}$$

where N is the concentration of the impurity. In the first, *limited source diffusion*, a fixed quantity of the impurity species is deposited in a thin layer on the wafer surface. This is modeled as an impulse function with the magnitude equal to the dose Q. The solution is given by the Gaussian distribution:

$$N(x,t) = (Q/\sqrt{\pi Dt}) \exp\left[-(x/2\sqrt{Dt})^2\right] \tag{4.3}$$

which is illustrated in Fig. 4.5(a). With increasing time, at any fixed temperature, the diffusion front moves into the wafer with a corresponding decrease of the surface concentration so that the area under the curves (dependent on the dose) remains constant. The second type is *constant source diffusion* or *error function diffusion*, where the impurity concentration is constant over time. In other words, an unlimited supply of dopant species is provided. Under this boundary condition the solution to

Vizio EX1016 Page 0067

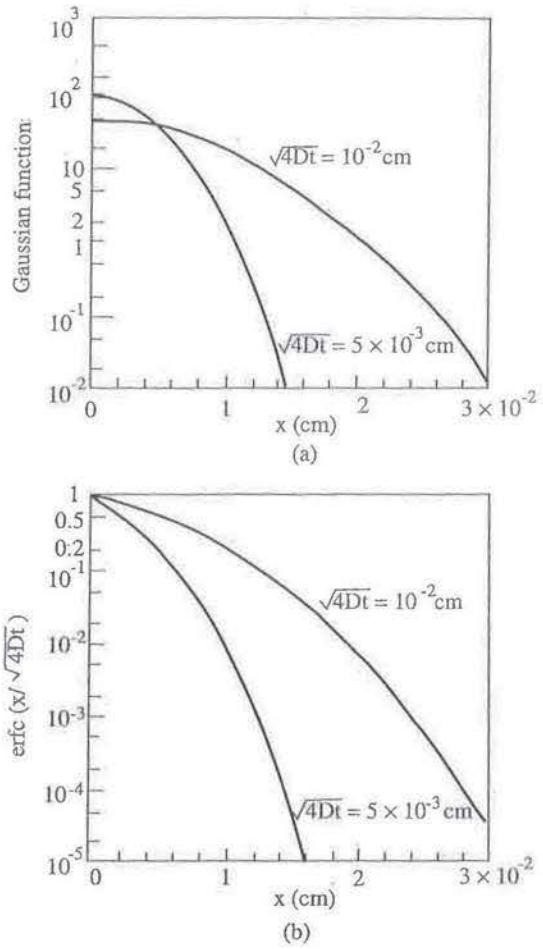


Figure 4.5 (a) Gaussian distribution and (b) complementary error function distribution for two different values of $\sqrt{4Dt}$ (from S. Wang, *Fundamentals of Semiconductor Theory and Device Physics*, Prentice Hall, Englewood Cliffs, NJ, 1989).

the diffusion equation is given by

$$N(x, t) = N_o \text{ erfc } (x/2\sqrt{\text{Dt}}) \tag{4.4}$$

where N_o is the impurity concentration at $x = 0$ (wafer surface). This solution is illustrated in Fig. 4.5 (b), where the impurity profile moves deeper into the wafer with increasing time, increasing temperature, or a combination of both. As an illustration, the impurity concentration profile, formed by either type of diffusion, and the net concentration profile of a p-n junction formed by diffusing acceptor impurities into an n-type wafer are shown in Fig. 4.6. At any given temperature, the upper limit to the amount of impurity that can be absorbed by the wafer is called the *solid solubility limit* for that species. Also, it should be remembered that at these high doping concentrations, only a fraction of the dopant species is electrically active (i.e., they are in substitutional lattice sites).

The third common technique of junction formation is *ion implantation* in which high-energy ions are projected onto the wafer surface. This technique of doping is more commonly used for the fabrication of electron devices such as metal-semicon-

Vizio EX1016 Page 0068

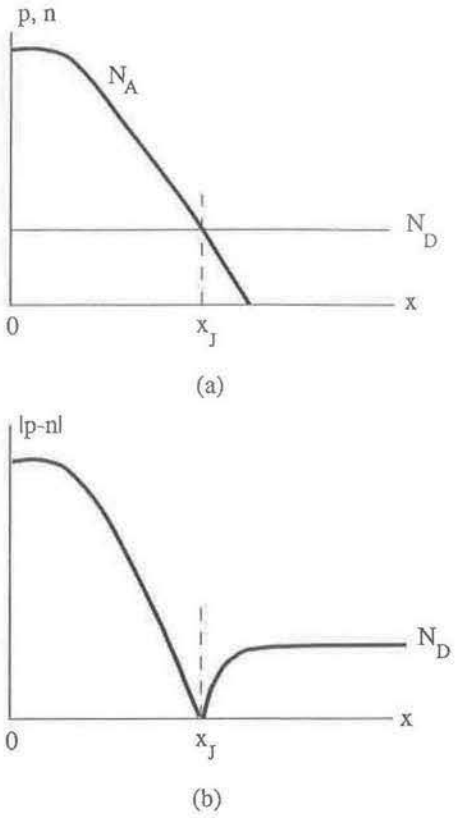


Figure 4.6 (a) Diffusion profile of acceptor species into an n-type semiconductor and (b) the corresponding free-carrier profile of the p-n junction. The point x_J corresponds to the metallurgical junction.

ductor field-effect transistors (MESFETs). It is also useful for dopant species that have very low vapor pressure and therefore cannot be effused during MBE. The ions are directed in a *nonchanneling* direction to ensure that their collision with the host lattice and their energy determine the average penetration depth, also called the *projected range*. A typical ion-implanter consists of a source chamber, accelerator, mass separator, and target chamber. The implant profile obeys a Gaussian law. Typical implant profiles for the case of single and multiple implants are shown schematically in Fig. 4.7. The latter is used for obtaining large junction depths with a uniform doping profile. The damage caused by the high-energy ions during implantation makes the

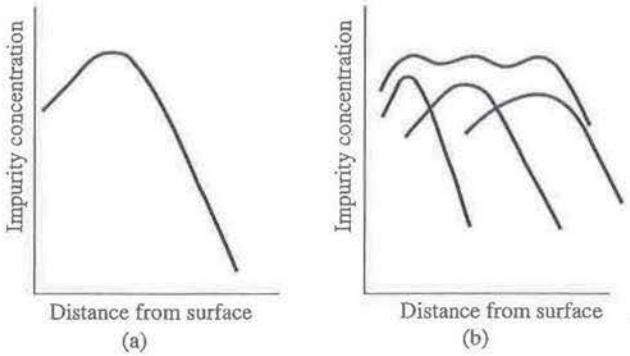


Figure 4.7 Schematic representation of implantation profile of dopant species in a semiconductor for (a) single implant and (b) multiple implants at successively higher energies. The typical range of energies in GaAs is 50–200 keV.

Vizio EX1016 Page 0069

material almost amorphous. To restore single crystal behavior, the implanted wafer is usually annealed, typically at temperatures ranging from 700° to 1000°C for periods ranging from 1/2 to 1 hour. Presently a technique of *flash lamp annealing* is also used in which the wafer is annealed at 1000°–1100°C for a few seconds in an inert atmosphere with or without encapsulation. A last point to note is that for most device fabrication, diffusion or implantation is done through on oxide mask to delineate the device area.

4.2.2 Electrostatics of the p-n junction: Contact Potential and Space Charge

A p-n junction is schematically shown in Fig. 4.8, which also depicts the equilibrium impurity concentrations and carrier densities in the n- and p-type semiconductors. The p- and n-type semiconductors have large concentrations of free holes and free electrons, respectively. Once they are joined, there is a large concentration gradient of holes from the p- to the n-side across the metallurgical junction and vice versa for electrons. Diffusion of free carriers takes place by virtue of this gradient. There are, however, two important points to take note of. First, we are dealing with the diffusion of charged particles, and second, electrons leaving the n-region leave behind positively ionized donor atoms and similarly holes leaving the p-region leave behind negatively ionized acceptor atoms. There is therefore the formation of two regions, consisting of *immobile* charge of opposite types, on the two sides of the metallurgical junction. As carrier diffusion proceeds, the extent (length) of these regions increases. The two

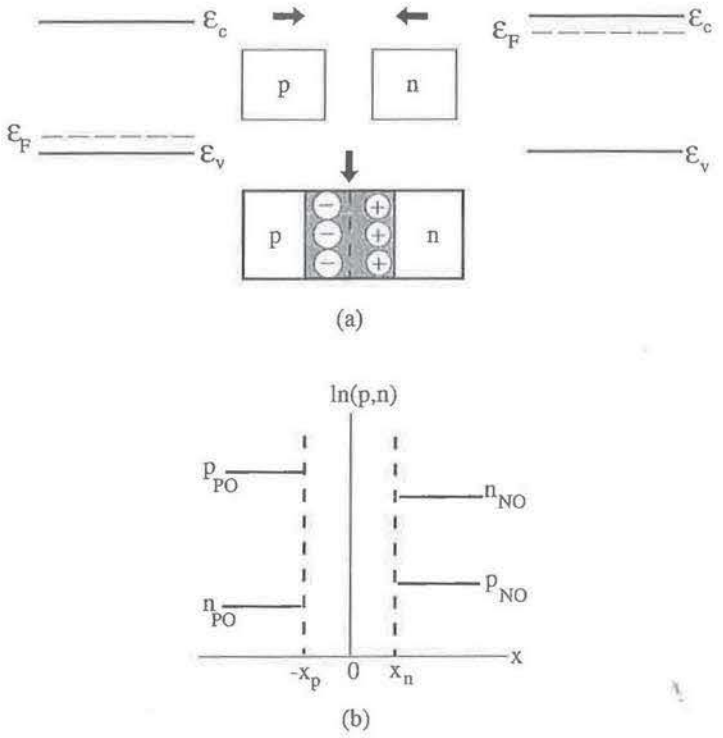


Figure 4.8 Electrostatics of the p-n junction: (a) formation of depletion region, (b) majority and minority carrier densities in neutral region, (c) space charge density, (d) electric field profile, and (e) potential profile.

Vizio EX1016 Page 0070

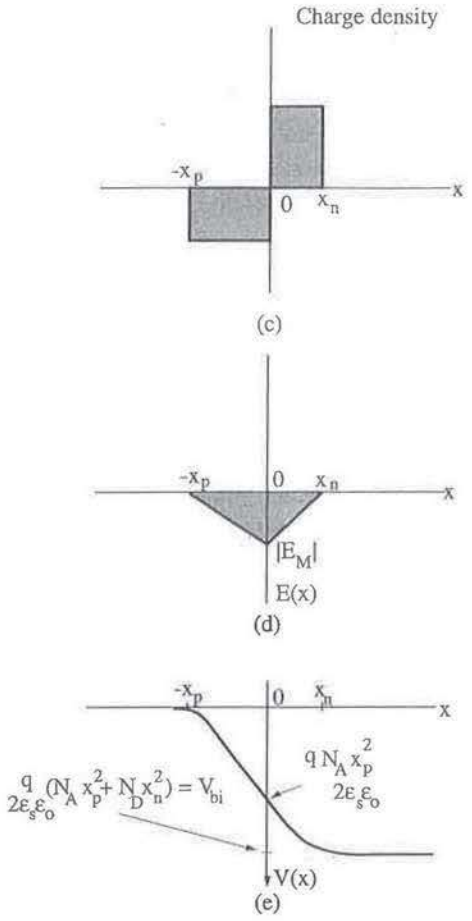


Figure 4.8 (*continued*)

regions of immobile charges result in an electric field, which is directed from the positive charge toward the negative charge. As this field builds up, it opposes further diffusion of carriers across the junction. In essence, the electric field generates drift currents of electrons and holes that exactly match the corresponding diffusion currents in equilibrium. This is consistent with the fact that under equilibrium conditions, with no applied bias, there is no *net* current flow. In equilibrium, the ionized region consisting of positive and negative charges on opposite sides is called the *depletion, space charge,* or *transition* region. The total depletion layer width $W = x_n + x_p$, where x_n and x_p are the space charge layer thicknesses in the two regions. The extent of the depletion region on each side is inversely dependent on the doping on that side. As a typical example, in GaAs doped with $n = 10^{15}$ cm^{-3}, the space charge region is approximately 0.4 μm thick.

The built-in electric field is a maximum at the junction and goes to zero at the edges of the depletion region on both sides. By virtue of this electric field E a potential gradient develops, according to

$$V_{bi} = -\int E dx \tag{4.5}$$

Vizio EX1016 Page 0071

The bands bend in the depletion region by an energy qV_{bi}, and this is called the contact or *built-in potential.* Looking at the picture simply, majority carriers diffusing due to contact formation are finally prevented from doing so because they cannot surmount this energy barrier. It is important to remember that minority carriers can easily go down the potential hills to the opposite sides, but these two components, due to electrons from the p-side and holes from the n-side, must exactly balance the majority carrier diffusion currents under equilibrium conditions. We are initially assuming an *abrupt* junction, in which a step change in impurity type occurs at the junction. Referring to Fig. 4.8, and the Gauss law, it follows that

$$qN_A x_p = qN_D x_n = \epsilon_s \epsilon_o E \tag{4.6}$$

where ϵ_s is the static dielectric constant of the semiconductor and N_A and N_D are the doping densities on the p- and n-sides respectively. Also, full ionization of the dopants are assumed for this analysis. Equation 4.6 signifies that the total ionized negative charge per unit area on the p-side must be equal to the total ionized positive charge per unit area on the n-side, and this is known as the *depletion approximation.* From Poisson's equation for the abrupt junction,

$$-\frac{d^2V}{dx^2} = \frac{dE}{dx} = \frac{q}{\epsilon_s \epsilon_o}\left[p(x) - n(x) + N_D^+(x) - N_A^-(x)\right] \tag{4.7}$$

Therefore,

$$-\frac{d^2V}{dx^2} \cong \frac{q}{\epsilon_s \epsilon_o} N_D \, , 0 < x \leq x_n \tag{4.8}$$

and

$$-\frac{d^2V}{dx^2} \cong \frac{q}{\epsilon_s \epsilon_o} N_A \, , -x_p \leq x < 0 \tag{4.9}$$

Upon integration, the electric field profile is given by

$$E(x) = \frac{qN_D}{\epsilon_s \epsilon_o}(x - x_n) \qquad \text{(n-side)}$$

$$= -\frac{qN_A}{\epsilon_s \epsilon_o}(x + x_p) \qquad \text{(p-side)} \tag{4.10}$$

Integrating Eqs. 4.8 and 4.9 twice gives

$$V(x) = E_M\left(x - \frac{x^2}{2W}\right) \tag{4.11}$$

where E_M is the maximum field at the junction. From Eq. 4.11, the contact potential distribution is obtained for different values of x, as indicated in Fig. 4.8, as

$$V(x) = -\frac{qN_A}{2\epsilon_s \epsilon_o}(x_p + x)^2 \qquad \text{(p-side)}$$

$$= -\frac{qN_D}{2\epsilon_s \epsilon_o}(x - x_n)^2 \qquad \text{(n-side)} \tag{4.12}$$

Vizio EX1016 Page 0072

and it follows that

$$V_{bi} = \frac{q}{2\epsilon_s \epsilon_o}\left[N_A x_p^2 + N_D x_n^2\right] \tag{4.13}$$

At the metallurgical junction,

$$V(x) = -\frac{qN_A\, x_p^2}{2\epsilon_s \epsilon_o} = -\frac{qN_D\, x_n^2}{2\epsilon_s \epsilon_o} \tag{4.14}$$

4.2.2.1 The Built-in Voltage. In equilibrium, when there is no further movement of carriers and the built-in voltage V_{bi} is established, the electron and hole currents have to be zero. That is, the drift and diffusion components of current for both electrons and holes must cancel. From the current density equations (Eq. 2.139), neglecting generation and recombination processes in the relatively thin depletion region,

$$J_e = q\left(n\mu_e E + D_e \frac{dn}{dx}\right) = 0$$

$$J_h = q\left(p\mu_h E - D_h \frac{dp}{dx}\right) = 0 \tag{4.15}$$

where, carrier flow is assumed in the x-direction. For zero net *hole* current, $E = \frac{D_h}{p\mu_h}\frac{dp}{dx}$ and remembering Eq. 4.5, we get

$$V_{bi} = \frac{D_h}{\mu_h}\int \frac{1}{p}dp \tag{4.16}$$

At any point x,

$$-V(x) = \frac{D_h}{\mu_h} |\ln\ p|_{p_{PO}}^{p(x)} \tag{4.17}$$

Thus,

$$V_{bi} = \frac{D_h}{\mu_h} \ln \frac{p_{PO}}{p_{NO}} \tag{4.18}$$

Using the Einstein relation, Eq. 2.138,

$$V_{bi} = \frac{k_B T}{q} \ln\left(\frac{p_{PO}}{p_{NO}}\right) \tag{4.19}$$

Similarly, starting with the equation for electrons,

$$V_{bi} = \frac{k_B T}{q} \ln\left(\frac{n_{NO}}{n_{PO}}\right) \tag{4.20}$$

from which it follows,

$$n_{PO} p_{PO} = n_{NO} p_{NO} \tag{4.21}$$

Vizio EX1016 Page 0073

The quantities in Eq. 4.21 represent equilibrium majority and minority carrier concentrations on both sides of the junction. Since in equilibrium the relation $n_o p_o = n_i^2$ holds, Eq. 4.19 or 4.20 may be rewritten as

$$V_{bi} = \frac{k_B T}{q} \ln\left(\frac{n_{NO} p_{PO}}{n_i^2}\right) \tag{4.22}$$

Again, assuming full ionization of the dopant impurity levels on either side (i.e., $p_{PO} = N_A$ and $n_{NO} = N_D$), one gets the familiar form,

$$V_{bi} = \frac{k_B T}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \quad (V) \tag{4.23}$$

In this equation V_{bi} is expressed in terms of known and measurable parameters of the materials forming the junction. The energy band diagrams of p- and n-type semiconductors and a p-n junction are depicted in Fig. 4.9(a).

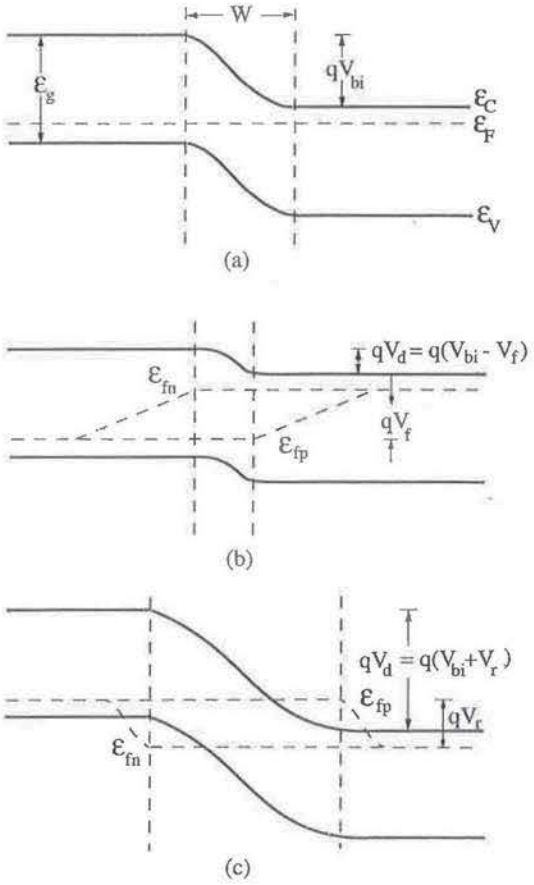


Figure 4.9 Energy band diagrams of a p-n junction (a) in equilibrium, (b) under forward bias, and (c) under reverse bias. The quasi-Fermi levels for the last two cases are also shown.

An alternate and simple way of deriving an expression for V_{bi} is as follows. For Fermi energies in the n- and p-type materials much larger than $k_B T$, the Boltzmann distribution function is valid. Assume the equilibrium majority and minority concentrations in the n- and p-sides to be n_{NO}, p_{NO}, and p_{PO}, p_{NO}, respectively. Then with

Vizio EX1016 Page 0074

respect to Fig. 4.9(a),

$$qV_{bi} = \mathcal{E}_g - (\mathcal{E}_C - \mathcal{E}_F) - (\mathcal{E}_F - \mathcal{E}_V) \tag{4.24}$$

It follows from Eqs. 2.78 and 2.79,

$$qV_{bi} = \mathcal{E}_g + k_B T \ln \frac{n_{NO}}{N_C} + k_B T \ln \frac{p_{PO}}{N_V} \tag{4.25}$$

Again, substituting for the first term on the r.h.s. from Eq. 2.83,

$$\begin{aligned} qV_{bi} &= k_B T \left(\ln \frac{N_C N_V}{n_i^2} + \ln \frac{n_{NO}}{N_C} + \ln \frac{p_{PO}}{N_V} \right) \\ &= k_B T \ln \left(\frac{n_{NO}\, p_{PO}}{n_i^2} \right) \\ &= k_B T \ln \left(\frac{N_A N_D}{n_i^2} \right) \end{aligned} \tag{4.26}$$

which is of the same form as Eq. 4.23. It is obvious that as N_A and N_D tend to n_i, V_{bi} tends to zero. On the other hand as N_A and N_D increase, V_{bi} increases to a maximum value very close to the bandgap energy $\mathcal{E}_g$ under equilibrium conditions.

Finally, a simple way to look at the formation of the potential barrier is to consider the motion of the carriers themselves, when the junction is formed. Holes diffusing from the p-side leave it negatively charged, thereby raising all energy levels. Similarly, electrons migrating in the opposite direction cause all levels in the n-type material to be lowered. As a consequence, the potential hill of height V_{bi} is formed. From the energy levels shown in Fig. 4.9(a) and the Boltzmann distribution it follows that

$$n_{NO} = N_C \exp \left[-(\mathcal{E}_g - \mathcal{E}_F)/k_B T \right] \tag{4.27}$$

and

$$n_{PO} = N_C \exp \left[-\left\{ (\mathcal{E}_g + qV_{bi}) - \mathcal{E}_F \right\} / k_B T \right] \tag{4.28}$$

Therefore,

$$\frac{n_{NO}}{n_{PO}} = \exp \left(\frac{qV_{bi}}{k_B T} \right) \tag{4.29}$$

or

$$V_{bi} = \frac{k_B T}{q} \ln \left(\frac{n_{NO}}{n_{PO}} \right) \tag{4.30}$$

which is of the same form as Eq. 4.20.

4.2.2.2 Depletion Layer Width. With respect to Fig. 4.8, in the space charge region where Poisson's equation is valid, the space charge density is equal to qN_D and qN_A on the n- and p-sides, respectively. Outside the depletion region, in neutral material on either side, the space charge density is zero. For an *abrupt* junction, the maximum electric field E_M is given by

$$\epsilon_s \epsilon_o E_M = qN_D x_n = qN_A x_p \tag{4.31}$$

Vizio EX1016 Page 0075

The area enclosed by the electric field profile is the built-in voltage, V_{bi}, which can therefore be expressed as

$$|-V_{bi}| = \frac{E_M}{2}(x_n + x_p) \tag{4.32}$$

It follows that

$$\begin{aligned} W &= x_n + x_p \\ &= \sqrt{\frac{2\epsilon_s\epsilon_o}{q}\left(\frac{1}{N_D} + \frac{1}{N_A}\right)|-V_{bi}|} \end{aligned} \tag{4.33}$$

If $N_A \gg N_D$, as in a one-sided step junction,

$$W = \sqrt{\frac{2\epsilon_s\epsilon_o}{qN_D}\left|-V_{bi}\right|} \tag{4.34}$$

Similarly, for $N_D \gg N_A$, N_D is replaced by N_A in the denominator of Eq. 4.34. Hence, the depletion layer extends more into the side with the lower doping. The variation of W with doping in Si, Ge and GaAs one-sided abrupt junctions is shown in Fig. 4.10.

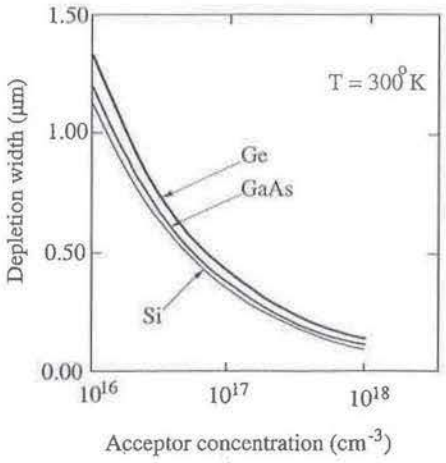


Figure 4.10 Depletion layer width of n^+-p junction for GaAs, Si, and Ge vs. the acceptor density at zero external bias. The donor density in the n^+ region is 5×10^{17} cm^{-3} and T = 300K (from M. Shur, *Physics of Semiconductor Devices*, Prentice Hall, Englewood Cliffs, NJ, 1990).

Until now we have assumed the depletion approximation to be true. In other words, the free-carrier densities abruptly end as step functions at the depletion layer boundaries on both sides. In practice, the free carriers have distributions as shown in Fig. 4.11, which in turn produce a space charge distribution at the depletion layer edge. For example, on the n-side, the space charge density is given by $q[N_D - n(x)]$ instead of qN_D. A similar equation holds for the p-side. As a result, it can be shown

Vizio EX1016 Page 0076

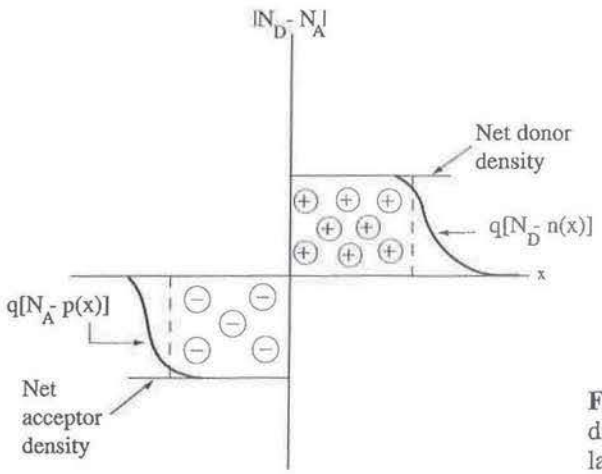


Figure 4.11 Space-charge distribution in a p-n junction with large free-carrier densities.

that (Problem 4.5)

$$W = \sqrt{\frac{2\epsilon_s\epsilon_o}{qN_B}\left(V_{bi} - \frac{2k_BT}{q}\right)}$$

$$= L_D\sqrt{2\left(\frac{qV_{bi}}{k_BT} - 2\right)} \tag{4.35}$$

where $N_B = N_D$ or N_A and L_D is the Debye length given by Eq. 2.114. The effect is therefore also called *Debye tailing*.

EXAMPLE 4.1

Objective. To calculate the extent of the depletion region on the n- and p-sides of an abrupt p-n junction in terms of the total depletion width W and the doping densities N_A and N_D.

From Eq. 4.31 it follows that

$$x_n = \frac{\epsilon_s\epsilon_o E_M}{qN_D},\ x_p = \frac{\epsilon_s\epsilon_o E_M}{qN_A} \tag{4.36}$$

The total depletion layer width

$$W = x_n + x_p$$

$$= \frac{\epsilon_s\epsilon_o E_M}{q}\left(\frac{1}{N_D} + \frac{1}{N_A}\right) \tag{4.37}$$

It follows from Eqs. 4.36 and 4.37

$$x_n = \frac{WN_A}{N_A + N_D},\ x_p = \frac{WN_D}{N_A + N_D} \tag{4.38}$$

Vizio EX1016 Page 0077

4.2.2.3 Junction Capacitance. The depletion layer of a p-n junction can be viewed as a parallel plate capacitor with its capacitance given by

$$C = \frac{\epsilon_s \epsilon_o A}{W} \tag{4.39}$$

where A is the junction area and W is the depletion-layer width. Using Eq. 4.33,

$$C = \frac{\epsilon_s \epsilon_o A}{\left[\frac{2\epsilon_s \epsilon_o}{q}(V_{bi} - V_a)\frac{N_A + N_D}{N_A N_D}\right]^{1/2}} \tag{4.40}$$

where V_a is the applied bias. For an asymmetric junction where $N_A \gg N_D$,

$$C \cong \frac{\epsilon_s \epsilon_o A}{\left[\frac{2\epsilon_s \epsilon_o}{qN_D}(V_{bi} - V_a)\right]^{1/2}} \tag{4.41}$$

These equations represent the bias dependence of the capacitance of a p-n junction. Eq. 4.41 can be rewritten in the form

$$\frac{1}{C^2} = \frac{2}{q\epsilon_s \epsilon_o N_D A^2}(V_{bi} + V_a) \tag{4.42}$$

Therefore, a linear plot of $1/C^2$ versus V_a can be obtained from direct measurements. The slope of this line is inversely proportional to the donor concentration N_D in a p^+-n junction, and its intercept on the voltage axis gives the value of V_{bi}. There are other techniques, however, to determine the built-in voltage of a junction, such as current-voltage and photoemission measurements.

EXAMPLE 4.2

Objective. To show that the peak field E_M is determined by the doping density of the lightly doped side of a p-n junction.

From Eqs. 4.34 and 4.38

$$x_n = \left[\frac{2\epsilon_s \epsilon_o V_{bi}}{qN_D}\left(\frac{1}{1 + N_D/N_A}\right)^2\right]^{1/2} \tag{4.43}$$

For a p^+-n junction in which $N_A \gg N_D$,

$$x_n \cong \left(\frac{2\epsilon_s \epsilon_o V_{bi}}{qN_D}\right)^{1/2}$$

and

$$E_M = \frac{q}{\epsilon_s \epsilon_0} N_D x_n$$

$$= \left(\frac{2qV_{bi}N_D}{\epsilon_s \epsilon_o}\right)^{1/2} \tag{4.44}$$

Therefore, for $N_A \gg N_D$, E_M is proportional to $N_D^{1/2}$. A similar result holds for a n^+-p junction, where $E_M \infty N_A^{1/2}$.

Vizio EX1016 Page 0078

4.2.3 Current-Voltage Relationship

There are four components of current that can flow across the junction. These are $J_h^{p\to n}$, $J_h^{n\to p}$, $J_e^{n\to p}$, and $J_e^{p\to n}$. $J_h^{n\to p}$ and $J_e^{p\to n}$ are due to thermal generation of minority carriers (electron-hole pairs) that travel down the hill by drift motion. $J_h^{p\to n}$ and $J_e^{n\to p}$ depend on the injection of majority carriers to the opposite side and their subsequent diffusion and are therefore exponentially dependent on applied bias. Under equilibrium conditions, the electron and hole currents must exactly balance. Thus, $J_h^{p\to n} = J_h^{n\to p}$ and $J_e^{n\to p} = J_e^{p\to n}$.

4.2.3.1 Junction under Forward Bias. Equilibrium is established at zero bias due to formation of a potential barrier, which restricts the motion of majority carriers. For any net current to flow, the barrier height V_{bi} has to be reduced, and this is achieved by applying a forward bias to the junction. A bias V_a applied to the diode will almost completely drop across the depletion region because of its high resistivity. So the effective bias across the depletion region for a forward bias V_f is $(V_{bi} - V_f)$ and the width of this region is also reduced. $(V_{bi} - V_f)$ is defined as the diffusion potential V_d. The band diagram of a forward-biased junction is shown in Fig. 4.9(b). Current flows across the junctions as drift and diffusion currents are no longer balanced. Majority carriers become minority carriers upon crossing to the other side. It is the change in the minority carrier density on both sides that is significant. The majority carrier densities are hardly affected. The diffusion of carriers cause excess minority carrier densities $\delta n(x)$ and $\delta p(x)$ on the p- and n-sides, respectively, as shown in Fig. 4.12. These excess carrier densities can be determined from a solution of the appropriate *continuity* equation (see Eqs. 8.59-8.62), governing the transport of *minority* carriers. For example, for holes in the n-side, this equation at steady state,

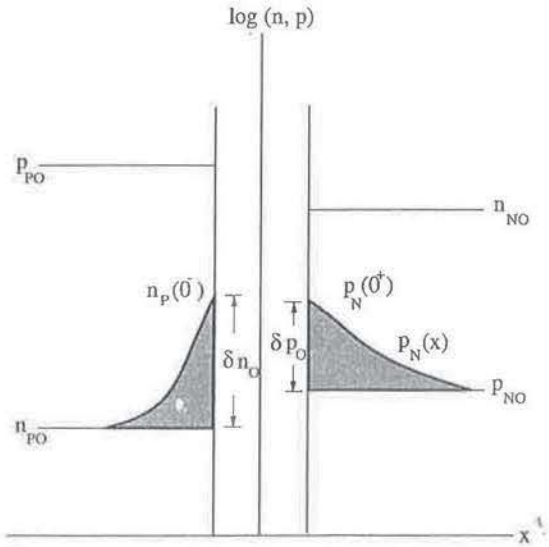


Figure 4.12 Injected carrier densities in a forward-biased junction diode.

Vizio EX1016 Page 0079

neglecting drift, is of the form

$$-\frac{p-p_o}{\tau_h}+D_h\frac{d^2p}{dx^2}=0 \tag{4.45}$$

where p_o is the equilibrium concentration and τ_h is the hole recombination lifetime in the n-region. The solution is of the form

$$\delta p(x)=\delta p_o\, e^{-x/L_h} \tag{4.46}$$

where $L_h=\sqrt{\tau_h D_h}$ is the diffusion length for holes and $\delta p_0=\delta p(0)$ is the excess hole density at the edge of the depletion region. With applied forward bias there is injection of both majority and minority carriers across the junction. With reference to Fig. 4.12,

$$p_{NO}=p_{PO}\, e^{-qV_{bi}/k_BT} \tag{4.47}$$

which signifies that the minority carrier density in one side of a junction is given by the majority carrier density of the other side and the built-in potential. In the following it is assumed that the depletion regions on both sides of the junction have negligible width compared to that of the neutral regions. The minority carrier density $p_N(0^+)$ at the edge of the depletion region on the n-side is given by

$$p_N(0^+)=p_{NO}\, e^{qV_f/k_BT} \tag{4.48}$$

By substitution of Eq. 4.47,

$$p_N(0^+)=p_{PO}\, e^{-q(V_{bi}-V_f)/k_BT} \tag{4.49}$$

Therefore, the excess minority carrier density at the edge of the depletion region is

$$\delta p(0)=p_N(0^+)-p_{NO}=p_{NO}\left(e^{qV_f/k_BT}-1\right) \tag{4.50}$$

and from Eq. 4.46,

$$\delta p(x)=p_{NO}\left(e^{qV_f/k_BT}-1\right)e^{-x/L_h} \tag{4.51}$$

The diffusion current resulting from these excess carriers is given by

$$J_h(x)\;=\;-qD_h\frac{d}{dx}\delta p(x) \tag{4.52}$$

At the origin, or the edge of the depletion region,

$$J_h(0)=J_h(0^+)=\frac{qD_hp_{NO}}{L_h}\left(e^{qV_f/k_BT}-1\right) \tag{4.53}$$

Similarly, the current resulting from the diffusion of electrons on the p-side is

$$J_e(0)\;=\;J_e(0^-)\;=\;\frac{qD_en_{PO}}{L_e}\left(e^{qV_f/k_BT}-1\right) \tag{4.54}$$

And the total current due to forward bias injection is then

$$J=J_h(0^+)+J_e(0^-)\;=\;q\left(\frac{D_hp_{NO}}{L_h}+\frac{D_en_{PO}}{L_e}\right)\left(e^{qV_f/k_BT}-1\right) \tag{4.55}$$

Vizio EX1016 Page 0080

It is important to note the following. First, $J_h(0^+) = J_h^{p\to n} - J_h^{n\to p}$, and $J_e(0^-) = J_e^{n\to p} - J_e^{p\to n}$. Second, that the prefactor in Eq. 4.55 before the exponential term is made up of parameters related to minority carriers.

4.2.3.2 Junction Under Reverse Bias. The band-bending in the junction under reverse bias is shown in Fig. 4.9(c). Again, the applied bias V_r drops almost entirely across the depletion region. The potentials on the p-side are raised and those on the n-side are lowered, resulting in a potential hill of increased height $(V_{bi} + V_r)$, which is now the diffusion potential V_d. Consequently, the depletion layer width is also increased. In terms of Eq. 4.55 above, if V_f is replaced by $-V_r$ and $|V_r|$ is large,

$$J \cong -q\left(\frac{D_h p_{NO}}{L_h} + \frac{D_e n_{PO}}{L_e}\right) = J_s \tag{4.56}$$

where J_s is termed the reverse saturation current of the diode. In terms of drift and diffusion current components, $J_s = J_h^{n\to p} + J_e^{p\to n}$. In other words, the reverse saturation current is composed of the drift current due to the thermally generated minority carriers across the depletion region. The injection of majority carriers and the associated diffusion currents are negligible, due to the large potential barrier. The current-voltage relation of a diode can therefore be expressed as

$$J = J_s\left(e^{qV_a/k_BT} - 1\right) \tag{4.57}$$

where V_a is the applied bias. Typically, for $|V_r|$ larger than a few millivolts, $J = -J_s$. A typical current-voltage plot of a p-n junction is shown in Fig. 4.13(a). The following may be noted about the current-voltage characteristics. Under forward bias greater than a few millivolts (usually $> 3k_BT$)

$$J \cong J_s\, e^{qV_a/k_BT} \tag{4.58}$$

An *ideality factor* n_f is usually incorporated in the exponent as

$$J \cong J_s\, e^{qV_a/n_f k_BT} \tag{4.59}$$

For an ideal diode n_f should be close to unity. Due to recombination and generation processes in the depletion region of real diodes $n_f > 1$. In fact, at low applied biases, n_f is usually larger than unity. Also, for large forward biases, there is a saturation effect, shown by the dashed line in Fig. 4.13(b), in the characteristics of real diodes. This is due to ohmic losses due to the finite resistance of the neutral n- and p-regions.

EXAMPLE 4.3

Objective. To find the *electron* current in the n-type region of a p-n junction under forward bias.

The total junction current I is given by Eq. 4.55. The hole current on the n-side is obtained from Eqs. 4.51 and 4.52 as

$$I_h(x) = qA\frac{D_h}{L_h}p_{NO}e^{-x/L_h}\left(e^{qV_f/k_BT} - 1\right) \tag{4.60}$$

Vizio EX1016 Page 0081

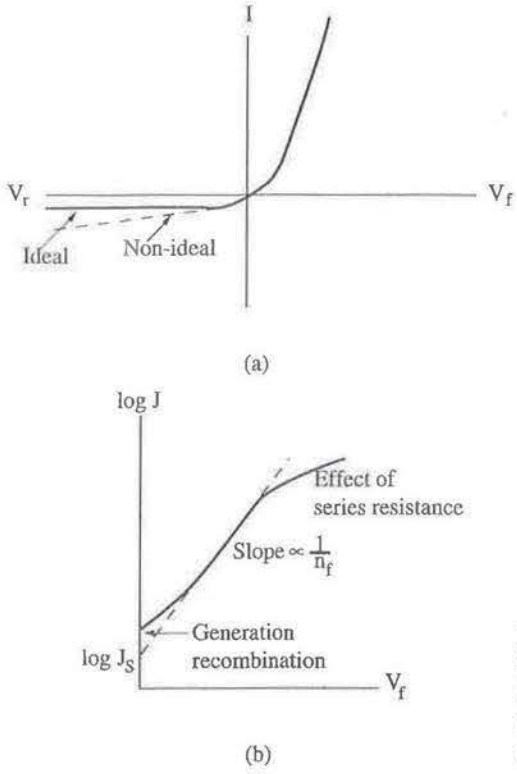


Figure 4.13 Schematic representation of (a) diode current-voltage characteristics and (b) forward-bias current-voltage characteristics in detail.

Therefore, the electron current in the n-region is

$$I_e(x) = I - I_h(x)$$
$$= qA\left[\frac{D_h}{L_h}\left(1 - e^{-x/L_h}\right)p_{NO} + \frac{D_e}{L_e}n_{PO}\right]\left(e^{qV_f/k_BT} - 1\right) \tag{4.61}$$

where A is the area of the diode. This current is the sum of two components. The first accounts for the supply of electrons for recombination with injected holes, and the second is related to the injection of electrons across the junction into the p-side. A similar expression can be written for the hole current on the p-side.

4.2.3.3 Current Densities across the Junction. Consider an abrupt p-n junction, in which the dopings on the p- and n-sides are $\sim 10^{16}$ cm^{-3}. Then the minority carrier densities in equilibrium are $\sim 10^4$ cm^{-3}, assuming the semiconductor to have a bandgap ~ 1 eV. We will examine the current components flowing across the junction under forward bias injection conditions. We will analyze the injection of holes in the n-side of the junction, and the situation on the p-side is exactly complementary. Consider the edge of the depletion region on the n-side to be the origin of the distance coordinates. At the injecting end ($x = 0$), there is a large concentration and concentration gradient of excess injected holes, whose density is given by Eq. 4.46. At this end, therefore, the drift component of current is zero and

Vizio EX1016 Page 0082

the hole current $J_h(x)$ is essentially a diffusion current, $J^h_{diff}(x)$. Therefore,

$$J^h_{diff}(x) = -qD_h\frac{d(\delta p)}{dx} = \frac{qD_h\delta p_o}{L_h}e^{-x/L_h} \tag{4.62}$$

Assuming the total current flowing through the diode is J_o, we know that

$$J_h(0) = J_o = \frac{qD_h\delta p_o}{L_h} \tag{4.63}$$

and

$$J^h_{diff}(x) = J_o e^{-x/L_h} \tag{4.64}$$

Since the total current $J_T = J_e(x) + J_h(x) = J_o$, the nature of $J_h(x)$ and $J_e(x)$ are as shown in Fig. 4.14. Here $J_e(x)$ is the current carried by electrons. On examining the nature of this electron current, it is clear that at the far end, where $J_e(x)$ tends to saturate, it is a drift current, since for diffusion a concentration gradient is necessary. However, closer to the origin, the electron current has both drift and diffusion components. To calculate the diffusion component of the electron current, $J^e_{diff}(x)$, it is necessary to know $\delta n(x)$, which is not simple. However, for charge neutrality to hold and to avoid the generation of high fields in neutral material, the condition $\delta n(x) \cong \delta p(x)$ must be valid. Thus,

$$J^e_{diff}(x) = qD_e\frac{d(\delta p)}{dx} = -\beta_D J_o e^{-x/L_h} \tag{4.65}$$

and

$$\beta_D = \frac{D_e}{D_h} = \frac{\mu_e}{\mu_h} \cong 2 - 10 \tag{4.66}$$

in most semiconductors. Referring to Fig. 4.14, in the region OA there is an electron diffusion current, which is of the same order of magnitude as the hole-diffusion current. Therefore, $J_h(x)$ is entirely a diffusion current and $J_e(x)$ has both drift and diffusion components. The drift component of the electron current

$$J^e_{dr} = qn\mu_e E = \sigma E \tag{4.67}$$

which is Ohm's law. Here σ is the conductivity of the n-region. Adding all the current components,

$$\sigma E - \beta_D J_o e^{-x/L_h} + J_o e^{-x/L_h} = J_o \tag{4.68}$$

from which

$$E = \frac{J_o}{\sigma} + (\beta_D - 1)\frac{J_o}{\sigma}e^{-x/L_h} \tag{4.69}$$

or

$$E = E_\infty\left[1 + (\beta_D - 1)e^{-x/L_h}\right] \tag{4.70}$$

where E_∞ is the electric field at the region far from $x = 0$, where the total current is predominantly an electron current. The electric field, excess carrier density, and electron and hole-current densities are shown in Fig. 4.14. Similar results can be obtained by considering the injection of electrons into the p-side. The carrier distributions and

Vizio EX1016 Page 0083

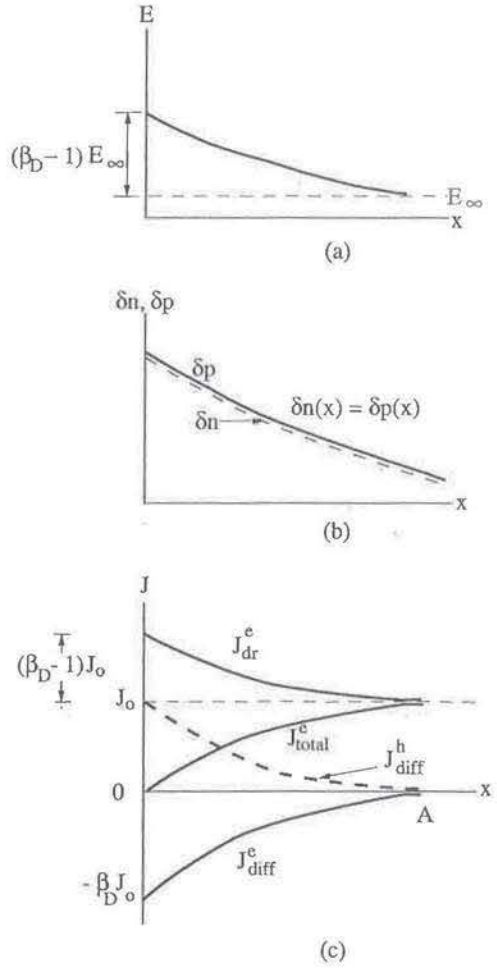


Figure 4.14 (a) The electric field, (b) excess carrier density, and (c) electron and hole current densities in the neutral n-region of a p-n junction under forward bias.

current densities in forward- and reverse-biased junctions are shown in Figs. 4.15(a) and (b).

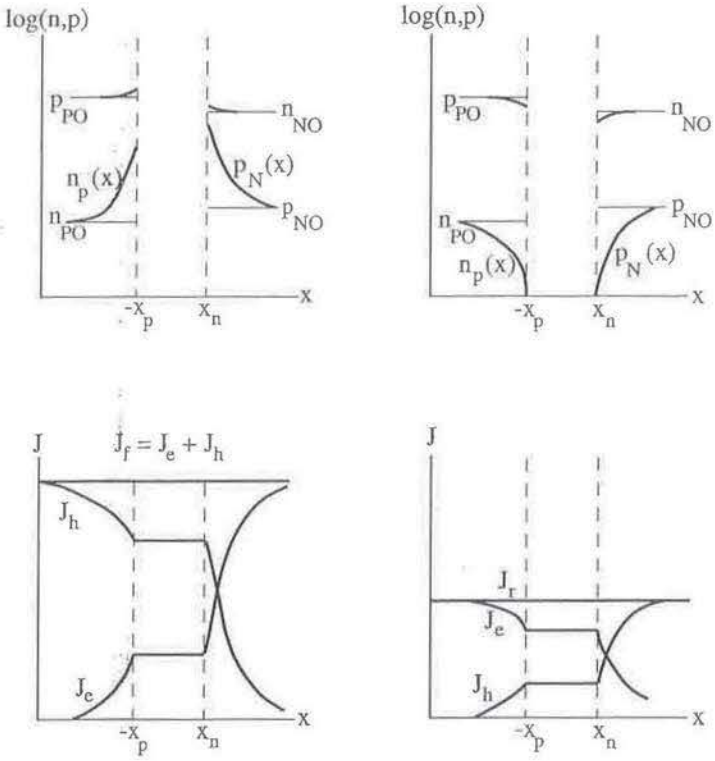
The nature of the electric field profile $E(x)$, shown in Fig. 4.14, merits some discussion. Far from the junction the equilibrium value E_∞ results from the applied bias and in neutral material E_∞ will have a small value. As the junction is approached $E(x)$ increases. The extra field is required to transport electrons from the bulk toward the junction so that the electron density is approximately equal to the injected hole density and the excess holes can recombine.

4.2.3.4 Temperature Dependence of the Reverse Saturation Current. In accordance with Eq. 4.56,

$$J_s = \frac{qD_h\,p_{NO}}{L_h} + \frac{qD_e\,n_{PO}}{L_e} \tag{4.71}$$

To examine the temperature dependence of J_s we consider a p^+-n one-sided abrupt junction. The arguments presented below hold equally for a two-sided junction. For

Vizio EX1016 Page 0084



(a) (b)

Figure 4.15 Carrier distributions and current densities in (a) forward-biased and (b) reverse-biased p-n junction.

the p^+-n junction $p_{NO} \gg n_{PO}$ and the second term in Eq. 4.71 can be neglected. In the first term D_h, p_{NO} and L_h are all functions of temperature. Thus,

$$J_s \cong \frac{q D_h\, p_{NO}}{L_h} = q\sqrt{\frac{D_h}{\tau_h}}\,\frac{n_i^2}{N_D} \tag{4.72}$$

where N_D is the donor density on the n-side and for full ionization $n_{NO} = N_D$. Assuming that the ratio D_h/τ_h has a temperature dependence of the form T^γ, where γ is an integer, and knowing that $n_i \propto [T^{3/2}\, e^{-\mathcal{E}_g/2k_BT}]$,

$$J_s(T) \propto T^{(3+\gamma/2)} e^{-\mathcal{E}_g/k_BT} \tag{4.73}$$

The temperature dependence of the $T^{(3+\gamma/2)}$ term is much weaker than the exponential term. Therefore, the slope of the plot of ln J_s (T) versus inverse temperature is determined by the bandgap $\mathcal{E}_g$. Under reverse bias $|J_r| \cong J_s \propto e^{-\mathcal{E}_g/k_BT}$. Under forward bias $J_f \simeq J_s e^{qV_f/k_BT}$ and the forward bias current is proportional to $e^{[-(\mathcal{E}_g - qV_f)/k_BT]}$.

4.2.4 Quasi-Fermi Levels and High-Level Injection

Until now, we have mostly considered the motion and distribution of carriers in the neutral n- and p-regions of the diode, *outside* the depletion region. As seen in the

Vizio EX1016 Page 0085

previous sections, the current-voltage characteristics can be determined by considering the motion of these carriers. For the sake of completeness, it is important to know what goes on within the depletion region. Under forward bias, as majority carriers are injected to the opposite side, these carriers drift through the depletion region, suffering some recombination. The lost carriers are replenished at the contacts, thereby increasing the diode current. Similarly, under reverse bias, there is generation of carriers in the depletion region. It is important to realize that within the depletion region there is a state of *quasiequilibrium*. It is easy to see the situation under forward bias, when there is a dynamic distribution of a large number of free carriers, both electrons and holes, whose densities as a function of distance cannot be determined by the equilibrium Fermi levels in the neutral materials outside the depletion region. The situation is similar to that created by shining light on a semiconductor, which produces an extra population of electron-hole pairs. Their density also, cannot be determined by the equilibrium Fermi level in the dark. In Chapter 3, the concept of quasi-Fermi levels was introduced to calculate the nonequilibrium density of photogenerated carriers. The carrier densities in terms of the quasi-Fermi levels $\mathcal{E}_{fn}$ and $\mathcal{E}_{fp}$ are given by Eqs. 3.5 and 3.6. Similar conditions exist in the depletion region under forward bias, as seen in Fig. 4.9(b). At equilibrium, $\mathcal{E}_{fn} = \mathcal{E}_{fp} = \mathcal{E}_F$. In terms of the intrinsic Fermi level $\mathcal{E}_{Fi}$,

$$n_i = N_C \exp\left(\frac{\mathcal{E}_{Fi} - \mathcal{E}_C}{k_B T}\right)$$

$$= N_V \exp\left(\frac{\mathcal{E}_V - \mathcal{E}_{Fi}}{k_B T}\right) \tag{4.74}$$

Combining Eqs. 3.5, 3.6, and 4.74, we get

$$n = n_i \exp\left(\frac{\mathcal{E}_{fn} - \mathcal{E}_{Fi}}{k_B T}\right) \tag{4.75}$$

$$p = n_i \exp\left(\frac{\mathcal{E}_{Fi} - \mathcal{E}_{fp}}{k_B T}\right) \tag{4.76}$$

These are general relations that hold for any semiconductor region under nonequilibrium conditions. Note that n and p are different from the equilibrium concentrations n_o and p_o. It follows from Eqs. 4.75 and 4.76,

$$np = n_i^2 \exp\left(\frac{\mathcal{E}_{fn} - \mathcal{E}_{fp}}{k_B T}\right) \tag{4.77}$$

Therefore, $\mathcal{E}_{fn}$ and $\mathcal{E}_{fp}$ are separate energy levels that define the nonequilibrium electron and hole populations, respectively.

Under forward bias $(\mathcal{E}_{fn} - \mathcal{E}_{fp}) > 0$ and $pn > n_i^2$ and for a reverse bias [Fig. 4.9(c)] $(\mathcal{E}_{fp} - \mathcal{E}_{fn}) > 0$ and $pn < n_i^2$. We have seen earlier from Eq. 4.49 that the minority carrier concentration at the edge of the depletion region on either side of a forward-biased junction is equal to the corresponding majority carrier concentration on the other side of the depletion layer, reduced by an exponential factor

Vizio EX1016 Page 0086

$\exp[q(V_{bi} - V_a)/k_BT]$, where $(V_{bi} - V_a)$ is the height of the barrier overcome by the carriers in crossing the space charge region. Thus, the quasi-Fermi levels do not vary in crossing the depletion layer. In other words,

$$\mathcal{E}_{fn}(-x_p) = \mathcal{E}_{fn}(x_n) \tag{4.78}$$

$$\mathcal{E}_{fp}(-x_p) = \mathcal{E}_{fp}(x_n) \tag{4.79}$$

where $(x_n + x_p)$ is the total depletion layer width. Similar arguments can be presented for the reverse-biased junction. It is interesting to follow the variation of the quasi Fermi levels beyond the depletion layer edges. For example, the density of excess holes on the n-side, shown in Fig. 4.12, is given by Eq. 4.46. As $p_N(x)$ approaches p_{NO}, $\mathcal{E}_{fp}$ varies linearly with x_n according to Eqs. 4.46 and 4.76. This linear variation continues for several diffusion lengths. In fact, when $\mathcal{E}_{fp}$ crosses the intrinsic level $\mathcal{E}_{Fi}$, the excess hole density is only n_i. $\mathcal{E}_{fn}(x_p)$ varies in a similar manner.

Under high-level injection conditions, the injected minority carrier densities become comparable to the majority carrier densities, and a substantial portion of the applied bias drops across the regions outside the space charge layer, which itself shrinks. Therefore, both drift and diffusion components of current become important. The current density equations (2.139) combined with Eqs. 4.75 and 4.76 and the fact that $\mathrm{E} = -\frac{1}{q}\frac{d\mathcal{E}_{Fi}}{dx}$, give

$$\begin{aligned} J_e(x) &= q\mu_e n\left(\frac{1}{q}\frac{d\mathcal{E}_{Fi}}{dx}\right) + \mu_e k_B T\left(\frac{n}{k_B T}\left[\frac{d\mathcal{E}_{fn}}{dx} - \frac{d\mathcal{E}_{Fi}}{dx}\right]\right) \\ &= n\mu_e\frac{d\mathcal{E}_{fn}}{dx} \end{aligned} \tag{4.80}$$

and

$$J_h(x) = p\mu_h\frac{d\mathcal{E}_{fp}}{dx} \tag{4.81}$$

Thus, the currents flowing through the diode are determined by the *gradients* of the quasi-Fermi levels. Appreciable gradients of these Fermi levels develop within the depletion region and outside, to account for the spatially dependent concentration of carriers. If the quasi Fermi levels are constant throughout, there is no current flow. From Figs. 4.9(b) and (c) it is clear that the separation between the quasi Fermi levels increases monotonically from zero to a maximum value of V_a. When the separation is zero, $\mathrm{pn} \cong \mathrm{n}_i^2$ and when it is nonzero

$$pn \leq n_i^2\, e^{qV_a/k_BT} \tag{4.82}$$

If $\mathrm{p} \cong \mathrm{n}$, then

$$p = n = n_i\, e^{qV_a/2k_BT} \tag{4.83}$$

This condition holds for high-injection conditions also.

Vizio EX1016 Page 0087

EXAMPLE 4.4

Objective. A GaAs p-n junction diode has $N_D = 2\times10^{16}$ cm^{-3} on the n-side and $N_A = 3\times10^{19}$ cm^{-3} on the p-side. To calculate the forward bias at which the injected hole concentration at the edge of the depletion region on the n-side becomes equal to the majority carrier (electron) concentration.

The condition of high-level injection is $p_N(0^+) = n_{NO}$. Using this equality and Eq. 4.49,

$$n_{NO} = p_{PO}e^{-q(V_{bi}-V_f)/k_BT} \tag{4.84}$$

Substituting the values of N_A, N_D and n_i (for GaAs at 300K) = 6.5 $\times10^6$ cm^{-3} in Eq. 4.23 gives $V_{bi} = 1.35$V. Substituting this value in Eq. 4.84 and knowing that $n_{NO} = N_D$, $p_{PO} = N_A$, we get $V_f = 1.1$V.

4.2.5 Graded Junctions

Practical p-n junctions made by diffusion are seldom abrupt. As we have seen earlier in Sec. 4.1, this is due to the profiles of the diffusing species. Consequently, a graded (doping) region is created on both sides of the metallurgical junction. With ion-implantation, by greater control of the implant profiles, it is possible to achieve near-abrupt behavior, and epitaxial techniques such as molecular beam epitaxy allow us to realize ideal diodes. Nonetheless, it is important to know the parameters for graded junctions. The analysis of such junctions is briefly outlined. We will consider the simplest case, the *linearly* graded junction. In this case, the net doping on both sides of the junction, and consequently the ionized charge densities in the depletion region, vary linearly. The electrostatic parameters of a linearly graded junction are depicted in Fig. 4.16.

For an impurity concentration gradient a_I [cm^{-4}] we have, in the depletion region,

$$-\frac{\partial^2 V}{\partial x^2} = \frac{\partial E}{\partial x} = \frac{q}{\epsilon_s\epsilon_o}(p - n + a_I x)$$

$$\cong \frac{q}{\epsilon_s\epsilon_o}a_I x, \qquad -\frac{W}{2} \le x \le \frac{W}{2} \tag{4.85}$$

where W is the width of the depletion region. Therefore,

$$E(x) = -\frac{qa_I}{\epsilon_s\epsilon_o}\frac{(W/2)^2 - x^2}{2} \tag{4.86}$$

and the maximum field E_M at $x = 0$ is given by

$$|E_M| = \frac{qa_I W^2}{8\epsilon_s\epsilon_o} \tag{4.87}$$

The built-in potential is given by

$$V_{bi} = \frac{qa_I W^3}{12\epsilon_s\epsilon_o} \tag{4.88}$$

Vizio EX1016 Page 0088

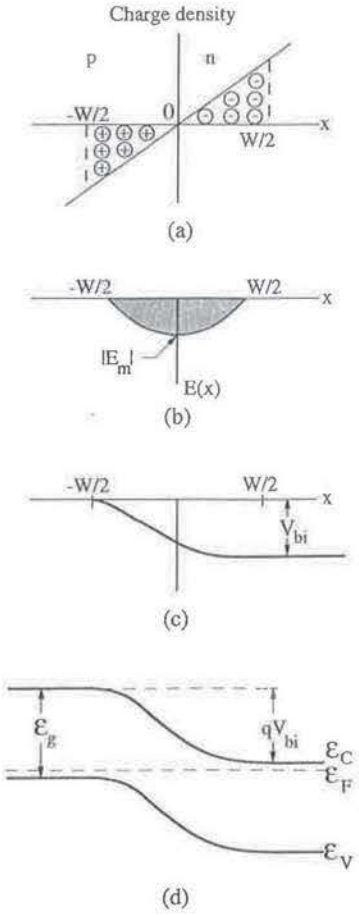


Figure 4.16 Electrostatics of linearly graded junction in thermal equilibrium: (a) space-charge distribution, (b) electric field profile, (c) potential profile, and (d) energy-band diagram.

from which

$$W = \left(\frac{12\epsilon_s\epsilon_o V_{bi}}{qa_I}\right)^{1/3} \tag{4.89}$$

Since the value of the impurity concentration at the edges of the depletion region are $a_I W/2$, by drawing analogy with Eq. 4.23, the built-in potential V_{bi} can be approximately written as

$$V_{bi} = \frac{k_B T}{q} \ln\left[\frac{(a_I W/2)^2}{n_i^2}\right]$$

$$= \frac{k_B T}{q} \ln\left[\frac{a_I W}{2n_i}\right]^2 \tag{4.90}$$

The depletion-layer capacitance as a function of applied bias is given by

$$C = \frac{\epsilon_s\epsilon_o}{W} = \left[\frac{qa_I(\epsilon_s\epsilon_o)^2}{12(V_{bi} \pm V_a)}\right]^{1/3} (\text{Farad.cm}^{-2}). \tag{4.91}$$

The difference of these expressions, with those for abrupt junctions, should be noted.

Vizio EX1016 Page 0089

4.2.6 AC Operation of Diodes: Diffusion Capacitance

The equivalent circuit of a reverse biased p-n junction is shown in Fig. 4.17(a). Here r_b is the total resistance of the bulk n- and p-regions, C_j is the bias-dependent depletion layer capacitance, and g_d is the conductance (usually small) of the junction due to the flow of the small reverse current. r_b and g_d are small compared to C_j. Under forward bias, however, the small-signal equivalent circuit changes to that shown in Fig. 4.17(b). The conductance changes to a different value G_d, C_j also changes and r_b remains fairly constant. However, there is an additional capacitance C_d, which results from the storage and motion of minority carriers on either side of the junction. Remember that it takes a finite time, of the order of 1 nanosecond, for the excess charge to recombine and dissipate. High-frequency operation of the diode is therefore limited by this minority carrier storage time. All bipolar devices, such as the bipolar junction transistor, suffer from this limitation. The operation of unipolar devices, such as the metal-semiconductor Schottky diode to be discussed shortly, depends on the transport of majority carriers only and are therefore not limited by minority carrier storage.

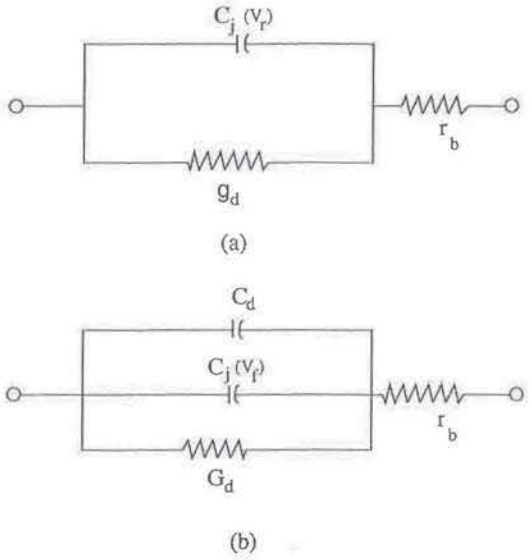


Figure 4.17 Equivalent circuits of a p-n junction under (a) reverse bias and (b) forward bias.

4.2.7 Breakdown Phenomena in Junction Diodes

It is generally assumed that the reverse current of a diode remains at a constant, low value for large values of reverse bias. We know that this is not true because of ongoing generation-recombination processes in the depletion region, which increase the current from the saturation value. In addition, there are ohmic leakage currents around the surface of the junction, if it is not properly encapsulated. Such encapsulation is usually done by the deposition of a dielectric such as SiO_x or Si_3N_4 after junction formation. Finally, and most importantly, at large reverse biases junction breakdown takes place. When this happens, the diode current is mostly limited by the resistance of the external

Vizio EX1016 Page 0090

circuit. In what follows, the various breakdown phenomena will be briefly described. Such breakdown usually depends on material parameters.

Zener breakdown occurs due to tunneling of carriers across the depletion region and the process is independent of temperature. If the doping densities in the p- and n-regions are high the depletion layer width is reduced and even at moderate reverse biases carriers can tunnel through the depletion barrier and add significantly to the reverse current. The process is shown Fig. 4.18(a). As the bias is increased, the tunneling component increases, and the diode eventually breaks down. In the atomic model it is the ionization of a covalent bond. It is important to remember that for Zener breakdown to occur before carrier avalanching, the doping densities should be high and consequently the depletion region should be narrow. If the diode is designed properly, the current in the diode will be *independent* of voltage after breakdown. Such diodes, usually called *Zener diodes*, are used as voltage regulators.

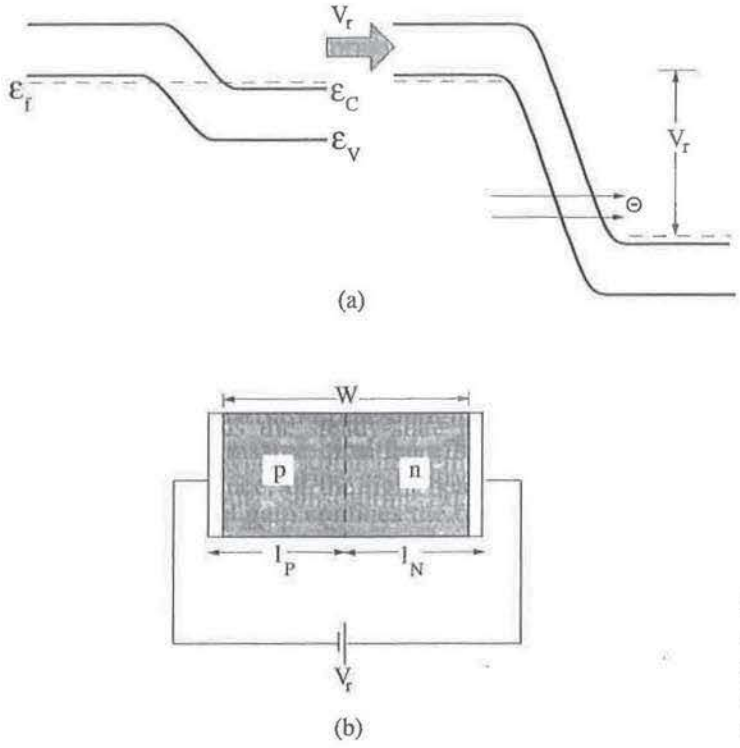


Figure 4.18 (a) Tunneling in a degenerately doped reverse-biased junction and (b) near-punchthrough conditions in a reverse-biased diode.

Avalanche breakdown occurs in diodes with moderate doping density, so that the depletion region is wider. With a large electric field across the depletion region, carriers moving across this region will collide with the lattice. At low fields the momentum and energy of the carriers are redirected, and such collisions are nonionizing. If the energy of the colliding carriers imparted by the field is large enough, then the collision can create an electron-hole pair by breaking a bond. In the band diagram the process is equivalent to raising an electron from the valence band to the conduction band, leaving a hole behind. It is important to realize that for the carriers to gain sufficient

Vizio EX1016 Page 0091

energy for ionizing collisions to occur, they must accelerate over some distance, and therefore the depletion region needs to be wide. After a collision event, the original carrier and the pair created by collision move in their respective directions under the field and suffer more ionizing collisions. Thus, a chain reaction sets in, as a result of which the reverse current increases and the diode eventually breaks down. The bias at which this occurs is called the breakdown voltage V_{BR}. The variation of the avalanche breakdown voltage as a function of carrier density in different semiconductors is shown in Fig. 4.19. If the temperature of the diode increases, the lifetime for lattice scattering decreases and that for ionizing collisions increases, resulting in an increase in V_{BR}. The reverse occurs with lowering of temperature. Thus, unlike Zener breakdown, the avalanche breakdown is a temperature-dependent phenomenon. The process just described—*impact ionization*—is characterized by impact ionization coefficients, α_e and α_h, for electrons and holes, respectively. The coefficients define the number of ionizing collisions per unit length. It may be realized that due to this bias-dependent carrier multiplication process, there is a net gain in the number of carriers. For example, if a photon incident on a photodiode produces an electron-hole pair, the gain is unity. But, if this pair creates many more by the avalanche multiplication process, then the external photocurrent is multiplied and a gain is obtained. This is the principle of operation of the avalanche photodiode, which will be discussed in detail in Chapter 8.

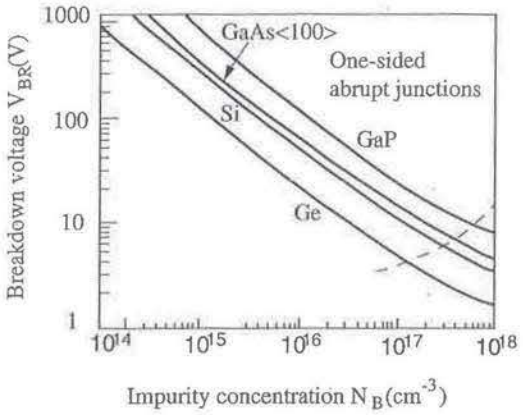


Figure 4.19 Breakdown voltage as a function of doping for one-sided abrupt GaP, GaAs, Si, and Ge p-n junctions (from S. M. Sze and G. Gibbons, *Appl. Phys. Lett.*, **8**, 111, 1966). The tunneling mechanism is dominant for the doping levels higher than corresponding to the dashed line in the low right corner of the figure.

Punchthrough is a breakdown phenomenon dependent primarily on the diode geometry. In a normal reverse-biased p-n junction, as shown in Fig. 4.18(b), holes injected from the contact into the n-region usually recombine in the neutral region before they reach the depletion layer. The same is true for electrons injected into the p-side. Now if the doping is light or if the length of the diode is small, then at some moderate reverse bias, the depletion layer may extend over the entire length. Under these conditions, the minority carriers injected by the contacts cannot recombine and build up in the depletion region and cause large increase of the diode current and eventual breakdown. It is important to remember that this process critically depends on device dimensions, which is not true for the breakdown mechanisms discussed

Vizio EX1016 Page 0092

earlier. For a simple analysis, assume an asymmetric p^+-n diode in which $N_A \gg N_D$. The width of the depletion region, which then extends almost entirely in the n region, is given by

$$W = \left[\frac{2\epsilon_s\epsilon_o}{qN_D}(V_{bi} + V_r)\right]^{1/2} \tag{4.92}$$

where N_D is the doping density on the n-side. If the length of the n-region is l_N, then for $V_R \gg V_{bi}$ punchthrough occurs when

$$W = l_N = \left[\frac{2\epsilon_s\epsilon_o V_r}{qN_D}\right]^{1/2} \tag{4.93}$$

The punchthrough voltage, V_{pt}, is given by

$$V_{pt} = \frac{qN_D l_N^2}{2\epsilon_s\epsilon_o}. \tag{4.94}$$

A nomogram for depletion layer width, junction capacitance and breakdown voltage for GaAs abrupt junctions is shown in Appendix 5.

4.3 SCHOTTKY BARRIERS AND OHMIC CONTACTS

4.3.1 Introduction

These junctions are special cases of the more general class of metal-semiconductor contacts, whose rectifying properties were first identified as early as 1874. It was realized later that metal-semiconductor contacts prepared in a special way, or in which the work functions of the metal and semiconductor have a required sign of difference, could serve as regions through which connections could be made to external power supplies, circuit elements, etc. A metal-semiconductor rectifying contact, or Schottky diode, was first realized as a point-contact and presently made as evaporated metal films on the semiconductor surface. The current-voltage characteristics of rectifying metal-semiconductor junctions are very similar to that of p-n junctions, although there are important differences in the mechanism of current flow and carrier (type) participation. Both electrons and holes contribute to the current in p-n junctions, whereas Schottky diodes are *unipolar* devices. The current is predominantly due to the thermionic emission of these carriers over the potential barrier that is created at the metal-semiconductor junction due to the difference in work function between them. We are more interested here in ohmic metal-semiconductor contacts, in which equal and large currents flow in both forward and reverse bias directions with small resistances. In other words, the potential drop across the junction should be negligible. The evolution of the I–V characteristics in going from a Schottky to an ohmic contact is shown in Fig. 4.20. The subject of both types of metal-semiconductor contacts has been discussed in detail in several texts. However, a brief discussion, covering the essential principles, is made here.

Vizio EX1016 Page 0093

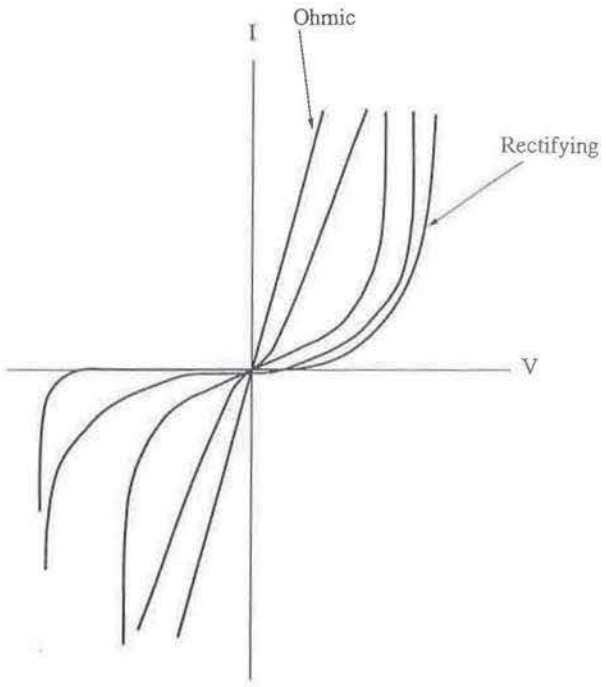


Figure 4.20 Evolution of the current-voltage characteristics of a diode from rectifying to ohmic behavior.

4.3.2 Schottky Barriers

Figure 4.21(a) shows the equilibrium band diagram of a metal of work function ϕ_m and an n-type semiconductor of work function ϕ_s. Note that in this case $\phi_m > \phi_s$. The electron affinity of the semiconductor, measured in energy from the edge of the conduction band to the vacuum level is χ_s. The electron affinity is the energy required to excite an electron from the bottom of the conduction band into vacuum. The work function is the energy needed to excite an electron from the Fermi level into vacuum. When a contact is made between the metal and semiconductor, electrons are transferred from the semiconductor to the metal until the Fermi levels are aligned (i.e., $\mathcal{E}_{Fm} = \mathcal{E}_{Fs}$). Exactly as in a p-n junction, a band-bending or energy barrier and a depletion region are created in the semiconductor, as shown in Fig. 4.21(b). The band-bending or built-in potential V_{bi} is equal to $\phi_m - \phi_s$, and the barrier height ϕ_b is equal to $\phi_m - \chi_s$. To maintain charge neutrality, the positively ionized donors in the depletion layer of the semiconductor are exactly balanced by the charge of a sheet of electrons in the metal at the junction. The barrier height ϕ_b is characteristic of a particular metal-semiconductor combination. Measurement of barrier heights have shown that ϕ_b is often almost independent of the metal species. Surface states resulting from surface imperfections, foreign atoms on the surface, etc., create a double layer on the free surface of a semiconductor, which tends to make the work function independent of the electron concentration in the bulk. If contact is made to a metal, the contact potential difference is compensated largely by a true surface charge rather than by space charge, so that the barrier height is largely independent of the metal. In

Vizio EX1016 Page 0094

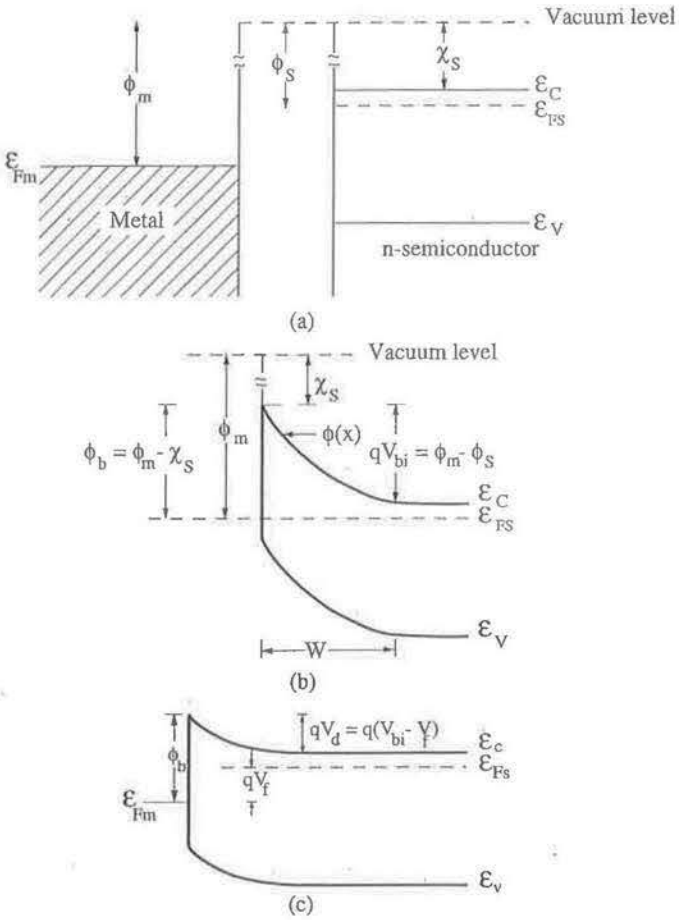


Figure 4.21 Band diagram of metal and n-type semiconductor ($\phi_m > \phi_s$): (a) before contact, (b) at zero applied bias, and (c) under forward bias.

the ideal case, without any surface or interface states, ϕ_b should be a linear function of ϕ_m, with a slope less than unity.

Assuming a uniform distribution of ionized impurities in the semiconductor, Poisson's equation yields the one-dimensional parabolic potential energy barrier described by

$$\phi(x) = \frac{qN_D x^2}{2\epsilon_s \epsilon_o} \quad (V) \tag{4.95}$$

for $0 < x < W$ as shown in Fig. 4.21(b). The depletion layer width W is related to the total diffusion potential, V_d, in a forward biased junction [Fig. 4.21(c)] by

$$V_d = \phi_b - \frac{\mathcal{E}_{Fs}}{q} - V_f = \frac{qN_D W^2}{2\epsilon_s \epsilon_o} \tag{4.96}$$

where V_f is the applied forward bias.

The current in a metal-semiconductor contact under bias is determined by the flow of electrons from the semiconductor to the metal and vice versa. In equilibrium, these two components are equal. Under forward bias, electron flow from semiconductor to

Vizio EX1016 Page 0095

metal is enhanced due to a reduced band-bending and potential barrier $V_d(V)$. The flow from metal to semiconductor remains unchanged, since ϕ_b remains unchanged. This results in a large net *current* flow from metal to semiconductor. Under reverse bias, due to the increase band bending and barrier, electron flow from semiconductor to metal is almost negligible, and therefore the nearly constant small reverse current is due to the flow of electrons from metal to semiconductor over an unchanged barrier ϕ_b. The reverse current therefore flows from semiconductor to metal.

Current flow and rectification in a metal-semiconductor contact are adequately described by thermionic emission theory, as opposed to drift and diffusion mechanisms. For example, under forward bias, the current results from thermionic emission of electrons over the potential barrier. There are, in general, three modes of current transport: thermionic emission, thermionic-field emission, and field emission. These processes are shown in Fig. 4.22. The latter two involve quantum mechanical tunneling through the barrier at energies below the top of the barrier and are more favored in heavily doped semiconductors and large bias voltages. The barrier changes from a rectifying to an ohmic one as these two modes of current transport dominate. In addition, there is recombination in the depletion region, which becomes important under reverse bias, and some recombination in the neutral semiconductor region. The total current density J for ideal thermionic emission over the barrier is given by

$$J = J_f - J_r = A^{**} T^2 \exp(-q\phi_b/k_B T)(\exp(qV_a/k_B T) - 1) \tag{4.97}$$

where J_f and J_r are, respectively, the forward and reverse bias current densities, $A^{**}(\cong 8$ A. cm^{-2}. K^{-2} for Au-GaAs) is the effective Richardson constant for the semiconductor, which takes account of the effective mass and some other corrections, and V_a is the applied bias. Departure from the simple thermionic emission model is accounted for by introducing the ideality factor n_f. Thus, under forward bias,

$$J \infty \exp(qV_a/n_f k_B T) \text{ for } V_a \geq 3k_B T/q. \tag{4.98}$$

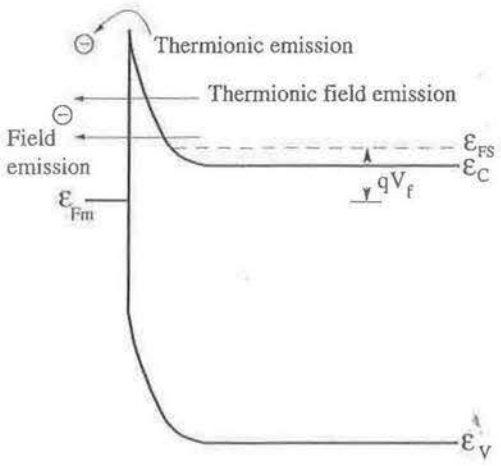


Figure 4.22 Possible conduction processes in a forward-biased Schottky diode with a degenerately doped n-type semiconductor.

Vizio EX1016 Page 0096

In other words,

$$n_f = \frac{q}{k_B T}\left(\frac{dV_a}{d \ln J}\right) \tag{4.99}$$

For a good rectifier $n_f \geq 1$. As shown in Fig. 4.13(b), the linear region of the $\ln J - V$ plot corresponds to $n_f = 1$. At very low biases, the slope of the curve deviates and $n_f \to 2$. Here the current is dominated by generation in the depletion region. Also, for large forward currents, the voltage drop across the series resistance R_s associated with the neutral region of the semiconductor causes the actual voltage drop across the depletion region to be less than the applied bias. Under these conditions the current density is given by

$$J \infty \exp\left[q(V_a - IR_s)/k_B T\right]^{-1} \tag{4.100}$$

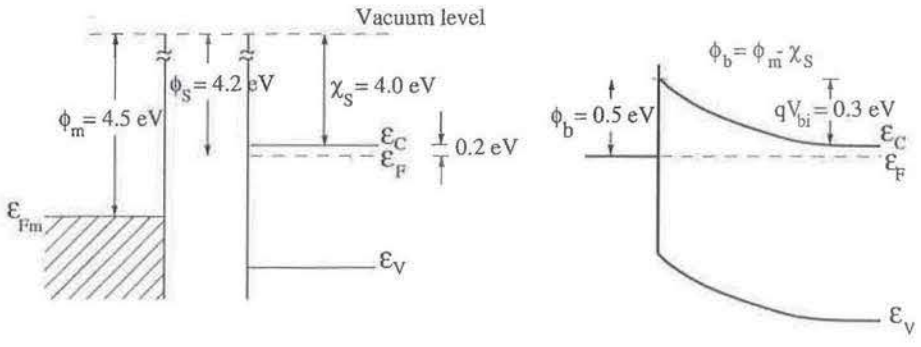
where I is the current through the diode. Under large forward biases, the voltage difference ΔV for any I gives the drop IR_s across the neutral region. Therefore, a plot of ΔV versus IR_s gives the value of R_s.

The discussion so far was related to a metal-(n) semiconductor junction with $\phi_m > \phi_s$. It is seen that under these conditions a rectifying diode is obtained. Similarly a metal-(p) semiconductor will be rectifying if $\phi_s > \phi_m$ and similar equations are valid.

EXAMPLE 4.5

Objective. An ideal Schottky barrier diode is formed on an n-type semiconductor with the following parameters at 300°K: $\mathcal{E}_g = 1.1$ eV, $N_C = 2.5 \times 10^{19}$ cm^{-3}, and $N_D = 10^{16}$ cm^{-3}. The metal work function is 4.5 eV and the electron affinity in the semiconductor is 4 eV. To draw the equilibrium band diagrams and show the energy positions of $\mathcal{E}_{FM}, \mathcal{E}_{FS}, \mathcal{E}_C$ and $\mathcal{E}_V$ before and after the diode is formed.

From Eq. 2.78, the energy $(\mathcal{E}_C - \mathcal{E}_F)$ is 0.2 eV. The band diagrams are as shown in Fig. 4.23.



(a) Before contact (b) After contact

Figure 4.23 Metal-(n) semiconductor junction with $\phi_m > \phi_S$.

Vizio EX1016 Page 0097

At this point it is important to discuss some important differences, including current conduction, between p-n junctions and Schottky barrier diodes. In a Schottky barrier diode made with an n-type semiconductor, the current under forward bias is entirely carried by electrons, even if the doping level is low. In a p-n junction the current is carried predominantly by the majority carrier in the heavily doped side. High-frequency operation of p-n junction diodes is limited by the minority-carrier storage problem. In other words, the minimum time required to dissipate the carriers injected by the forward bias is dictated by the recombination lifetime. In a Schottky diode, electrons are injected from the semiconductor into the metal under forward bias if the semiconductor is n-type. In principle, these "hot" electrons can diffuse back to the semiconductor on the application of a reverse bias. However, they thermalize very rapidly ($\sim 10^{-14}s$) by carrier-carrier collisions, and this time is negligible compared to the minority-carrier recombination lifetime. Finally, Schottky diodes are more suitable for low-voltage applications than p-n junctions. This is because of two reasons. First, the thermionic emission process is much more efficient than the diffusion process and therefore for a given built-in voltage V_{bi}, the saturation current in a Schottky diode is several orders of magnitude higher than in p-n junctions. In addition, the built-in voltage of a Schottky diode is smaller than that of a p-n junction with the same semiconductor. These two factors taken together lead to the fact that to attain the same forward current density, a smaller bias needs to be applied to the Schottky diode than to the p-n junction.

4.3.3 Ohmic Contacts

The band diagram for a metal-(n) semiconductor junction with $\phi_s > \phi_m$ is shown in Fig. 4.24. It is seen that there is no potential barrier, and equal currents will flow under forward and reverse bias. The same is true for a metal-(p) semiconductor with $\phi_m > \phi_s$. Such junctions are termed *ohmic*. Due to the absence of a depletion region, the potential drop across these junctions is negligible at any bias and, therefore, they provide useful regions for contacting a semiconductor device. Unfortunately, metal and semiconductor combinations with the desired relationship between ϕ_m and ϕ_s are hard to come by, and therefore practical contacts cannot be conveniently realized in this way.

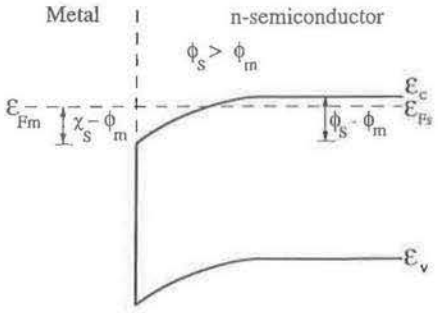


Figure 4.24 Equilibrium band diagram of metal-(n) semiconductor contact with $\phi_s > \phi_m$.

As the doping in the semiconductor is increased, the depletion region becomes narrower. With reference to Fig. 4.22 it is clear that under a reverse bias, when the

Vizio EX1016 Page 0098

depletion region becomes very narrow due to band-bending, thermionic field emission and field emission may become dominant. This can result in a rapid increase of current with reverse bias, which is analogous to Zener breakdown in p-n junctions. A similar mechanism of conduction can prevail under large forward biases. These mechanisms are utilized for realizing practical ohmic contacts. The common technique of forming reliable ohmic contacts to semiconductors is to heavily dope the region under the metal contact by diffusion, implantation, or alloying. For this, a metal alloy is first evaporated and then heated above the eutectic temperature. For example, to realize ohmic contacts to n-type GaAs, Au-Ge is evaporated and alloyed at $\sim$ 400°C. Ge diffuses into the region under the contact and dopes the semiconductor degenerately n-type. A layer of Ni is usually evaporated on top of the Au-Ge to improve the morphology of the contact.

What is important in an ohmic contact is the differential resistance around zero bias, $R_{CT} = (dV/dI)_{V=0}$. For low doping, in the thermionic emission regime, R_{CT} is independent of N_D, while for highly doped material R_{CT} is proportional to $N_D^{-1/2}$ and conduction is predominantly by field emission. For intermediate values of doping thermionic field emission plays a dominant role. It has been found that in practical contacts $R_{CT} \infty 1/N_D$. This has been explained on the basis of current spreading through nonuniform metal-semiconductor regions that resemble protrusions.

More recently, epitaxial techniques such as MBE and OMVPE have allowed the realization of graded bandgap nonalloyed ohmic contacts. For example, highly doped graded $In_xGa_{1-x}As$ $(0 \leq x \leq 1)$ is grown as the contact region and a metal layer is evaporated on top. Such a contact does not require annealing, since the highly doped, low bandgap semiconductor reduces the barrier heights sufficiently to allow ohmic conduction. Low-resistance ohmic contacts are absolutely essential for high-performance devices. For example, if the contact resistances of lasers are not very low, the threshold currents are high, which lead to device heating and inferior performance.

The measurement of contact resistance is technologically very important. The total resistance R_T between a small top contact and a very large bottom contact of a uniform piece of semiconductor is given by

$$R_T = r_{spr} + R_{CT} + R_o \tag{4.101}$$

where r_{spr} is called the spreading resistance under the top contact and accounts for the bulk sample resistance. R_{CT} is the contact resistance of the top contact to be determined and R_o is the resistance of the bottom contact, which is frequently neglected. There are several techniques for measuring contact resistance and a common one is known as the *transfer length method*, originally proposed by Shockley. For a front contact resistance measurement with the geometry shown in Fig. 4.25(a), the total resistance between any two contacts can be shown to be given by

$$R_T = \frac{\rho_s d}{Z} + 2R_{CT} \tag{4.102}$$

where Z and d are indicated in the figure and ρ_s is the sheet resistance in ohm/□. The measured R_T is plotted as a function of d, as shown schematically in Fig. 4.25(b).

Vizio EX1016 Page 0099

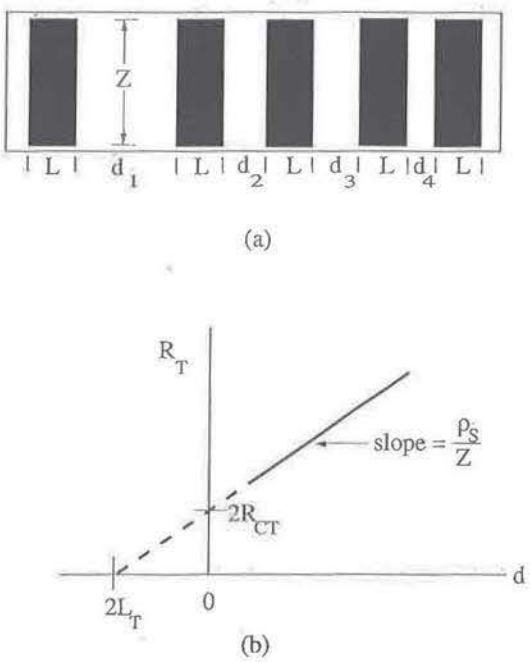


Figure 4.25 Schematic illustration of the transfer length method (TLM) for measuring contact resistance: (a) test structure with contact geometry and (b) plot of total resistance as a function of contact spacing.

Typically $L \sim 50\mu$m, $Z \sim 100\mu$m and $d \sim 5 - 50\mu$m. The intercept at $d = 0$ is $R_T = 2R_{CT}$, giving the contact resistance. The slope of the plot gives the value of ρ_s. A quantity $L_T = \sqrt{\rho_C/\rho_s}$ is defined as the *transfer length*, where ρ_C is the specific contact resistance in ohm.cm^2, which can also be approximately expressed as $\rho_C = R_{CT} L_T Z$ for $L > 1.5L_T$. From these two equations, ρ_C can be calculated by eliminating L_T and knowing ρ_S. A major limitation of this technique is the uncertainty of the sheet resistance under the contacts.†

4.4 SEMICONDUCTOR HETEROJUNCTIONS

4.4.1 Introduction

A heterojunction is formed when two semiconductors with different bandgaps and lattice constants are brought together, usually by epitaxy. Heterojunctions form essential constituents of almost all electronic and optoelectronic devices. For example, in semiconductor lasers heterojunctions provide both optical and carrier confinement. In a bipolar junction transistor, the incorporation of a heterojunction drastically improves the emitter injection efficiency and the current gain of the device. A heterojunction forms a two-dimensional channel of carriers at the interface with superior transport properties, which has been exploited to make high-performance field-effect transistors.

A heterojunction is formed by chemical bonding at the interface. In general, the lattice constants of the two semiconductors are different. If the heterojunction

†For a more detailed discussion, see D. K. Schroder, *Semiconductor Material and Device Characterization*, Wiley, New York, 1990.

Vizio EX1016 Page 0100

is formed by epitaxy, as is usually the case, then there is a misfit between the two semiconductors, A and B. We are more interested in the perfect heterostructure, in which the lattice constants of semiconductors A and B are perfectly matched. The $Al_xGa_{1-x}As$ alloy system, for the entire composition range, $0 \leq x \leq 1$, is almost perfectly lattice matched to GaAs. It therefore forms a useful and important heterostructure system. Lattice matching can also be achieved by growing the appropriate semiconductors with the right compositions. As we have seen in Chapter 1, the ternary alloys $In_{0.53}Ga_{0.47}As$ ($\mathcal{E}_g$ = 0.74 eV) and $In_{0.52}Al_{0.48}As$ ($\mathcal{E}_g$ = 1.45 eV) are lattice matched to InP ($\mathcal{E}_g$ = 1.35 eV). Therefore, very important heterojunctions, useful for optoelectronic devices working in the 1.3–1.6 μm range can be formed with these semiconductors.

4.4.2 The Ideal Heterojunction

Before analyzing a heterojunction in accordance with the model developed by Anderson, we will first qualitatively examine some interesting features of an isotype n-N heterojunction, formed, for example by n-type GaAs and n-type AlGaAs with a larger bandgap. The band diagrams of the two semiconductors, with their Fermi energies, work functions ϕ_1 and ϕ_2, and electron affinities χ_1 and χ_2, are shown in Fig. 4.26(a). Once the heterojunction is formed, electrons transfer from AlGaAs, because of its smaller electron affinity, to GaAs creating a depletion region in the former and an accumulation region in the latter. The formation of this depletion-accumulation region at the interface region is characteristic of heterojunctions. The band diagram of the heterojunction is shown in Fig. 4.26(b). The band offsets $\Delta\mathcal{E}_C$ and $\Delta\mathcal{E}_V$ in the conduction and valence bands, respectively, will be discussed later. It is important to note that the accumulation region in GaAs, depending on the density of electrons transferred from AlGaAs, forms a triangular potential well with a lateral extent of 100–150 Å. The electrons in this well form a two dimensional electron gas and their energy is quantized, just as in a rectangular quantum well. For a p-type isotype heterojunction, a two-dimensional hole gas will be formed in the well created by the valence band discontinuity.

We will now analyze a heterojunction in accordance with Anderson's model,[†] which assumes an ideal interface free of interfacial defect states. A p-N or anisotype heterojunction is considered, consisting of a small bandgap p-type semiconductor and a large bandgap n-type layer. The band diagrams of the two semiconductors before joining, together with the vacuum level, work function, electron affinity, and Fermi energy, are shown in Fig. 4.27(a). The band diagram after formation, with respect to the vacuum level, is shown in Fig. 4.27(b). The band-bending at the junction is equal to the built-in potential V_{bi}, which is given by

$$\begin{aligned} qV_{bi} &= \phi_1 - \phi_2 = (\mathcal{E}_{g1} + \chi_1 - \Delta\mathcal{E}_{F1}) - (\chi_2 + \Delta\mathcal{E}_{F2}) \\ &= \mathcal{E}_{g1} + \Delta\chi - \mathcal{E}_{F1} - \Delta\mathcal{E}_{F2} \end{aligned} \tag{4.103}$$

[†]R. L. Anderson, *Solid State Electronics*, 5, 341, 1962.

Vizio EX1016 Page 0101

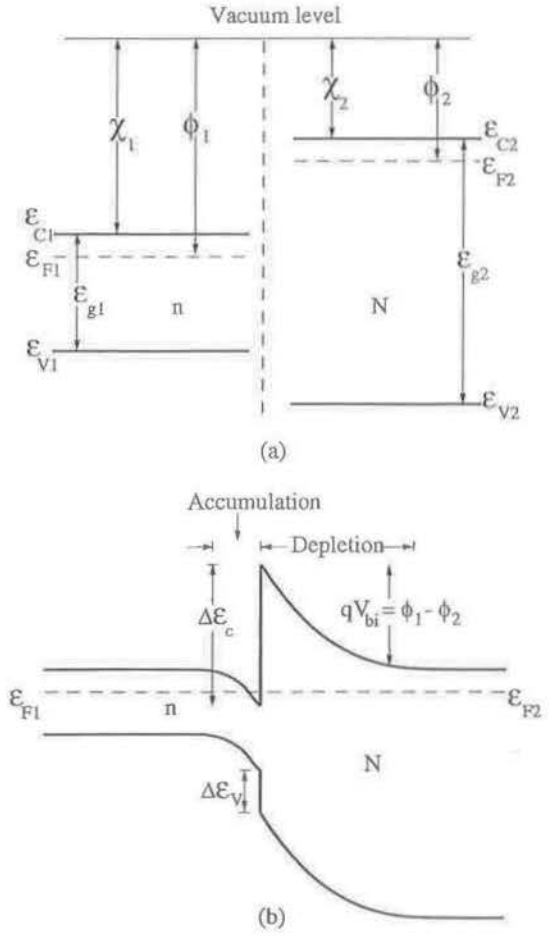


Figure 4.26 The isotype heterojunction: (a) two n-type semiconductors with bandgaps $\mathcal{E}_{g1}$ and $\mathcal{E}_{g2}$ before junction formation and (b) equilibrium band diagram after formation.

According to the Anderson model, the conduction band offset, $\Delta\mathcal{E}_c$, is given by

$$\Delta\mathcal{E}_c = \chi_1 - \chi_2 = \Delta\chi \tag{4.104}$$

Then the valence band offset is given by

$$\Delta\mathcal{E}_V = (\chi_2 + \mathcal{E}_{g2}) - (\chi_1 + \mathcal{E}_{g1}) = \Delta\mathcal{E}_g - \Delta\chi \tag{4.105}$$

Therefore,

$$\Delta\mathcal{E}_C + \Delta\mathcal{E}_V = \Delta\mathcal{E}_g \tag{4.106}$$

An important and interesting case, applicable to heterojunction bipolar transistors, is as follows. If $\chi_1 = \chi_2$, the offset in the conduction band disappears, and $\Delta\mathcal{E}_V = \Delta\mathcal{E}_g$. In this case

$$\begin{aligned} qV_{bi} &= \mathcal{E}_{g1} - \Delta\mathcal{E}_{F1} - \Delta\mathcal{E}_{F2} \\ &= \mathcal{E}_{g1} - k_B T \ln\left(\frac{N_C}{n}\right) \doteq k_B T \ln\left(\frac{N_V}{p}\right) \end{aligned} \tag{4.107}$$

Vizio EX1016 Page 0102

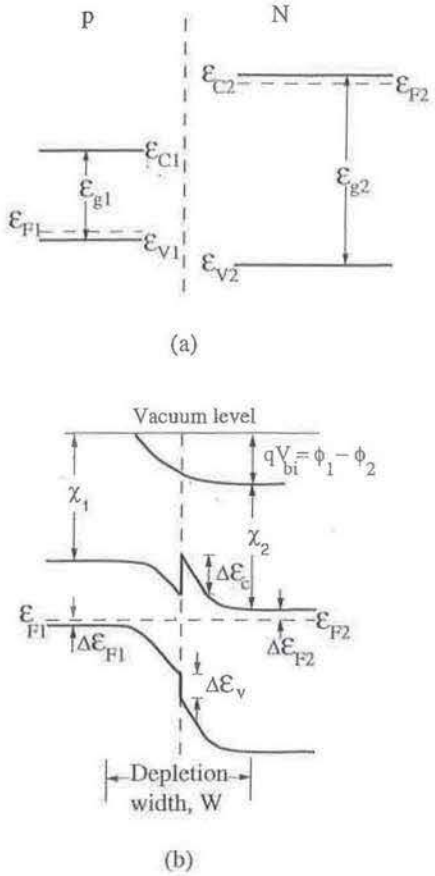


Figure 4.27 Band diagrams of p-N anisotype heterojunction(a) before and (b) after contact.

where n and p are the free-carrier concentrations in the n- and p-type semiconductors, respectively. Equation 4.107 for the built-in voltage is identical to that for a homojunction. However, there is one important difference in this case. The barrier to hole injection is increased by $\Delta\mathcal{E}_V = \Delta\mathcal{E}_g$, compared to a homojunction. Therefore, the electron injection over the barrier is a much more efficient process than hole injection. This effect is utilized in the design of heterojunction bipolar transistors and phototransistors, to be discussed later.

The total built-in voltage V_{bi} at the junction is divided between the p- and n-regions

$$V_{bi} = V_{bi,p} + V_{bi,n} \tag{4.108}$$

where

$$V_{bi,p} = \frac{\epsilon_2 N_D}{\epsilon_2 N_D + \epsilon_1 N_A} V_{bi}$$

$$V_{bi,n} = \frac{\epsilon_1 N_A}{\epsilon_2 N_D + \epsilon_1 N_A} V_{bi} \tag{4.109}$$

Vizio EX1016 Page 0103

The corresponding depletion layer thicknesses in the p- and n-regions are given by

$$x_{dp} = \left[\frac{2\epsilon_1\epsilon_2\epsilon_0 N_D V_{bi}}{qN_A(\epsilon_2 N_D + \epsilon_1 N_A)}\right]^{1/2}$$

$$x_{dn} = \left[\frac{2\epsilon_1\epsilon_2\epsilon_0 N_A V_{bi}}{qN_D(\epsilon_2 N_D + \epsilon_1 N_A)}\right]^{1/2} \tag{4.110}$$

The capacitance-voltage relationship is

$$C(V) = \left[\frac{q\epsilon_1\epsilon_2\epsilon_0 N_A N_D}{2(\epsilon_1 N_A + \epsilon_2 N_D)(V_{bi} - V)}\right]^{1/2} \text{(Farad·cm}^{-2}\text{)} \tag{4.111}$$

It is assumed here that the n- and p-regions are uncompensated and there is full ionization of the donor and acceptor centers in these regions. Note that these relations reduce to those for a homojunction when the dielectric constants (and bandgaps) of the two semiconductors become identical (i.e., $\epsilon_1 = \epsilon_2$).

The interfacial spikes, produced by the band offsets, may sometimes be undesirable for ideal device operation. Their magnitude can be reduced by compositional grading at the interface, which reduces the electron affinity difference at the interface and therefore the field magnitude. This is shown in Fig. 4.28 for an n-P heterojunction.

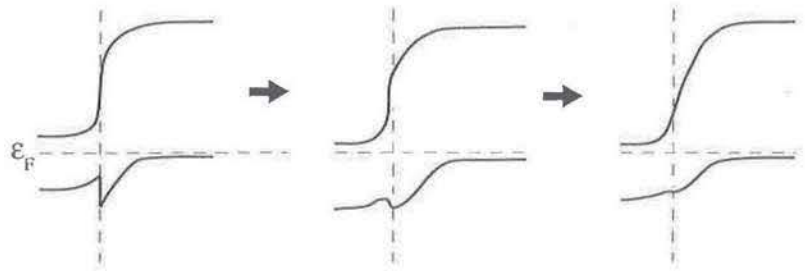


Figure 4.28 Gradual smoothing of the interfacial spike of an n-P heterojunction by compositional grading.

4.4.3 Current-Voltage Characteristics

Under applied forward or reverse bias, the change in the band diagram of a heterojunction at the interface region is very similar to that of a homojunction. Complexities are introduced, however, due to the band offset spikes, and the interfacial defects and traps that are present due to lattice mismatch. There is therefore some uncertainty concerning the current transport mechanism. These include thermionic emission over the barrier, interfacial recombination, and tunneling through the spikes and defect states. Assuming thermionic emission over the heterojunction barrier to be the dominant mechanism, similar to that in a Schottky diode, the current-voltage relation can be expressed as

$$J = \frac{qA^{**}T\ V_{bi}}{k_B} \exp\left[\frac{-qV_{bi}}{k_B T}\right]\left[\exp\left(\frac{qV_a}{k_B T}\right) - 1\right] \tag{4.112}$$

Vizio EX1016 Page 0104

For a more accurate formulation of the current-voltage characteristics, tunneling, drift, and diffusion effects need to be taken into account.

4.4.4 Real Heterojunction Band Offsets

Anderson's model for band offsets assumes a perfect heterojunction without any interfacial defects. The model states that the conduction band discontinuity is given by the difference of the electron affinities of the two semiconductors. The electron affinities of the common compound semiconductors are listed in Table 4.1. A quick examination reveals that the band offsets measured experimentally are very different from those obtained from the electron affinity rule. The "common anion rule" was proposed to explain experimental data. This rule states that the valence band discontinuity $\Delta\mathcal{E}_V$ at the interface will be "very small" for semiconductors with the same anion arising from the fact that the top of valence band states are predominantly anion states. This rule also failed to explain experimental results. What has been observed, however, is that in heterojunctions made up of two semiconductors with different anions, such as $In_{0.53}Ga_{0.47}As$ and InP (lattice matched) or $Ga_{0.5}In_{0.5}P$ and GaAs (lattice matched), $\Delta\mathcal{E}_V > \Delta\mathcal{E}_C$. The trend is in accordance with the common anion rule.

TABLE 4.1 BANDGAP AND ELECTRON AFFINITY OF SOME IMPORTANT SEMICONDUCTORS.

	$\mathcal{E}_g$(eV)	χ (eV)
GaAs	1.43	4.07
AlAs	2.16	2.62
GaP	2.21	4.3
AlSb	1.65	3.65
GaSb	0.73	4.06
InAs	0.36	4.9
InSb	0.17	4.59
InP	1.35	4.35
ZnTe	2.26	3.5
CdTe	1.44	4.28
ZnSe	2.67	3.9
CdS	2.42	4.87
Ge	0.66	4.13
Si	1.11	4.01

More recently Tersoff[†] has proposed a model, based on analogy with a metal-metal junction, which states that the conduction band discontinuity is given by the difference in the Schottky barrier heights of the two semiconductors forming the heterojunction. In other words,

$$\Delta\mathcal{E}_C = \phi_{b1} - \phi_{b2}. \tag{4.113}$$

[†]J. Tersoff, *Physical Review*, **B30**, 4874, 1984.

Vizio EX1016 Page 0105

In this model, it is postulated that a dipole layer is formed at the interface due to electron tunneling from one material into the energy gap of the other. Reasonable agreement with this model has been observed for lattice-matched GaAs/$Al_xGa_{1-x}As$ heterostructures.

Experimentally, early determination of the band offsets was made by Dingle and co-workers in which the band offsets in GaAs/AlGaAs quantum wells was derived from analysis of exciton recombination energies in the luminescence spectra. The fits suggested $\Delta\mathcal{E}_C/\Delta\mathcal{E}_g = 0.85$ for this heterostructure system, and this 85:15 ($\Delta\mathcal{E}_C$: $\Delta\mathcal{E}_V$) rule prevailed for sometime. However, closer examination showed that other fractions may fit the data better and, more significantly, that the fit is quite insensitive to the choice of band offset. A considerably smaller value of $\Delta\mathcal{E}_C/\Delta\mathcal{E}_g \cong 0.6$ was later found by analyzing excitonic energies in photoluminescence excitation spectra from parabolic and square quantum wells. This 60:40 distribution of the discontinuity between the conduction and valence bands has been corroborated by the results of several recent electrical measurements.

Measurements of the conduction band offsets in $In_xGa_{1-x}As/Al_{0.3}Ga_{0.7}As$ quantum wells were made by the author by treating the quantum well as a deep-level trap. The thermal ionization energy of electrons trapped in the well are measured and the band offset is derived from them. Figure 4.29 shows the measured variation of $\Delta\mathcal{E}_C/\Delta\mathcal{E}_g$ with In composition x. It should be cautioned, however, that the quantum wells are biaxially strained and pseudomorphic. The contribution of this strain to the measured data is not known exactly. The offsets are in close agreement to the 60:40 rule.

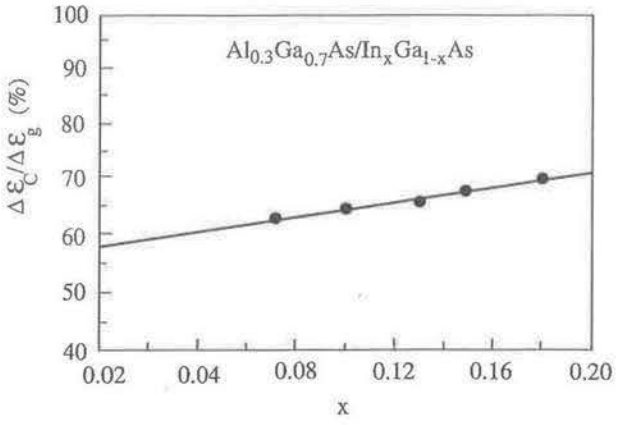


Figure 4.29 Conduction band offsets of $In_xGa_{1-x}As/Al_{0.3}Ga_{0.7}As$ quantum wells measured by the author and co-workers (from N. Debbar et al., *Physical Review*, **B-40**, 1058, 1989).

4.4.5 Application of Heterojunctions to Bipolar Transistors

In Chapter 9 we will be discussing the operating principles of phototransistors. The optical gain of this device is the product of the current gain β_T and the quantum efficiency of the device. It is therefore important that the current gain of the device is reasonably high. In order to put the subject in proper perspective, we will reiterate a

Vizio EX1016 Page 0106

few definitions and relations for a p(emitter)- n(base)- p(collector) bipolar transistor. The schematic of a bipolar transistor is shown in Fig. 4.30, where i_E, i_B, and i_C are the emitter, base, and collector currents, respectively. The emitter-base junction is forward biased, while the collector-base junction is reverse biased. Holes are injected from the emitter to the base and after traversing this region they reach the collector where they contribute to the collector current. There are at least three mechanisms that contribute to the base current i_B, but the dominant one is due to the replenishment of electrons in the n-type base region that recombine with the holes injected from the emitter. Also, electrons emitted from the n-type base region to the p-type emitter must be replenished. The terminal currents can be related by several important factors. Neglecting the collector junction reverse saturation current and recombination in the base-collector depletion region, the collector current is made up *almost* entirely of holes injected from the emitter, which do not recombine in the base region. In other words,

$$i_C = B_T i_{E(p)} \tag{4.114}$$

where

$$i_E = i_{E(p)} + i_{E(n)} \tag{4.115}$$

taking into account the electron and hole injection at the forward-biased emitter junction. B_T is the *base transport factor* and the *emitter injection efficiency*, γ_E, is given by

$$\gamma_E = \frac{i_{E(p)}}{i_{E(p)} + i_{E(n)}} \tag{4.116}$$

We also define the *current transfer ratio* as

$$\alpha_T = B_T \gamma_E \tag{4.117}$$

Finally, we define the current gain,

$$\frac{i_C}{i_B} = \beta_T = \frac{\alpha_T}{1 - \alpha_T}. \tag{4.118}$$

Therefore, for high-current gain α_T, B_T, and γ_E should all be as close to unity as possible. Kroemer defined another measure of transistor performance as the *injection deficit* given by

$$\frac{1 - \gamma_E}{\gamma_E} = \frac{i_{E(n)}}{i_{E(p)}} \tag{4.119}$$

In a good transistor the injection deficit should be small. It is possible to lower the injection deficit by use of a heterojunction (i.e., by using an emitter material with a wider

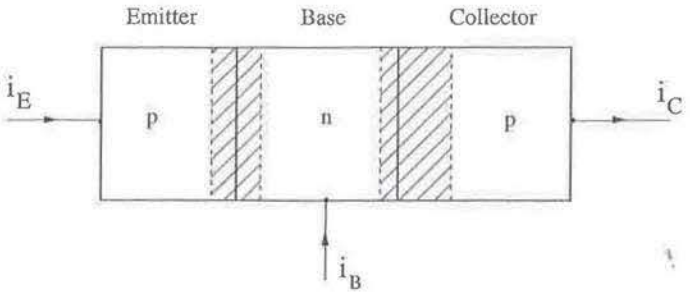


Figure 4.30 Schematic representation of a p-n-p bipolar transistor.

Vizio EX1016 Page 0107

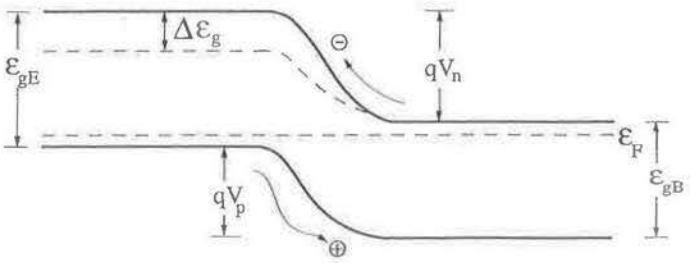


Figure 4.31 (p) emitter- (n) base heterojunction with $\Delta\mathcal{E}_C = \Delta\mathcal{E}_g$. qV_n and qV_p are the potential barriers for electrons and holes, respectively.

bandgap than the base material). The reason becomes clear on considering Fig. 4.31. The potential barrier for electrons flowing from the base to the emitter is larger than the barrier for holes flowing from the emitter to the base. For an applied bias V_a,

$$i_{E(p)} = \frac{qD_h p_o}{L_h}\left[\exp\left(\frac{qV_a}{k_B T}\right) - 1\right]$$

$$i_{E(n)} = \frac{qD_e n_o}{L_e}\left[\exp\left(\frac{qV_a}{k_B T}\right) - 1\right] \tag{4.120}$$

where D_h and L_h are the diffusion constants and diffusion lengths, respectively, for holes in the base and D_e and L_e are the same parameters for electrons in the emitter. n_o and p_o are the equilibrium minority carrier densities in the emitter and base regions. Therefore,

$$\frac{i_{E(n)}}{i_{E(p)}} = \frac{D_e L_h n_o}{D_h L_e p_o} \tag{4.121}$$

We know that

$$n_o = \frac{n_{iE}^2}{N_A}, \quad p_o = \frac{n_{iB}^2}{N_D} \tag{4.122}$$

where n_{iE} and n_{iB} are the intrinsic carrier concentrations in the emitter and base regions, respectively, and N_A and N_D are the majority carrier (doping) concentrations in these regions. Furthermore,

$$\frac{n_{iE}^2}{n_{iB}^2} = \left(\frac{m_{eE}^* m_{hE}^*}{m_{eB}^* m_{hB}^*}\right)^{3/2} \exp\left[-\Delta\mathcal{E}_g/k_B T\right] \tag{4.123}$$

where m_e^* and m_h^* are the effective masses in the two regions and $\Delta\mathcal{E}_g = \mathcal{E}_{gE} - \mathcal{E}_{gB}$. Therefore,

$$\frac{i_{E(n)}}{i_{E(p)}} = \frac{D_e L_h N_D}{D_h L_e N_A}\left(\frac{m_{eE}^* m_{hE}^*}{m_{eB}^* m_{hB}^*}\right)^{3/2} \exp\left[-\Delta\mathcal{E}_g/k_B T\right] \tag{4.124}$$

Remembering that for a homojunction

$$\frac{i_{E(n)}}{I_{E(p)}} = \frac{D_e L_h N_D}{D_h L_e N_A} \tag{4.125}$$

the injection deficit in a heterojunction transistor is largely controlled by the exponential factor. For example, even for $\Delta\mathcal{E}_g = 0.2$ eV, at room temperature, with $k_B T \simeq 0.025$ eV, the injection deficit is *decreased* by a factor of almost 3000.

Vizio EX1016 Page 0108

There are other advantages of having a heterojunction at the emitter-base junction. One can afford to greatly increase the base doping and decrease the emitter doping, without sacrificing the emitting injection efficiency. As a consequence of the higher base doping the base spreading resistance in the direction transverse to current flow is lower. The lower emitter region doping makes the emitter-base junction capacitance small. Both these factors contribute to a higher speed of operation of the transistor.

PROBLEMS

4.1 **(a)** Discuss in a few words the following:

1. quasi-Fermi levels
2. Debye tailing in a p-i-n junction

(b) Sketch the *hole* quasi-Fermi level in the quasi-neutral n-region of a p^+-n junction at

1. equilibrium
2. reverse bias
3. forward bias

4.2 A semiconductor sample is composed of two regions as shown in Fig. P4.2. Consider regions I and II to be far away from the p-n junction, so that these regions are in equilibrium. At a particular temperature the hole concentration in region I is determined to be $p_1 = 1.1 \times 10^{16}$ cm^{-3}.

(a) Find the intrinsic concentration n_i for the semiconductor at this temperature.

(b) Find the equilibrium electron concentration n in region II at this temperature.

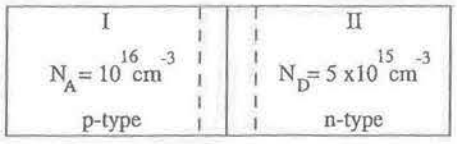


Figure P4.2

4.3 $Al_{0.3}Ga_{0.7}As$ is doped as a $n_1 - n_2$ step junction shown in Fig. P4.3, where $N_{D1} < N_{D2}$. Draw the energy band diagram of such a junction and indicate V_{bi} on this diagram. Also draw the charge density and electric field profile for such a junction.

4.4 A Si p-n junction diode has a rectangular cross-sectional area of 0.5 mm × 1.0 mm and the thickness of the p- and n-regions are 1.0 mm each. Calculate the forward bias necessary to drive 1 mA and 10 mA through the diode. What is the voltage drop across the bulk p- and n-regions for these

Vizio EX1016 Page 0109

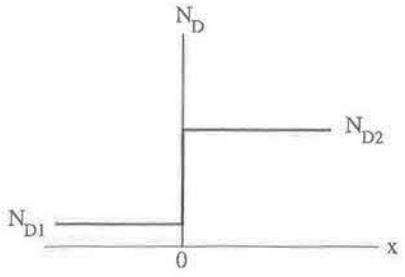


Figure P4.3

forward currents?

[Given: $J_S = 10^{-15} A/cm^2$, $N_A = 10^{16}$ cm^{-3}, $N_D = 10^{17}$ cm^{-3}.]

4.5 Derive Eq. 4.35 under the conditions of Debye tailing in the depletion layer of a junction diode.

4.6 An abrupt p^+-n junction is formed with InP having $N_D = 10^{15}$ cm^{-3}. The avalanche breakdown voltage is -80V. What is the minimum thickness of the n-region to ensure avalanche breakdown rather than punchthrough?

4.7 **(a)** Draw the band diagrams of a metal-n semiconductor ($\phi_m > \phi_s$) under a small forward and a large reverse bias. Show clearly the positions of the Fermi level and the different potentials.

(b) A metal and p-type Si ($\phi_s > \phi_m$) are brought together to form a junction. Will the junction be ohmic or rectifying? If the doping in the Si layer is p = 10^{17} cm^{-3}, $\phi_m = 3$ eV, $\chi_s = 2.5$ eV, $\mathcal{E}_g$ (Si) = 1.1 eV, $n_i = 1.5 \times 10^{10}$ cm^{-3}, calculate: (1) V_{bi}, (2) ϕ_b.

4.8 Alternate heavily doped p-type and n-type layers of the same semiconductor, each 100 Å thick, are grown by molecular beam epitaxy. Draw the energy band diagram of this structure and discuss some of its special properties, including electron-hole recombination rates.

4.9 List and explain the differences in the operation of a p-n junction and a metal-n (semiconductor) Schottky diode with $\phi_m > \phi_s$.

4.10 Consider a p-n junction with the following parameters:

$$\text{conductivity } \sigma_p \gg \sigma_n = 1(\Omega - cm)^{-1}$$
$$D_p = 50cm^2/sec$$
$$\mu_n = 1000cm^2/V.s.$$
$$L_h = 1cm$$
$$I = 100mA$$
$$V \gg k_B T/q$$
$$\text{Area A} = 1.25 \times 10^{-2} cm^2$$

You may assume that L_e is of the same order of magnitude as L_h. Is the low injection assumption valid for this device? Explain in detail, including analytical and/or numerical calculations to justify your answer.

Vizio EX1016 Page 0110

4.11 We know that the capacitance of an abrupt p-n junction varies as $V_r^{-1/2}$ for $V_r \gg V_{bi}$ where V_{bi} is the built-in potential and V_r is the reverse bias applied across the junction. The capacitance of a linearly graded junction varies as $V_r^{-1/3}$. In a certain application a capacitance that varies as V_r^{-1} for $V_r \gg V_{bi}$ is needed. Qualitatively, discuss the general form of doping profile needed. Indicate in each of the three cases the variation of the depletion-region width with bias.

4.12 Using the Shockley-Read-Hall theory discussed in Chapter 2, derive an expression for the generation current of a reverse-biased p-n junction diode. What is the temperature dependence of this current?

4.13 In an experiment with a Schottky diode, its current-voltage characteristics are measured as a function of temperature. What parameters can be extracted from such data? Is there any difference between forward-bias and reverse-bias data? If so, explain why.

4.14 Schottky diodes are majority-carrier devices, and therefore can operate at higher speeds compared to p-n junctions. There is still a preference for the p-n junction over the Schottky diode. State what you think may be the reasons.

4.15 The capacitance C of a Schottky diode of 1 cm^2 area is measured as a function of forward and reverse biases. A linear relationship is found that can be expressed as $1/C^2 = A - BV$, where V is the applied bias and A and B are constants. Derive expressions for built-in voltage V_{bi} and the donor density $N_D(cm^{-3})$ in the semiconductor.

4.16 A p-n junction diode has a small alternating voltage superimposed on a steady state forward bias applied to it. Assuming small signal conditions and a low frequency of the modulating bias, derive an expression for the diffusion capacitance C_d.

READING LIST

MILNES, A. G. *Semiconductor Devices and Integrated Electronics*. Van Nostrand Reinhold, New York, 1980.

RHODERICK, E. H. *Metal-Semiconductor Contacts*. Clarendon Press, Oxford, 1978.

SHARMA, B. L., and PUROHIT, R. K. *Semiconductor Heterojunctions*. Pergamon Press, London, 1974.

SHUR, M. *Physics of Semiconductor Devices*. Prentice Hall, Englewood Cliffs, NJ, 1990.

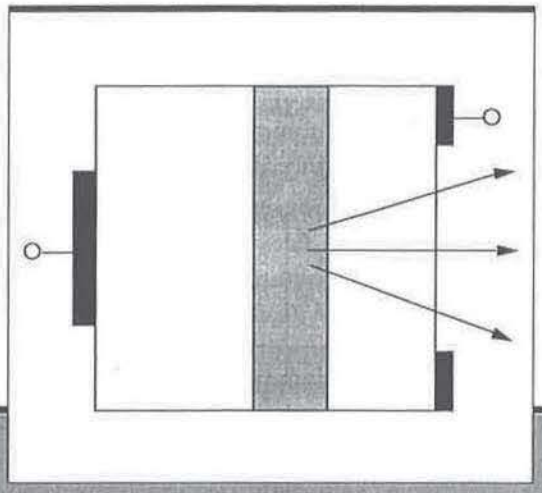
SZE, S. M. *Physics of Semiconductor Devices*, 2nd ed. Wiley, New York, 1981.

TIWARI, S. *Compound Semiconductor Device Physics*. Academic Press, San Diego, 1992.

Vizio EX1016 Page 0111

5

Light-Emitting Diodes



Chapter Contents

Vizio EX1016 Page 0113

5.1 INTRODUCTION

The light-emitting diode (LED) was discovered in 1904. The only semiconductor diode in existence at that time was the silicon carbide (SiC) point contact rectifier. Needless to say, the level of light emission from such devices was very small. Since then, light emission from semiconductor devices has attracted the interest of researchers all over the world. However, it is only recently, with the development of wider bandgap III–V compounds, that visible light emission from a biased p-n junction has been demonstrated. The LED has found wide applications in the form of indicator lamps, display elements, and sensors.

In a junction LED photons of near-bandgap energy are generated by the process of *injection luminescence* or *electroluminescence*, in which a large population of electrons, injected into a normally empty conduction band by forward bias, recombine with holes in the valence band. The device emits light by spontaneous emission, which is different from *stimulated emission* operative in junction lasers. The spontaneous emission process requires a smaller forward bias, and therefore LEDs operate at lower current densities than lasers. An LED also does not require an optical cavity and mirror facets to provide feedback of photons. The emitted photons have random phases, and therefore the LED is an incoherent light source. The linewidth of the spontaneous emission is approximately equal to the photoluminescence linewidth, which is a few times $k_B T$ and is typically 30–50 nm at room temperature. Thus, many optical modes are supported and therefore the LED is a convenient multimode optical source, suitable for use with multimode fibers.

Some of the obvious advantages of the LED as a light source are simpler fabrication procedures, lower cost, and simpler drive circuitry. Light-emitting diodes are, in general, more reliable than lasers because they operate at lower powers and therefore do not suffer catastrophic degradation or even gradual degradation prevalent in lasers. Ideally the LED exhibits a linear light output–current characteristic, unlike that of an injection laser, and is therefore very suitable for analog modulation. The light-current characteristic is also less sensitive to temperature than that in a laser. Some of the obvious disadvantages are the low power output, harmonic distortion due to the multimodal output, and a smaller modulation bandwidth. However, superradiant LEDs are being developed, which have output powers comparable to that of lasers.

5.2 THE ELECTROLUMINESCENT PROCESS

The LED converts input electrical energy into output optical radiation in the visible or infrared portion of the spectrum, depending on the semiconductor material. The energy conversion takes place in two stages: first, the energy of carriers in the semiconductors are raised above their equilibrium value by electrical input energy, and second, most of these carriers after having lived a mean lifetime in the higher energy state, give up their energy as spontaneous emission of photons with energy nearly equal to the bandgap $\mathcal{E}_g$ of the semiconductor. If the semiconductor can be doped p- and n-type, then the energy of the current carriers can be increased by applying a forward bias to

Vizio EX1016 Page 0114

a p-n junction. As we have seen in Chapter 4, under forward bias, minority carriers are injected on both sides of the junction and these excess minority carriers diffuse away from the junction, recombining with the majority carriers as they do so. This process is illustrated in Fig. 5.1. The excess hole concentration in the n-side, away from the depletion region, is given by Eq. 4.46 and a similar equation is valid for excess electrons on the p-side. Most of the excess minority carriers on both sides of the junction recombine radiatively with the majority carriers to create photons of frequency ν given by

$$h\nu \cong \mathcal{E}_g \tag{5.1}$$

A small fraction of the excess minority carriers do recombine nonradiatively, and the excess energy of these carriers is dissipated as heat in the lattice. The rate of radiative recombination is normally proportional to the forward bias injection rate, and hence to the diode current under such conditions, given by Eq. 4.55. Most of the recombination occurs close to the junction, although some of the minority carriers diffuse away and hence recombination can occur in regions further away from the junction. The *internal quantum efficiency* of the device is defined as the rate of emission of photons divided by the rate of supply of electrons. As mentioned above, all the injected electrons do not recombine radiatively with holes, and therefore the quantum efficiency may be less than 100%. Under reverse bias applied to the diode, no injection takes place across the depletion region, and consequently no light is emitted.

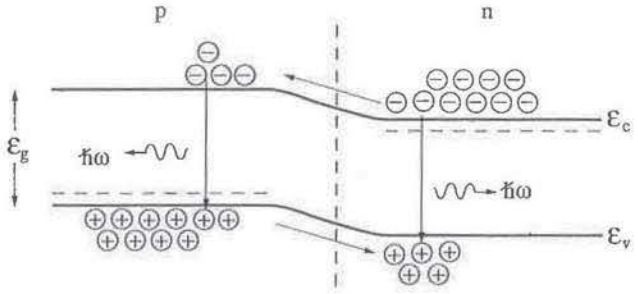


Figure 5.1 Injection of minority carriers in a forward-biased p-n junction leading to spontaneous emission of photons.

In using a semiconductor material for light emission, a wavelength (or energy) of emission close to the bandgap is of importance. As we have seen in Chapter 3, the transitions of interest are band-to-band, impurity-band, donor-to-acceptor, and excitonic transitions. The peak in the density distribution of electrons in the conduction band occurs at an energy $k_BT/2$ above the band minimum. Therefore, the band-to-band transition energy is slightly larger than $\mathcal{E}_g$, and some of the emitted photons are reabsorbed. This self-absorption does not exist in indirect bandgap semiconductors due to the involvement of phonons. However, because of phonons participating in the radiative transition in indirect semiconductors, their radiative efficiency is low. Fortunately, most radiative transitions involve energy levels within the bandgap for which the emission energy is smaller than the bandgap.

Impurity-induced radiative transitions are usually impurity-band and donor-acceptor transitions, both of which have been discussed in Chapter 3. However, there is an interesting aspect of impurity levels, which is important for the operation of LEDs.

Vizio EX1016 Page 0115

A carrier in an impurity state is "trapped" in the potential of the impurity, and is therefore localized in space. As the impurity level becomes deeper, or its binding energy increases, the localization of the carriers is enhanced. It may be noted that the impurity potential is Coulombic for shallow levels and more like a delta function for deep levels. In either case, as a consequence of the Heisenberg uncertainty principle, a spatial localization of the carriers in the impurity levels causes a spread in their momentum (or k) values. The spread of the impurity states in k-space is sufficient to allow a significant rate of transitions between the impurity states and the band edges *without* the involvement of phonons. This, in effect, increases the radiative efficiency of indirect bandgap semiconductors. An example, which is of tremendous commercial importance, is the presence of O and N increasing the radiative efficiency of GaP. Oxygen produces a deep donor level 0.8 eV below the conduction band, while N produces a shallow isoelectronic trap. The impurity-related transitions are the most important ones for LED operation, since most junction LEDs are made of doped materials. Since the radiation resulting from impurity-related transitions has a lower photon energy than the bandgap, it is not subject to reabsorption to the same extent as band-to-band transitions.

Excitonic transitions have been discussed in Chapter 3. Since the design of LEDs involves doped materials, the radiative transitions of interest are those involving bound excitons. The binding energy of the bound exciton centers depends on the binding energy of the impurity to which the excitons are associated with, and is given by

$$\mathcal{E}_{ex} \cong 0.1\mathcal{E}_i \tag{5.2}$$

where $\mathcal{E}_i$ is the binding energy of the impurity level. Again, the excitons can be bound to neutral or ionized impurities depending on the doping level and the temperature of the sample. Bound excitons may also be sufficiently localized so that electron-hole recombination can take place in indirect bandgap semiconductors via these states, without involving phonons to conserve momentum.

Having described the various near-bandgap transitions via which electrons and holes may recombine radiatively, it is important to answer the question *Which transition is dominant in the operation of a junction LED*? As one may guess at this point, all the transitions described above may be involved, depending primarily on the doping in the sample and also to some extent on the impurity levels and the bias applied. Values of the spontaneous recombination coefficient, B_r, in indirect semiconductors are significantly lower than those in direct bandgap materials. Thus band-to-band transitions are not efficient in indirect semiconductors and usually impurities producing deep levels are incorporated to increase the radiative efficiency. An example, as already mentioned, is O in GaP and GaAsP.

5.3 CHOICE OF LED MATERIALS

The wavelength of light emission required usually dictates the semiconductor materials required, in terms of their bandgap energy. An equally important factor is the ability to heavily dope these materials n- and p-type and thereby fabricate a junction diode.

Vizio EX1016 Page 0116

Lower bandgap materials are required for infrared and far-infrared applications, and larger bandgap materials are needed for a light source in the visible part of the spectrum (around 2 eV). The higher the bandgap, the higher is the melting point of the material, and the lower is their structural stability. Also, the higher bandgap materials have higher resistivity and cannot be easily doped to high levels, as may be desirable for LED fabrication.

The most important III–V compounds for application to LEDs are the binary compounds GaAs and GaP, and their ternary derivative $GaAs_{1-x}P_x$. The properties of these materials were discussed in Chapter 1. GaAs has a direct bandgap of 1.43 eV (λ = 860 nm) at room temperature and therefore emits in the near-infrared region. Light-emitting diodes are usually made by Zn-diffusion into pulled n-GaAs or Si-doping of solution grown GaAs. Si, being amphoteric in nature, can incorporate into the GaAs crystal as a donor or acceptor, depending on growth temperature. In Zn-doped devices, the main radiative transition is a band-to-band process, and consequently there is a large reabsorption of the emitted radiation. The emission wavelength of Si-doped diodes is $\sim$ 940 nm and reabsorption is much less severe. Here the main recombination transition is from an (impurity-related) defect to the conduction band, and the complex defect behaving as an acceptor is believed to have a binding energy of $\sim$ 0.1 eV. Typical external quantum efficiencies of these diodes are of the order of 10%. However, Si-doped GaAs LEDs are more expensive to manufacture than the Zn-doped ones.

GaP is an indirect bandgap semiconductor with $\mathcal{E}_g$ = 2.26 eV (λ = 549 nm), and therefore impurity levels in the bandgap are needed for it to be used as an efficient light emitter. GaP is commonly used for red and green emission. Nitrogen acts as a donor-like isoelectronic trap in GaP with a binding energy of 8 meV, remembering that it belongs to the same group in the periodic table as P. This isoelectronic trap forms a bound exciton, which emits at a near-bandgap wavelength of 550 nm (green). The quantum efficiency is, however, even smaller than 1%. It has been found, however, that higher levels of N doping ($\geq 10^{14}$ cm^{-3}) give rise to defect levels in the bandgap due to the formation of N-N molecular complexes. The corresponding emission is at 590 nm (yellow). Red emission is obtained from GaP simultaneously doped with Zn and O, which occupy, respectively, substitutional Ga and P sites in the lattice. Oxygen as an impurity in GaP gives rise to a deep donor with a binding energy of 0.8 eV. However, when Zn and O atoms occur as nearest neighbors in the lattice, they form an isoelectronic trap with a binding energy of $\sim$ 0.3 eV. The bound exciton associated with this trap level recombines to produce emission at λ = 690 nm.

The ternary $GaAs_{1-x}P_x$ changes from direct to indirect bandgap material for x = 0.45, at which composition $\mathcal{E}_g$ = 2.1 eV. Thus, the direct bandgap and indirect bandgap materials with N doping can be used for the fabrication of diodes to emit red ($GaAs_{0.6}P_{0.4}$), orange ($GaAs_{0.35}P_{0.65}$: N), and yellow ($GaAs_{0.15}P_{0.85}$: N) light.

For longer wavelengths in the near-infrared region of the spectrum, InP-based compounds such as the ternary $In_{0.53}Ga_{0.47}As$ and the quaternaries $In_xGa_{1-x}P_yAs_{1-y}$ are important. These are newer materials, and studies are in progress to fully understand their radiative properties. Devices made of these materials operate in the 1.1–1.6 μm range, which overlaps with the spectral region of low loss and minimum

Vizio EX1016 Page 0117

dispersion in optical fibers. They are therefore important for optical-fiber communication. For the far-infrared region of the spectrum, the antimonides such as GaSb, InSb, and their alloys will eventually be important. However, the use of these compounds is hampered by the lack of suitable lattice-matched substrates. Similarly at the other end in the range of bandgaps, the binary gallium nitride (GaN) ($\mathcal{E}_g$ = 3.5 eV) is extremely important for emission of blue light, which would be useful for satellite communication. However, doping the material p-type is a problem. Other promising semiconductor materials that are being developed for making blue LEDs are the IV–IV compound silicon carbide (SiC) and the II–VI compounds. The materials and device technology of SiC are difficult due to its mechanical and chemical properties. The high melting point makes growth of the compound difficult. Doping with Al, B, Be, and Sc give rise to emission in various parts of the spectrum and currently yellow-emitting LEDs are commercially available. Diamond as an LED material also has considerable potential. The II–VI compounds such as ZnS and ZnSe have promise as visible LED materials, but p-type doping of these compounds still eludes success. Similarly compounds such as CdTe, HgTe, ZnTe, and their ternary derivatives with which heterojunctions can also be formed are under intensive study and will be applicable for commercial visible LEDs.

5.4 DEVICE CONFIGURATION AND EFFICIENCY

The schematic of a typical surface-emitting LED is shown in Fig. 5.2. The junction is usually formed by diffusion, or the whole structure is grown by epitaxial techniques. The device is usually designed such that most of the radiative recombination takes place in the side of the junction nearest the surface whereby the probability of reabsorption is greatly reduced. This is ensured by arranging most of the current flowing across the diode to be carried by those carriers (type) that are injected into the surface layer. As shown in the figure, such conditions are achieved in an n^+-p junction where the p-layer is closer to the surface.

The processes occurring in a junction LED can be divided into three stages. The first is the *excitation* or *injection* process, in which the energy of carriers is raised by forward bias injection. Next is the *recombination* process, during which most of these carriers give up their excess energy as photons. Finally, the generated photons must leave the semiconductor and provide the desired optical stimulus to the eye or produce a photocurrent in a detector. This is the *extraction* process. Each of these processes have a characteristic efficiency, and the overall device efficiency, η_o, may then be expressed as

$$\eta_o = \eta_{in}\eta_r\eta_e \tag{5.3}$$

where η_{in}, η_r, and η_e respectively are the injection, radiative recombination, and extraction efficiency. η_o can also be termed the external conversion efficiency.

5.4.1 Injection Efficiency

In an asymmetric n^+-p junction diode, such as the devices shown in Fig. 5.2, electron injection is much more dominant than hole injection. For small carrier diffusion

Vizio EX1016 Page 0118

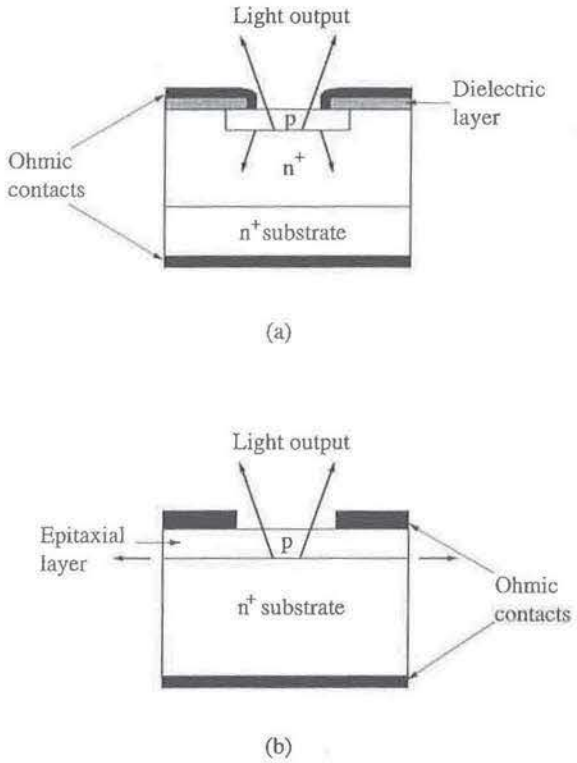


Figure 5.2 Planar surface-emitting LED structure made by (a) diffusion and (b) epitaxial growth.

lengths, which is usually the case in most III–V compounds, it is almost impossible to determine the injection efficiency experimentally. However, η_{in} can be calculated from junction theory seen in Chapter 4. The fraction of the total diode current that is carried by diffusion of electrons being injected into the p-side of the junction is given by

$$\eta_{in} = \frac{\frac{D_e n_{PO}}{L_e}}{\frac{D_e n_{PO}}{L_e} + \frac{D_h p_{NO}}{L_h}} \tag{5.4}$$

from which we get

$$\eta_{in} = \left(1 + \frac{D_h L_e p_{NO}}{D_e L_h n_{PO}}\right)^{-1} \tag{5.5}$$

where the terms are as defined in Chapter 4.

Using the Einstein relation and the equality $n_o p_o = n_i^2$, Eq. 5.5 becomes

$$\eta_{in} = \left(1 + \frac{\mu_h p_{PO} L_e}{\mu_e n_{NO} L_h}\right)^{-1} \tag{5.6}$$

where, for full ionization $p_{PO} = N_A$, the acceptor doping on the p-side, and $n_{NO} = N_D$, the donor doping on the n-side. In most III–V compounds $\mu_e \gg \mu_h$ and L_e is of the same order as L_h. Under these conditions $\eta_{in} \cong 1$. From Eq. 5.6 it is evident that a high-injection efficiency can be ensured by making $N_D \gg N_A$ (i.e., an

Vizio EX1016 Page 0119

asymmetric junction). It should be noted that, in calculating the injection efficiency the recombination in the space charge (depletion) layer of the junction is neglected. From Shockley-Read-Hall theory the current due to space charge recombination is given by

$$I_{rec} = \frac{qW}{2} s\vartheta_{th} N_T n_i A \left(e^{qV_f/2k_BT} - 1\right) \tag{5.7}$$

where s and N_T are, respectively, the capture cross section and density of the recombination center or traps that are present, $\vartheta_{th} = (3k_BT/m^*)^{1/2}$ is the thermal velocity of the carriers, n_i is the intrinsic carrier concentration, A and W are the area and depletion-layer width of the diode, respectively, and V_f is the applied forward bias. In small bandgap materials the recombination current is negligible due to a lower density of deep-level traps. Note that the product $(s\vartheta_{th}N_T)$ is equal to the inverse of the lifetime of the deep-level recombination center. As the bandgap increases this component of current density increases and can be as high as 10^{-2}A/cm^2 in large bandgap materials. Also, under certain special circumstances, such as in the presence of a large density of interface traps caused by impurities, or dislocations as in a lattice-mismatched heterojunction, space charge recombination may represent a significant shunt path at low forward biases. In general, L_e and L_h are functions of N_A and N_D, respectively. For example, in GaAs, L_e and L_h are fairly constant for N_A, $N_D \leq 10^{18}$ cm^{-3}, but for higher doping, L_e and L_h are inversely proportional to the respective dopings. In GaAs, values of $\eta_{in} \sim 0.8$ can be easily achieved, while that in GaP ranges from 0.6 to 0.8. However, η_{in} is not important in determining η_o, since its value is higher than the other two efficiencies in Eq. 5.3 and a moderate variation in its value does not have a significant effect on the value of η_o.

5.4.2 Recombination Efficiency

The operation of the junction LED relies on the spontaneous emission of photons and the absence of any optical amplification process through stimulated emission, as in a laser, limits the *internal quantum efficiency* (η_i) or radiative efficiency (η_r). The latter is defined as the ratio of the number of photons generated to the number of injected electrons. The radiative efficiency will increase if the possible nonradiative processes can be eliminated. The main nonradiative processes that are operative in an LED structure are Shockley-Read-Hall recombination, band-to-band and trap-assisted Auger recombination, and recombination via surface states.

In direct bandgap materials the radiative recombination efficiency is usually $\sim$ 50% for homojunction devices. In indirect bandgap semiconductors this efficiency is usually very low unless, as we have seen earlier, recombination can take place via an impurity level. In order to achieve a reasonable radiative efficiency in an indirect semiconductor with a substitutional impurity, carriers must be tightly bound to the core of the impurity atom. This requires a deep level in the energy gap and a transition related to such a level will be at a wavelength much larger than that corresponding to the energy gap. For III–V compounds such as GaP or GaAsP, this implies emission in the near-infrared or infrared parts of the spectrum. The best measured values of η_r for GaP: Zn, O, and GaP: N are 30% and 3%, respectively.

Vizio EX1016 Page 0120

Nonradiative surface recombination is more important in direct bandgap materials than in indirect bandgap materials. However, with the use of double heterojunction (DH) structures, as in lasers, the radiative efficiency or internal quantum efficiency can be as high as 60–80%.

5.4.3 Extraction Efficiency and External Conversion Efficiency

It is ultimately the amount of light emerging from the device that determines the overall external efficiency of the device. This efficiency can be low in spite of high values of injection and radiative efficiencies if the light is not extracted efficiently. Several factors determine the extraction efficiency. The absorption coefficient of the semiconductor at the emission wavelength plays an important role. This is dictated, to a large extent, by whether the material has a direct or indirect bandgap. The absorption and luminescence spectra of GaAs and GaP: Zn,O are approximately shown in Figs. 5.3(a) and (b). In GaAs the absorption coefficient at the peak of the output is $\sim 10^3$ cm^{-1}, which implies that after passing through as little as 2 μm of material about

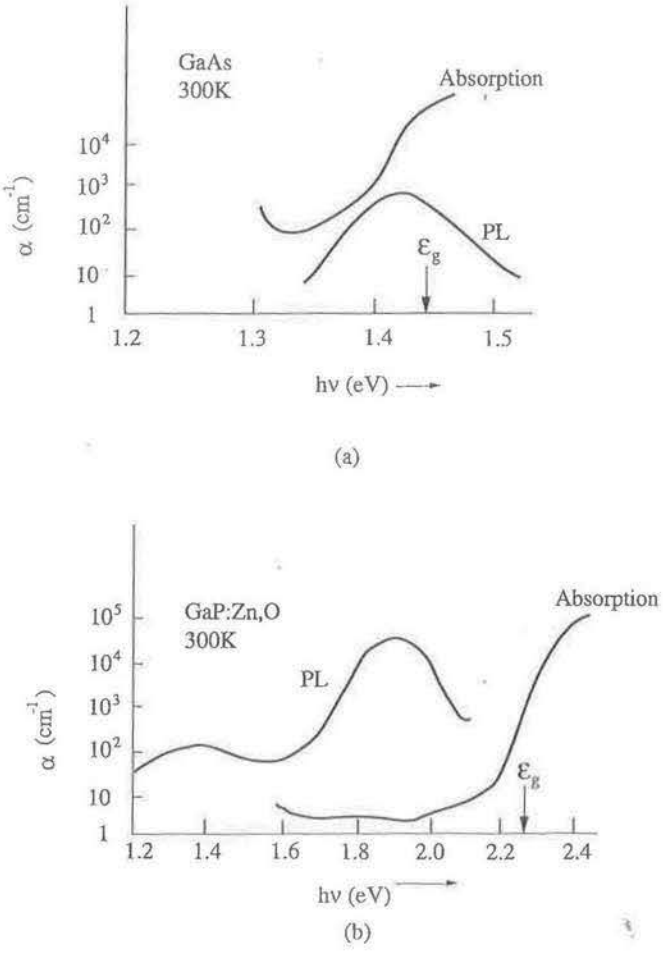


Figure 5.3 Apsorption and luminescence spectra of (a) GaAs and (b) GaP:Zn, O at room temperature.

Vizio EX1016 Page 0121

half of the emitted radiation is reabsorbed, since $\phi(x) = \phi_o e^{-\alpha x} = \phi_o e^{-0.6} \cong \phi_o/2$ for $\alpha = 3 \times 10^3$ cm^{-1}. This may be contrasted with the case of GaP: Zn,O in which $\alpha \simeq 3$ cm^{-1} at the peak of the emission spectrum and the reabsorption in a few microns of material is negligible. The obvious answer seems to be to place the light-emitting junction very close to the surface in a direct bandgap device. Unfortunately, surface states will reduce the radiative recombination efficiency as the junction approaches the surface. As we shall see in the next section, the use of a heterojunction alleviates this problem to a large extent.

Another cause for the lowering of the external quantum efficiency is the radiation geometry of the LED, which emits through a planar surface. The intensity distribution of the radiation is Lambertian, as shown in Fig. 5.4(a), in which the power radiated from a unit area into a unit solid angle, expressed in units of W sr^{-1} m^{-2} is constant in all directions. However, most of the emitted radiation strikes the semiconductor-air interface at angles greater than the critical angle, θ_c, and so remains trapped by total internal reflection. The high dielectric constant (refractive index) of compound semiconductors makes the critical angle very small. The critical angle is given by

$$\theta_c = \sin^{-1}\left(\frac{n_{r2}}{n_{r1}}\right) \tag{5.8}$$

where n_{r2}(air) = 1 and n_{r1} is the refractive index of the semiconductor. Therefore light originating near the junction will be radiated isotropically, of which only that

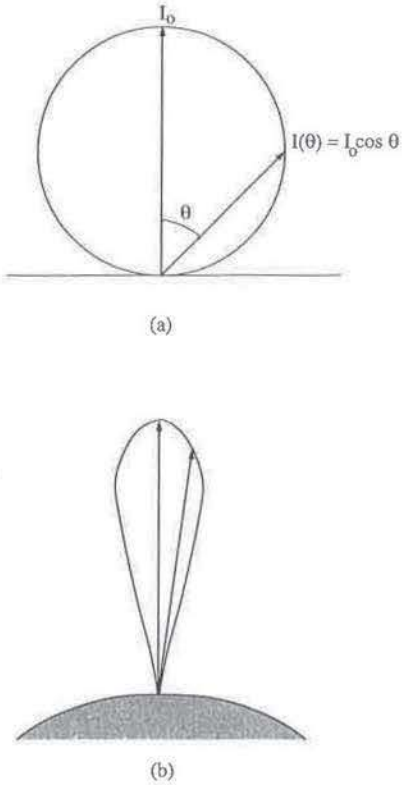


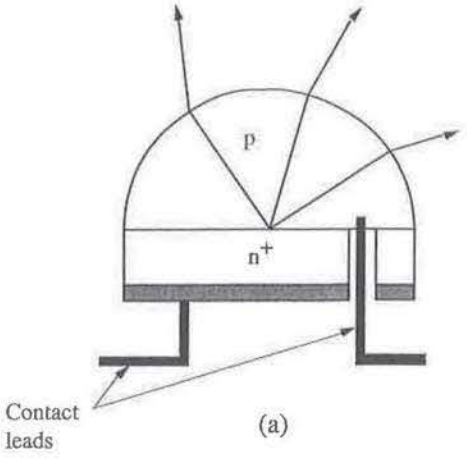
Figure 5.4 Intensity distribution of radiation from a light-emitting diode: (a) Lambertian and (b) directional.

Vizio EX1016 Page 0122

within a cone of semi-angle θ_c will leave the device surface. It can be shown that the transmission F_T (light radiated to the total light generated) at the semiconductor-air interface is approximately given by (Problem 5.2):

$$F_T = \frac{1}{4}\left(\frac{n_{r2}}{n_{r1}}\right)^2\left[1-\left(\frac{n_{r1}-n_{r2}}{n_{r1}+n_{r2}}\right)^2\right] \tag{5.9}$$

Two techniques that can be used to increase F_T are shown in Fig. 5.5. In the first, shown in Fig. 5.5(a), the semiconductor surface is shaped as a hemisphere. This is, however, expensive and impractical for most applications. The second, and cheaper technique, shown in Fig. 5.5(b), is to use dielectric encapsulation shaped as a dome with a transparent material of high refractive index. For example, with $n_{r1} = 3.6$



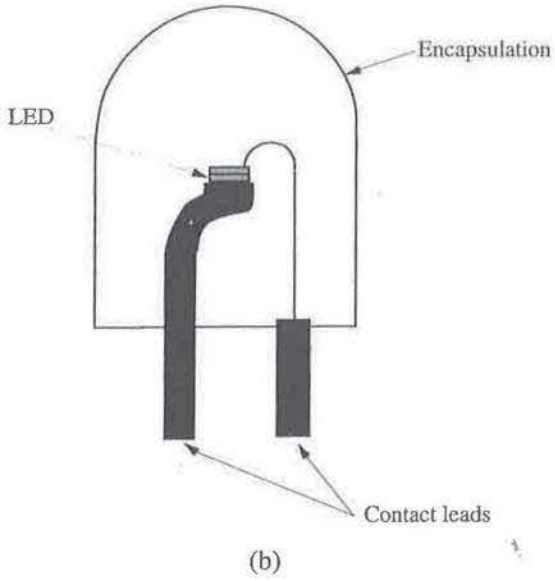


Figure 5.5 Two types of LED construction to enhance light transmission from device: (a) hemispherically shaped device surface and (b) encapsulated device or dome LED.

Vizio EX1016 Page 0123

and $n_{r2} = 1.6$, a fivefold increase in light output is obtained over the planar semiconductor/air interface. These are called *dome LEDs*, and are useful for high-volume, low-price applications such as intruder alarms, displays, calculators, etc. Some planar and dome LEDs are also used in optical communication in the near-infrared with fiber bundles. The radiation patterns from these structures are much more directional, as shown in Fig. 5.4(b).

The external power efficiency, or conversion efficiency, η_o, is defined as the ratio of the optical power output P_o to the electrical power input P_e. In other words,

$$\eta_o = \frac{P_o}{P_e} \times 100\%. \tag{5.10}$$

It may be noted that the losses encountered in the extraction process is included in η_o. Typical values for η_o in common LEDs are in the range of 1% to 5%.

Finally, in an LED for daily and common applications it is important to take into account the physiological impact of the device. For this purpose it is more appropriate to express the performance in photometric units, which normalizes the output spectrum of the device with the eye response. Thus, the *luminous efficiency*, or *luminosity*, is expressed as

$$\eta_L \propto \frac{\int_\lambda V(\lambda) P_o(\lambda) d\lambda}{\int_\lambda P_o(\lambda) d\lambda} \tag{5.11}$$

where $P_o(\lambda)$ is the LED emission spectrum and $V(\lambda)$ is the eye response shown in Fig. 5.6. It is apparent that a diode emitting at the peak of the eye response in the green or yellow will appear much brighter than a device emitting an equal amount of energy in the red or blue. As an example, a green-emitting LED (λ = 515 nm) will appear nearly 30 times brighter than an LED emitting in the red (λ = 630 nm) but having the same quantum efficiency. Taking the eye response into account

$$\eta_o = \eta_{in} \eta_r \eta_e \eta_L \tag{5.12}$$

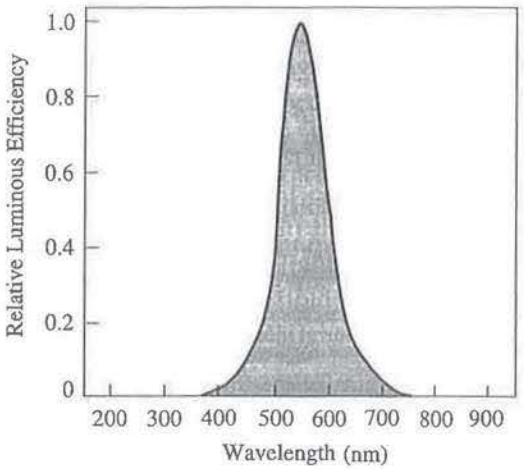


Figure 5.6 Approximate normalized spectral response of the human eye, also known as the photopic visual curve.

Vizio EX1016 Page 0124

5.4.4 Coupling Loss

In most practical devices the light output is coupled into a fiber. This process introduces a further *coupling loss*. Consider the geometry shown in Fig. 5.7. The acceptance angle θ_a of a step index fiber with core index n_{r1} and cladding index n_{r2} ($n_{r1} > n_{r2}$) is given by (Problem 5.3):

$$\theta_a = \sin^{-1}(n_{r1}^2 - n_{r2}^2)^{1/2} = \sin^{-1}(A_n) \tag{5.13}$$

where A_n is the numerical aperture of the fiber. It is assumed that all the light incident on the exposed end of the fiber core within the angle θ_a is coupled by internal reflection. Also, for a Lambertian distribution of the radiated intensity,

$$\Im(\theta) = \Im_o \cos\theta \tag{5.14}$$

Considering a device that is smaller than, and in close proximity to, the fiber core, and assuming a cylindrical symmetry, the *coupling efficiency* η_c can be expressed as

$$\eta_c = \frac{\int_o^{\theta_a} \Im(\theta) \sin\theta d\theta}{\int_o^{\pi/2} \Im(\theta) \sin\theta d\theta} \tag{5.15}$$

which, by substitution of Eq. 5.14 becomes

$$\eta_c = \frac{\int_o^{\theta_a} \Im_o \sin 2\theta d\theta}{\int_o^{\pi/2} \Im_o \sin 2\theta d\theta}$$

$$= \sin^2\theta_a. \tag{5.16}$$

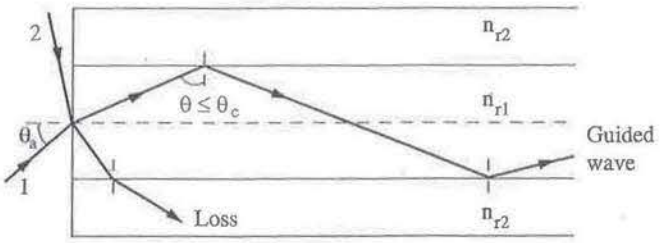


Figure 5.7 Geometry of light coupling into an optical fiber.

EXAMPLE 5.1

Objective. To determine the coupling characteristics of a silica optical fiber with core and cladding refractive indices of 1.48 and 1.50, respectively.

From Eq. 5.8, the critical angle θ_c is equal to 89.6°. From Eq. 5.13 $A_n = 0.244$. Therefore, the acceptance angle of the fiber, $\theta_a = 15.7°$.

Vizio EX1016 Page 0125

5.5 LIGHT OUTPUT FROM LED

We will now calculate the amount of light emitted by an LED due to forward-bias injection. The photons are emitted, with respect to Fig. 3.2(a), due to the recombination of an electron at energy $\mathcal{E}(\mathbf{k}'') = \mathcal{E}_2$ with a hole at energy $\mathcal{E}(\mathbf{k}') = \mathcal{E}_1$. We assume that $\mathbf{k}'' = \mathbf{k}' = \mathbf{k}$ and $(\mathcal{E}_2 - \mathcal{E}_1) = h\nu = \mathcal{E}$, the energy of the emitted photon. From Eqs. 3.28 and 3.29, we can write

$$k^2 = \frac{2m_r^*}{\hbar^2}(\mathcal{E} - \mathcal{E}_g) \tag{5.17}$$

where m_r^* is the reduced mass given by Eq. 3.36. Therefore, near the bottom of the conduction band

$$\mathcal{E}_2 = \mathcal{E}_g + \frac{m_r^*}{m_e^*}(\mathcal{E} - \mathcal{E}_g) \tag{5.18}$$

Emission of a photon involves the density of states in both the conduction and valence bands. Conservation of energy and momentum forces the photon to interact with electrons and holes of specific energy and momentum. This interaction is described by the *joint density of states*, $N_J(\mathcal{E})$. Equation 5.18 describes a one-to-one correspondence between $\mathcal{E}_2$ and the photon energy $\mathcal{E}$ and allows us to write the incremental relation

$$N(\mathcal{E}_2)d\mathcal{E}_2 = N_J(\mathcal{E})d\mathcal{E} \tag{5.19}$$

from which it follows that

$$N_J(\mathcal{E}) = \frac{m_r^*}{m_e^*}N(\mathcal{E}_2) \tag{5.20}$$

Substitution of Eq. 2.53 for $N(\mathcal{E}_2)$ leads to

$$N_J(\mathcal{E}) = \frac{(2m_r^*)^{3/2}}{2\pi^2\hbar^3}(\mathcal{E} - \mathcal{E}_g)^{1/2} \quad ((eV)^{-1}.cm^{-3}) \tag{5.21}$$

for parabolic electron-hole bands. The spontaneous emission rate $r_{sp}(\mathcal{E})$ and total spontaneous emission rate R_{sp} per unit volume, taking the selection rule into account, are defined by

$$r_{sp}(\mathcal{E}) = P_{em}N_J(\mathcal{E})f_n(\mathcal{E}_2)[1 - f_p(\mathcal{E}_1)] \quad (s^{-1}.(eV)^{-1}.cm^{-3})$$

$$R_{sp} = \int_{\mathcal{E}} r_{sp}(\mathcal{E})d\mathcal{E} \quad (s^{-1}.cm^{-3}) \tag{5.22a}$$

or in an alternate form[†]

$$R_{sp} = \int d\mathcal{E}\frac{n_r q^2 \mathcal{E}}{3\pi\epsilon_o m_o^2\hbar^2 c^3}\left\{\int \frac{1}{(2\pi)^3}d^3k p_{CV}^2 \delta(\mathcal{E}_2(\mathbf{k}) - \mathcal{E}_1(\mathbf{k}) - \mathcal{E})\right.$$

$$\left.\times f_n(\mathcal{E}_2)[1 - f_p(\mathcal{E}_1)]\right\} \tag{5.22b}$$

[†]J. Singh, *Physics of Semiconductors and Their Heterostructures*, McGraw-Hill, New York, 1993.

Vizio EX1016 Page 0126

where the integral over d^3k assures that all occupied electron and hole states are included and the δ-function assures that energy is conserved. The integration therefore gives the joint density of states, and Eqs. 5.22(a) and (b) are equivalent. The quantity P_{em} is the emission probability given by

$$P_{em} = \frac{n_r q^2 \mathcal{E} p_{CV}^2 (1 + u_\mathcal{E})}{3\pi \epsilon_o m_o^2 \hbar^2 c^3} \tag{5.23}$$

from Eqs. 3.42–3.44. P_{em} also includes the factor $n_\mathcal{E}$, which is the photon mode density, as defined in Appendix 4. It should also be noted that the emission probability has a factor $(1 + u_\mathcal{E})$ where $u_\mathcal{E}$ is the number of photons per mode, or the photon occupation, as defined in Appendix 4. In the case of spontaneous emission, where the photons created by recombination escape, $u_\mathcal{E}$ is very small and the factor is unity. For example, in GaAs, where $2p_{CV}^2/m_o = 23$ eV, and assuming $\mathcal{E} = 1.5$ eV, we get $P_{em} = 1.8 \times 10^9 s^{-1}$. Thus, the recombination lifetime $\tau_r = 1/P_{em} = 0.6$ ns. The Fermi functions in Eq. 5.22(a) and (b) ensure that a state in the conduction band is filled and a state in the valence band is empty. The distribution functions $f_n(\mathcal{E}_2)$ and $f_p(\mathcal{E}_1)$ are defined by Eqs. 3.1 and 3.2 and involve the electron and hole quasi-Fermi levels $\mathcal{E}_{fn}$ and $\mathcal{E}_{fp}$, respectively. For example, under thermal equilibrium conditions, the quasi-Fermi levels merge into a single Fermi level within the bandgap. The equilibrium distribution can then be used. Furthermore, if the Boltzmann approximation is valid, then $f_n(\mathcal{E}_2)[1 - f_p(\mathcal{E}_1)] \cong \exp\left(-\frac{\mathcal{E}}{k_B T}\right)$. Under these conditions,

$$r_{sp}(\mathcal{E}) = \frac{(2m_r^*)^{3/2}}{2\pi^2 \hbar^3 \tau_r} (\mathcal{E} - \mathcal{E}_g)^{1/2} e^{-\frac{\mathcal{E}_g}{k_B T}} e^{-\left(\frac{\mathcal{E} - \mathcal{E}_g}{k_B T}\right)} \tag{5.24}$$

Next consider the case of weak injection in an LED, such that the quasi-Fermi levels are still several $k_B T$ away from the bandedges and are within the bandgap. The distribution functions are then given by the exponential approximations of Eqs. 3.3 and 3.4. Under these conditions

$$r_{sp}(\mathcal{E}) = \frac{(2m_r^*)^{3/2}}{2\pi^2 \hbar^3 \tau_r} (\mathcal{E} - \mathcal{E}_g)^{1/2} e^{\left(\frac{\mathcal{E}_{fn} - \mathcal{E}_{fp} - \mathcal{E}_g}{k_B T}\right)} e^{\left(\frac{\mathcal{E} - \mathcal{E}_g}{k_B T}\right)} \tag{5.25}$$

A plot of this function is shown in Fig. 5.8. The total photon flux is obtained by

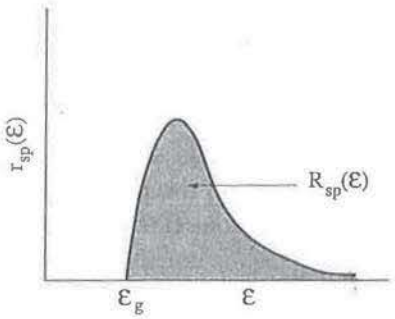


Figure 5.8 Spectral distribution of the injection electroluminescent rate. Also see Problem 5.11.

Vizio EX1016 Page 0127

integration of Eq. 5.25:

$$\Phi_o = V_l \int_o^\infty r_{sp}(\mathcal{E})d\mathcal{E}$$

$$= \frac{V_l}{\sqrt{2}\hbar^3\tau_r}\left(\frac{m_r^*}{\pi}\right)^{3/2}(k_BT)^{3/2}e^{\left(\frac{\mathcal{E}_{fn}-\mathcal{E}_{fp}-\mathcal{E}_g}{k_BT}\right)}\left(s^{-1}\right) \tag{5.26}$$

where V_l is the active region volume. In performing this integration P_{em}, a function of photon energy $\mathcal{E}$, has been replaced by $1/\tau_r$. This is because the effect of the value of $\mathcal{E}$ is small compared to the change in $(\mathcal{E} - \mathcal{E}_g)$. When the injection level is increased, the quasi-Fermi levels move into the bands and the probability function $f_n(\mathcal{E}_2)[1 - f_p(\mathcal{E}_1)]$ increases. As a consequence the light output increases.

The responsivity, $\mathcal{R}$, of an LED is defined as the ratio of the emitted optical power $P_o = \Phi_o h\nu$ to the injection current I. In other words,

$$\mathcal{R} = \frac{\Phi_o h\nu}{I} = \eta_o\frac{h\nu}{q}$$

$$= \frac{1.24\eta_o}{\lambda(\mu m)}\ (W/A) \tag{5.27}$$

EXAMPLE 5.2

Objective. To calculate the responsivity of an LED with $\eta_o = 3\%$ at $\lambda = 1\mu$m.

$$\mathcal{R} = 0.03 \times 1.24$$

$$= 37.2\,\mu W/mA$$

According to Eq. 5.27 the output power is proportional to drive current. In this range the responsivity has a constant value. As the injection current is increased, saturation effects take over, and the slope decreases.

5.6 LED STRUCTURES

The structures being discussed here are used in a variety of applications where both low and high radiance are required.

5.6.1 Heterojunction LED

The devices shown in Fig. 5.2 and discussed until now are simple homojunctions. There are two main problems with this structure, which limit the internal quantum efficiency. First, the surface states on the p-layer close to the light-emitting junction give rise to nonradiative recombination. If the surface is far removed from the junction, then the probability of reabsorption is enhanced. Also, the electrons injected from the n^+ region into the p-region as minority carriers diffuse away from the junction and

Vizio EX1016 Page 0128

gradually recombine with the majority carriers. Thus, the same number of photons are created in a greater volume of material, thereby reducing the internal quantum efficiency and enhancing reabsorption. Both these problems can be solved in a double heterostructure diode, whose schematic and band diagram are shown in Fig. 5.9. The advantages of this configuration are obvious. The GaAs layer into which electrons are injected from the top n^+ AlGaAs layer can be kept thin ($\sim$ 1000–2000 Å), and these electrons are contained within this thickness by the potential barrier of the p-GaAs/p-AlGaAs heterojunction. The top AlGaAs layer serves several important functions. First, the interface state density at this lattice-matched junction is orders of magnitude lower than the free surface state density, thereby drastically reducing nonradiative recombination. Second, being of a larger bandgap, it acts as a *window* layer in which the emitted photons, with energy close to the GaAs bandgap, are not reabsorbed. Finally, the injection efficiency is higher. These factors help to increase both the internal and external quantum efficiencies. Such heterostructure LEDs are usually realized by epitaxial growth and have also been made with InP-based materials such as InGaAsP/InP or InGaAs/InGaAsP/InP, which emit at longer wavelengths.

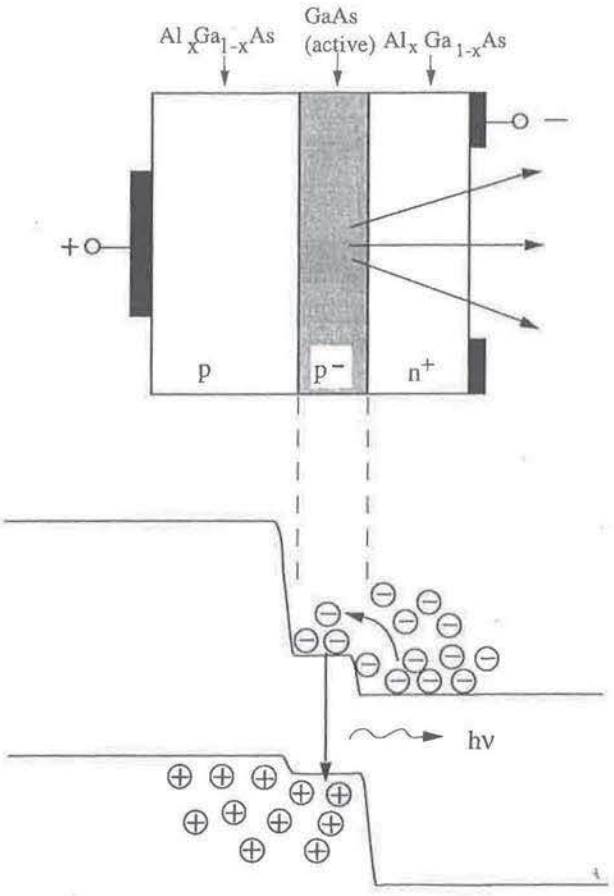


Figure 5.9 Device structure and band diagram of a forward-biased double heterostructure (DH) LED.

Vizio EX1016 Page 0129

5.6.2 Burrus Surface-Emitting LED

This structure was conceived and realized by Burrus and Dawson in order to obtain high radiance and efficient coupling of the emitted light to a fiber. It is essentially a double heterostructure, and is shown schematically in Fig. 5.10. The bottom p^+ GaAs and top n^+ GaAs layers are included for the realization of low-resistance ohmic contacts. However, to avoid reabsorption of the emitted radiation in the top n-GaAs layer, a deep well is etched to reach the top n-AlGaAs layer. This can be done conveniently by a selective etchant that etches GaAs but not AlGaAs. The well is also used to support the fiber, which is butt-coupled to the device and is held in place with an epoxy resin of appropriate refractive index to enhance the external power efficiency of the device. The thin SiO_2 layer at the back isolates the contact layer from the gold heat sink. Photons are generated in the thin p-GaAs region and emission from the top surface is ensured by the heterostructure and reflection from the back crystal face. Thus, the forward radiance of these devices is very high. The top n-GaAs contact layer ensures low contact resistance and thermal resistance, thereby allowing for high current densities and high radiation intensity. The fiber is properly aligned to optimize coupling of the emitted radiation. Nonetheless, there is some loss due to the Lambertian distribution of the radiation intensity and the coupling efficiency is typically 1%–2%.

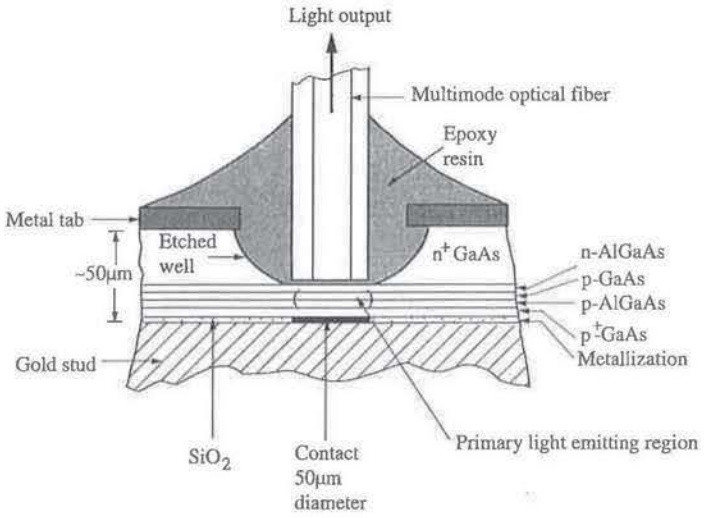


Figure 5.10 The structure of an A1GaAs DH surface-emitting LED (Burrus type) (from C. A. Burrus and B. I. Miller, *Opt. Commun.*, **4**, 307, 1971).

Several other types of lens coupling are used with surface-emitting heterostructures to improve the coupling efficiency. These are schematically shown in Figs. 5.11 (a)–(c). Figure 5.11(a) shows the end of a fiber that is polished into a spherical lens. This collects and collimates the divergent radiation from the LED. The same purpose is served by a glass microlens inserted between the device and the fiber, as shown in Fig. 5.11(b). An integrated lens structure is obtained by etching and polishing the top surface layer of the LED into a spherical lens, as shown in Fig. 5.11(c). This procedure does away with a lot of cumbersome fixing and alignment, but is difficult to implement. It is important to realize that in spite of the problems associated with large

Vizio EX1016 Page 0130

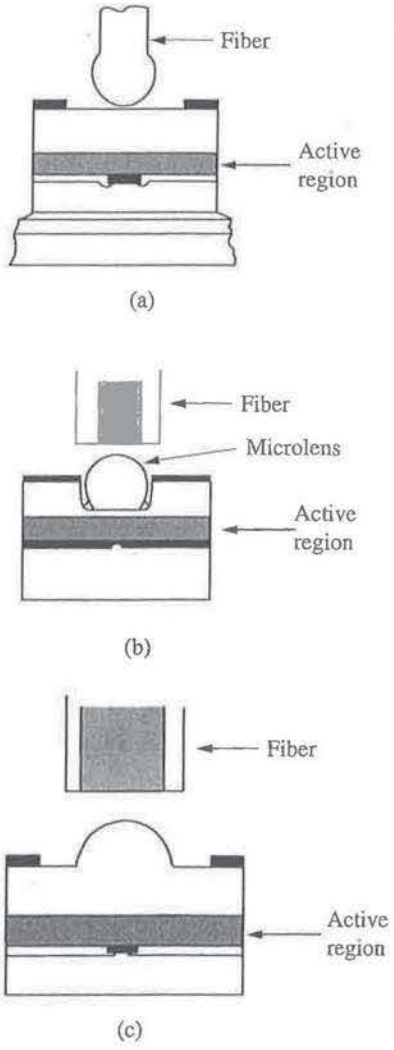


Figure 5.11 Different schemes of lens-coupled LEDs: (a) spherical-ended fiber-coupled device, (b) microlens coupling, and (c) integrated semiconductor lens structure.

divergence and coupling of the radiant beam, surface-emitting structures will remain important not only for large-volume, low-cost applications, but also for short-distance chip-to-chip communication where surface-emitting sources are essential. For optical computing applications also, surface-emitting sources may play an important role. With the different techniques of lens coupling discussed above, coupling efficiencies of 5%–15% have been achieved.

5.6.3 Guided Wave or Edge-Emitting LED

The surface-emitting LEDs just described are useful for common and less demanding applications. However, in optical communication, where tight coupling of the emitted light to a fiber or waveguide is required, a more collimated light emission is desirable. The basic edge-emitting LED is shown in Fig. 5.12. Generally for superradiant LEDs, a stripe geometry is incorporated, as in a junction laser, to improve the injection efficiency. The active layer is usually lightly doped or undoped,

Vizio EX1016 Page 0131

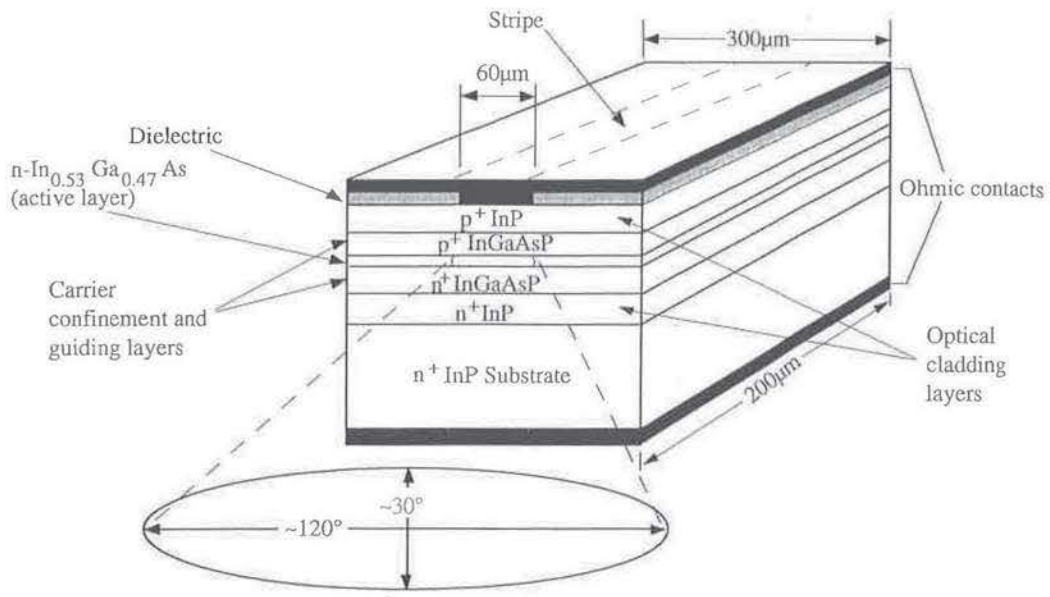


Figure 5.12 Schematic illustration of the structure of a stripe geometry DH InGaAs/InGaAs/InP edge-emitting LED.

and a very large population of carriers for recombination is created in this region by forward-bias injection. The two InGaAsP layers on both sides serve as carrier confinement layers, on the outer sides of which are doped InP layers. These serve as the *cladding* layers, and the region in between forms an optical waveguide. Thus, the wide bandgap layers serve the dual purpose of optical and carrier confinement. The photons are generated in the very thin active region and spread into the guiding layers, without reabsorption, because of their larger bandgaps. It is important to realize that the edge-emitting structure allows a thin active region without sacrificing the radiative recombination efficiency, in comparison with a homojunction surface-emitting LED. The stripe geometry, made by selective metallization on the top surface through a window opened in an SiO_2 layer, allows higher carrier injection densities for the same drive current. In other words, the power conversion efficiency is improved.

It is clear that the edge-emitting LED is very similar to a junction laser. However, it does not have a feedback cavity and is still an incoherent light source. Most of the light is made to come out of one edge of the structure. This is ensured by cleaving and putting a reflective coating at the nonemitting end and putting an antireflective coating at the emitting end. The waveguide vastly reduces the divergence of the emitted radiation. In the direction perpendicular to the plane of the layers, the half-power width is approximately 30°. However, in the plane of the layers the output is still Lambertian and the half-power width is approximately 120°. Optical confinement in the transverse direction can be created by selective etching and regrowth of a higher bandgap material. This technique is more common in laser fabrication and will be discussed in Chapter 7. The waveguiding and reduced beam divergence allows more

Vizio EX1016 Page 0132

efficient coupling of the radiated beam into fibers. The larger operating current density in a smaller structure can cause heat-sinking problems. However, it is anticipated that edge-emitting LEDs with InP-based materials will play an important role as high-power sources in short-distance optical communication, where dispersion in the fibers will not be too severe.

5.6.4 Drive Circuitry

The LED is operated under forward-bias injection conditions. The amount of forward bias depends on the forward cut-in voltage, which, in turn, depends on the bandgap of the semiconductor. Typical biases are therefore between 1 and 2V. Under operating conditions the current drawn by the LED is $\sim$ 100 mA. An LED can be operated by a direct current (dc) or alternating current (ac) drive circuit. A simple dc drive circuit is shown in Fig. 5.13(a), where the current through the diode provided by the voltage source V is limited by the series resistance R_S. Under operating conditions, the voltage drop across the LED is V_L, the operating voltage of the device. If the operating current is I_L, then the circuit can be described by

$$V - V_L = I_L R_S \tag{5.28}$$

It may often be desirable or necessary to place an LED in an ac circuit. There can be several kinds of circuits, depending on the application, and a common form is shown in Fig. 5.13(b). Here a diode with reversed polarity is placed in parallel with the LED to prevent it from high reverse bias voltages. A capacitor and current-limiting resistor are in series with the two diodes. In this circuit, for current of both polarities the voltage drop across the diode is restricted to a small value. Then the current flowing through the circuit is approximately the supply voltage divided by the impedances due to the resistance and capacitance. The series resistor also prevents the LED from high turn-on transients.

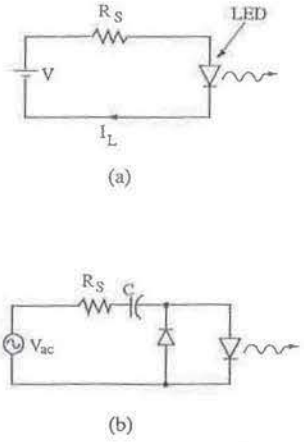


Figure 5.13 Simple (a) dc and (b) ac drive circuits for light-emitting diodes.

For most communication applications a high-frequency signal is impressed upon the cw output of the light source. This is usually done by internal or external modulation. In external modulation, a modulator is placed in the path of the cw signal

Vizio EX1016 Page 0133

from the source, and the signal coming in as an electrical input varies the transmission of the modulator at a high speed. The properties of such optical modulators will be discussed in Chapter 11. Here we will discuss internal modulation, which is usually achieved by varying the drive current of the LED with the signal. Hence it is also called current modulation. In Fig. 5.14 a simple modulation circuit is shown, in which the signal is applied to the base circuit of an n-p-n transistor connected in series with the LED and a current-limiting resistor. It is advantageous to use the transistor because of the gain it provides. The variation of the base current with the signal varies the current flow and thus the forward injection of the LED. A modulated light output is obtained from the diode. The circuit is usually operated such that both the LED and the transistor operate in their linear output regimes. Under these conditions changes in the current flowing through the LED are directly proportional to input signal voltage in the base circuit.

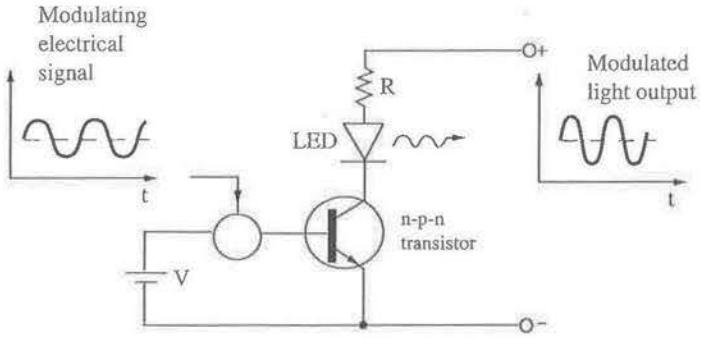


Figure 5.14 Current (internal) modulation circuit for light-emitting diode.

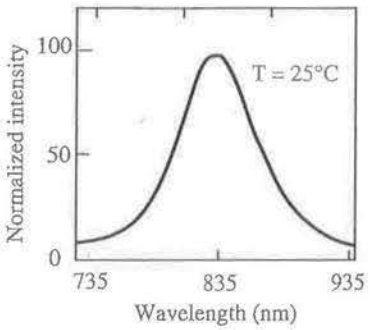
5.7 DEVICE PERFORMANCE CHARACTERISTICS

Once an LED is fabricated, measurements are made to determine its electrical and optical characteristics. For commercial diodes, these characteristics are sometimes provided by the manufacturer. A typical set of performance curves for stripe-geometry AlGaAs and InGaAsP double heterojunction LEDs are shown in Fig. 5.15.

5.7.1 Spectral Response

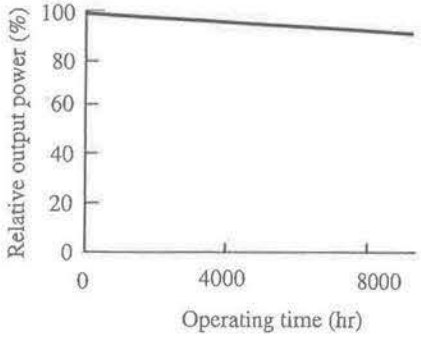
The spectral response of the AlGaAs LED is shown in Fig. 5.15(a) and is the plot of the relative output intensity against the wavelength in nanometers. The peak intensity is approximately at 835 nm. The spectral linewidth, measured in units of wavelength between the half maximum intensity points—termed the full width at half maximum (FWHM)—is around 40 nm. This is typical of LEDs operating around 0.8 μm. In LEDs emitting at longer wavelengths (1.1–1.6 μm) the material used is typically InGaAsP/InP, and the linewidth of the spectral response is broadened by alloy scattering. The linewidth is also shifted and broadened by heavy doping in the active region. Therefore, in surface-emitting LEDs where the active region is more heavily doped, the emission wavelength moves to longer wavelengths (lower

Vizio EX1016 Page 0134



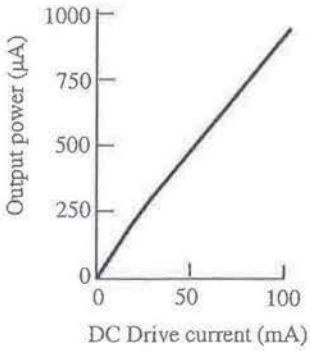
Typical spectral output

(a) AlGaAs



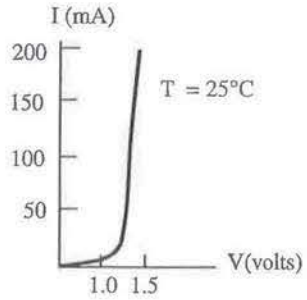
Output power vs. typical operating time for an edge-emitting device

(b) AlGaAs



Output power vs forward current

(c) AlGaAs



Forward I-V characteristics

(d) InGaAsP (1300 nm)

Figure 5.15 Typical performance curves of A1GaAs and InGaAsP light-emitting diodes (courtesy of D. Pooladdej, Laser Diode Inc., New Brunswick, NJ).

energies) and the spectral response curve is broadened. Fluctuations in temperature also affect the response characteristics. The bandgap of most III–V semiconductors decrease at a rate of 5 meV/degree increase of temperature. The linewidth of the spectrum also increases with increase of temperature due to a larger energy spread of the carrier distribution and phonon scattering. Therefore, for applications where the emission wavelength and linewidth need to remain invariant, heat sinking and cooling arrangements are incorporated with the device.

5.7.2 Output Power-Time Characteristics

The lifetime of an electroluminescent device or a source for communication applications is an important parameter. Therefore, curves such as the one shown in

Vizio EX1016 Page 0135

Fig. 5.15(b) are important. This curve, for the AlGaAs LED, shows a power loss of a few percent after operating for 8,000 hours.

5.7.3 Light(Power)-Current Characteristics

These are perhaps the single most important characteristics of an LED. The characteristics of an edge-emitting LED is shown in Fig. 5.15(c). The linearity of the light-current characteristics is important for modulation in analog transmission. Some LEDs can exhibit nonlinearity depending on material properties and device configuration. It is also important to note that edge emitters radiate less optical power into air than surface emitters, although the radiation intensity at the emitting face can be very high in the former. This is because interfacial recombination and reabsorption can play a more dominant role in edge-emitting LEDs. However, because of the higher radiation intensity from them, coupling into fibers with smaller numerical apertures is more efficient. The light-current characteristics shown in Fig. 5.16 illustrate the above points.

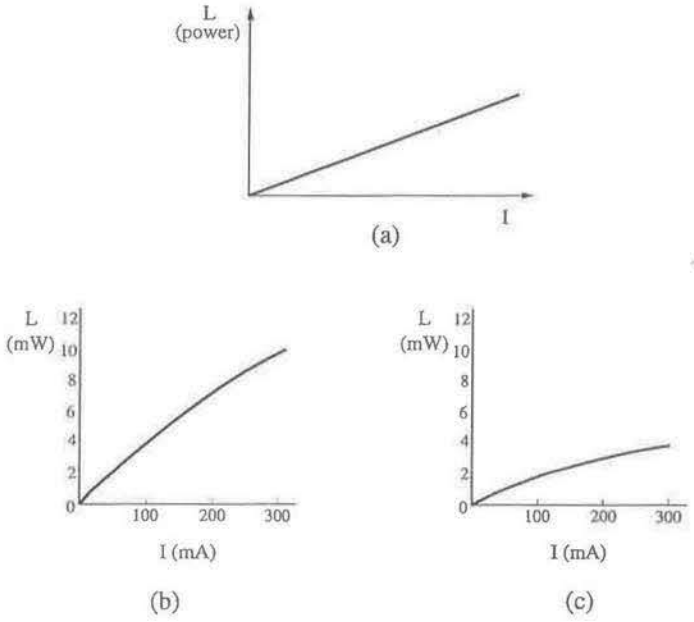


Figure 5.16 (a) Schematic illustration of linear light-current characteristics of LED, (b) an A1GaAs surface emitter with a 50μm diameter dot contact, and (c) and A1GaAs edge emitter with a 65 μm wide stripe and 100 μm length. (D. Botez and M. Ettenburg, *IEEE Trans. Electron Devices*, **ED-26(3)**, 1230, ©1979 IEEE).

5.7.4 Diode Current-Voltage Characteristics

These describe the electric operating characteristics of the device. Typical current-voltage characteristics are shown in Fig. 5.15(d). For efficient operation it is important to know the forward threshold voltage, the reverse breakdown voltage and the reverse leakage current at breakdown. It may be noted that the reverse breakdown voltage (5–15V) in LEDs is not very high, since the junction is usually formed of heavily doped materials.

Vizio EX1016 Page 0136

5.8 FREQUENCY RESPONSE AND MODULATION BANDWIDTH

The efficient utilization of a communication system calls for minimizing the transmission time. In other words, the maximum amount of information needs to be transmitted in the least possible time. Rapid transmission of signals can be achieved if the signal changes rapidly with time. A common and convenient measure of signal speed is its bandwidth, which is the width of the signal spectrum in frequency domain. High-speed communication calls for wideband signals to represent the information and wideband systems to accommodate the signals. In optical communication a high-frequency lightwave is modulated by the wideband information signal, thereby reducing the *fractional* bandwidth and simplifying equipment design. For everyday commercial applications and in particular as display devices, the response time of an LED to a drive current pulse is not a limiting factor. The response time of most commercial LEDs is $\leq 1\mu s$ and this is more than adequate. However, for communication applications in which high-frequency internal modulation of the LED as a source is desirable, response times of the order of a nanosecond or less are desirable. Therefore it is important to understand the processes that limit the response time and to explore techniques to reduce it.

There are two main factors, extrinsic and intrinsic, that limit the response time of a LED. The extrinsic factor is the junction capacitance of the diode. This capacitance, together with the resistance in the diode circuit, gives rise to a characteristic RC time constant. The capacitance can be decreased by decreasing the diode area and increasing the reverse bias. The latter is related to the doping. The intrinsic limitation to the response time arises from the charge storage and diffusion capacitance of a p-n junction under forward bias. When the applied bias to the diode is modulated at high speed, the stored charge cannot respond. The excess injected charge during forward bias is removed by diffusion and recombination with minority carriers. Therefore, the recombination lifetime in the junction material plays a major role in determining the modulation bandwidth. As shown in Appendix 6, the frequency response is given by

$$r(\omega) = \frac{1}{[1 + \omega^2\tau^2]^{1/2}} \tag{5.29}$$

where $r(\omega)$ is the response (power output) at angular frequency ω and τ is the minority carrier lifetime in the injected region. It is evident that for good high-frequency response the recombination lifetime τ should be made small. As we have seen in Chapter 3, the value of τ is also related to doping and injection level. For low-level injection into a p-type material, the value of τ ($= \tau_r$ if nonradiative recombination is negligible) is given by Eq. 3.23. Thus, τ_r decreases with increasing doping and the best value that can be obtained is approximately 1 ns. The limit is set by the breakdown and leakage current of the diode and the solubility limit of the dopant specie. An alternate approach is to operate in the bi-molecular recombination regime in lightly doped material. For example, for high-level injection into a thin active region, the value of τ_r is expressed by Eq. 3.22. Under these conditions the injected carrier density Δn into the active region is much greater than the equilibrium value. A

Vizio EX1016 Page 0137

simple expression for τ_r as a function of the injection current density J can be derived as follows. If J is entirely a recombination current, then the number of recombinations per unit volume per second in an active region of width d is J/qd. This must be equal to $\Delta n/\tau_r$. Therefore, by using Eq. 3.22,

$$\tau_r = \left(\frac{qd}{JB_r}\right)^{1/2} \tag{5.30}$$

Thus, τ_r may be reduced by reducing d and increasing J. However, a very high value of J may lead to heat-sinking problems and distortion of the modulated signal. It should also be remembered that in conventional surface-emitting LEDs with thin active regions, the recombination lifetime is limited by the surface recombination rate. Edge-emitting LEDs with a thin undoped active region, operating in the bi-molecular recombination region promise the best high-frequency response. It is important to note that all the material and device parameters that are to be adjusted for realizing the best high-frequency response are interrelated.

An alternate way of expressing the response of an LED to a high-frequency current drive is the *modulation bandwidth*. This is defined as the frequency at which the output optical power received at a detector is reduced by 3 dB with respect to power transmitted. As the modulation frequency increases, the LED does not emit the total amount of light, since all the injected carriers cannot recombine. If we assume a linear light(power)-current relationship for both the LED and the detector, then we may define the bandwidth as follows:

$$\Delta f(dB) = 10\log_{10}\frac{P_{out}(f)}{P_{out}(dc)}$$

$$= 10\log_{10}\frac{I_{out}(f)}{I_{out}(dc)} \tag{5.31}$$

where I_{out} is the current at the detector. The 3 dB point occurs when the ratio of the currents is equal to 1/2. The bandwidth is schematically shown in Fig. 5.17. It

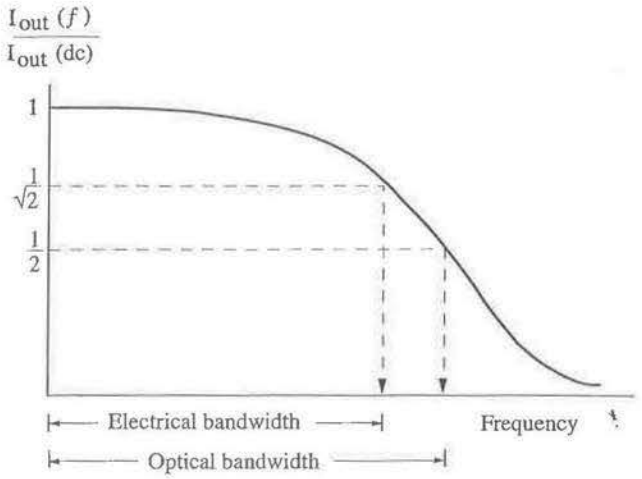


Figure 5.17 Frequency response of a light-emitting diode.

Vizio EX1016 Page 0138

is sometimes also possible to define an electrical 3 dB bandwidth as the ratio of the electrical powers. The 3 dB bandwidth occurs when $[I_{out}\ (f)/I_{out}\ (dc)]^2 = 1/2$ or $I_{out}\ (f)/I_{out}\ (dc) = 0.707$. Therefore, the electrical bandwidth is smaller than the optical bandwidth.

In conclusion, the modulation bandwidth of an LED depends on the device configuration (surface- or edge-emitting) doping level in the active layer, the lifetime of the injected carriers, and the parasitic capacitance and resistance in the circuit. Finally, it may be remembered that in LEDs the photons are generated by spontaneous recombination, and the radiative lifetime under these conditions is always longer than that under stimulated emission, as in a laser. Therefore, lasers have higher modulation bandwidths. Nevertheless, bandwidths over 1 GHz can now be obtained in practical LEDs. The intensity modulation by variation of the drive current can be applied with both analog and digital signals. The former requires a larger signal-to-noise ratio and is therefore more suited for short-distance applications. Digital modulation has better noise immunity and is therefore more suited for long-distance optical-fiber communication.

5.9 MANUFACTURING PROCESS AND APPLICATIONS

The LED represents the simplest and most robust of the electroluminescent devices. The cross section of a standard red-emitting LED is shown in Fig. 5.18. The device consists of layers of GaAsP grown on a GaAs substrate. Some LEDs (orange-emitting, for example) are grown on GaP substrates. The substrate material is the same as that used for other electronic and optoelectronic devices and is grown in ingots by the Horizontal-Bridgemann or Czochralski techniques. For P-containing compounds, the resulting pressure at the melting point is very high (e.g., 40 atmospheres for GaP). Also, the B_2O_3 used for encapsulation introduces B impurities in the single crystal. For high-quality substrates the defect density varies in the range 10^2–10^5 cm^{-2}.

The epitaxial layers are grown by LPE or VPE. For red LEDs, the n-type layer is doped with Te and for the orange LED the dopant is S. For p-type layers, Zn is the best dopant. NH_3 is used for creating N isoelectronic traps. If Zn-O pairs are needed

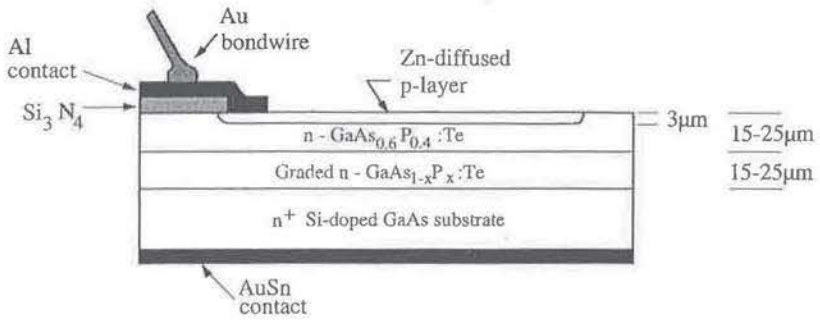


Figure 5.18 Schematic of a red-emitting $GaAs_{0.6}P_{0.4}$ LED with typical dimensions.

Vizio EX1016 Page 0139

to create isoelectronic traps, then Ga_2O_3 is passed through the material, which has already been doped with Zn. GaAsP grown on GaAs substrates may produce misfit dislocations, and therefore graded composition layers are sometimes incorporated between the substrate and the active region. Also, as shown in Fig. 5.18, an n-type substrate and a top p-type Zn-diffused layer are the most common configuration due to the relative ease with which Zn can be incorporated by diffusion. This diffusion is done selectively through an Si_3N_4 mask to form the p-n junction. Contact layers are usually formed of Al and AuSn on the p and n sides, respectively. Alternately, AuBe and AuGe are also used on the p and n sides. Contact to the bias source is made via gold bonds, as shown. To make the light scatter and radiate uniformly from the junction, the metallization on it is made such that the current distribution over the active region is uniform.

The completed LED is encapsulated in a dome, as shown in Fig. 5.5(b), for display and counting applications. For visible LEDs the bottom surface can cause spurious reflections, and therefore a special coating is applied. The most common application of visible LEDs are as lamps and displays. The choice of an LED as a lamp is decided by the color required for the particular application and the requirements of luminous intensity, uniform visibility from all directions, and a contrasting background. LEDs have proved to be very useful for pocket calculators, where red-emitting LEDs are generally used. For this application, several individual chips are required, and each of them should be capable of forming all the numerals. To reduce the size of each chip and yet keep the numerals large, a lens or reflector is used, as shown in Fig. 5.11. In addition, a narrowband filter is placed in the path of the radiated light to improve the contrast ratio. In the fabrication of these devices cost is an important factor.

In applications such as pocket calculators, parallel connections of many small-area LEDs are used to display alphanumerics. Two schemes are generally used. In the first scheme, shown in Fig. 5.19(a), a seven-segment system is used, where each segment can consist of several LEDs. A digital logic control system is used to feed current to the desired segment. The negative terminal of all the LEDs is connected to a common ground terminal. In the second scheme, shown in Fig. 5.19(b), a matrix of LEDs is used, and logic control directs current to specific LEDs to produce a specific display. In both schemes, failure of one or a small number of LEDs will not ruin

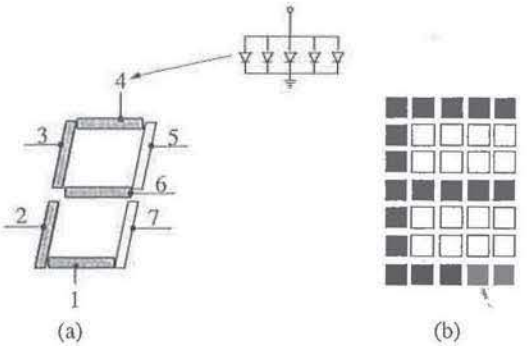


Figure 5.19 (a) Segment architecture and (b) matrix architecture of LEDs for alphanumeric display. In the first, each segment consists of multiple LEDs.

Vizio EX1016 Page 0140

the display. There are competing technologies, such as liquid crystal displays, which have also found wide applications in computing machines and watches.

For sensing and communication applications, infrared LEDs are generally used. For the former GaAs is the most common material. InP-based devices are being used for the latter.

5.10 DEFECTS AND DEVICE RELIABILITY

One of the foremost requirement of LEDs, both for common and specialized applications, is reliability over a long lifetime. Unlike junction lasers, which operate under much larger drive currents, LEDs generally do not suffer catastrophic degradation. But their performance characteristics are known to change with time, and they exhibit gradual degradation. There are two classes of defects that are known to degrade the performance of these devices. The first is dislocations, which may propagate from the substrate during epitaxy into the active region. Currently the dislocation in doped substrates is ~ 10–10^2 cm^{-2} and the density in the active layer can be reduced by small geometry and incorporation of appropriate buffer layers, that act as dislocation filters. The second is deep-level defects in the forbidden energy gap. Deep levels can arise from point defects such as vacancies, interstitial defects, impurity atoms, or complexes formed from combinations of these.

Dislocations in the active area give rise to *dark line defects* and *dark spots*, which cause a more rapid degradation of the device. In essence, the dislocations form absorbing regions. The degradation is dependent on the operating or device temperature, injection current density, and impurity concentration in the active region.

Deep levels, as we have already seen, reduce the radiative efficiency of the LED. In addition, point defects and impurity migration into the active region may be created or enhanced by the recombination process. These give rise to a slow or long-term degradation.[†] The degradation rate is related to the activation energy $\mathcal{E}_d$ of the degradation process by an Arrhenius equation:

$$\gamma_d = C \exp(-\mathcal{E}_d / k_B T) \tag{5.32}$$

where γ_d is the degradation rate (s^{-1}) and C is a proportionality constant. The activation energy $\mathcal{E}_d$ represents an energy barrier to impurity migration or defect generation. Consequently, the larger is the value of $\mathcal{E}_d$, smaller is the degradation rate. $\mathcal{E}_d$ is related to materials parameters, epitaxial techniques, and the device processing and geometry. The radiant output power as a function of time, $P(t)$, may then be expressed as

$$P(t) = P(0) \exp(-\gamma_d t). \tag{5.33}$$

Based on these equations and values of $\mathcal{E}_d$ in the range of 0.6–1.0 eV, operation lifetimes of $\sim 10^6$–10^9 hours of continuous wave (cw) operation can be expected,

[†]S. Yamakoshi et al., *IEEE J. Quantum Electronics, QE-17*(2), 167–173, 1981.

Vizio EX1016 Page 0141

depending on the material system. The realization of these lifetimes depends on materials synthesis and device fabrication, and operation with very efficient heat sinking. Without the latter, the junction temperature can quickly become much higher than the ambient temperature.

EXAMPLE 5.3

Objective. A GaAs/AlGaAs LED has an activation barrier of energy $\mathcal{E}_d = 0.6$ eV to long-term degradation. If the value of C is $2 \times 10^2 hr^{-1}$, to find the time after which the radiant output power will fall to half its initial value for room temperature operation.

From Eq. 5.32, $\gamma_d = 7.6 \times 10^{-9} hr^{-1}$. Substituting this value in Eq. 5.33,

$$\frac{P(t)}{P(0)} = 0.5 = \exp(-7.6 \times 10^{-9} t) \tag{5.34}$$

from which $t = 9 \times 10^7 hr$.

It must be remembered, however, that there are other modes of degradation, which can further reduce the operation lifetime.

PROBLEMS

5.1 Discuss the different processes and their efficiencies that combine to produce the external conversion efficiency of an LED. Which one(s) are the most significant? Give reasons for your answer.

5.2 Derive Eq. 5.9.

5.3 Derive Eq. 5.13.

5.4 Without knowing too much about injection lasers at this point, briefly discuss some of the advantages and disadvantages of LEDs in comparison with lasers for long-haul optical-fiber communication.

5.5 A GaAs p-n junction LED has $N_A = N_D = 10^{18}$ cm^{-3}. Light emission predominantly results from electron injection and its recombination in the p-region. Calculate the radiative lifetime. A deep-level recombination center with a lifetime of 5×10^{-9} sec is also present in the active region. Calculate the electron lifetime in the p-region and the radiative recombination efficiency. Assuming that the hole lifetime is equal to the electron lifetime, calculate the electron injection efficiency η_{in} if $D_e = 120$ cm^2/s and $D_h = 0.01 D_e$. What will be the value of η_{in} under a forward bias of 1 V if the recombination current I_{rec} due to the deep level is also taken in account?
[Hint: Taking I_{rec} into account, $\eta_{in} = \frac{I_e}{I_e + I_h + I_{rec}}$. Also, use a value of $B_r = 7.2 \times 10^{-10}$ cm$^3 . s^{-1}$.]

Vizio EX1016 Page 0142

5.6 For the LED of Problem 5.5 and under identical operating conditions (neglect I_{rec}) determine the value of the extraction efficiency η_e, if the overall device efficiency is one percent.

5.7 The Zn-O transition is used for light emission at 6900 Å in a GaP LED. Zinc impurities give rise to an acceptor level with $\mathcal{E}_A = 40$ meV and O atoms give rise to a deep donor with $\mathcal{E}_D = 0.8$ eV. Estimate the average spacing between the donor and acceptor atoms. From this result can you estimate the approximate doping levels of Zn and O?

5.8 List the parameters that need to be optimized and the techniques to maximize the external quantum efficiency of an LED.

5.9 Show that the numerical aperture of a step-index fiber is approximately given by the relation

$$A_n = \frac{1}{n_{r0}}\sqrt{2n_{r1}\Delta n_r}$$

where n_{r0} is the index of the medium from which light is coupled, n_{r1} is the index of the fiber core, and Δn_r is the difference of the indices of the fiber core and cladding (n_{r2}). A GaAs LED is coupled through air ($n_r = 1$) to a step-index glass fiber ($n_{r1} = 1.55$). Is the acceptance angle for rays still inside the LED affected by using a dielectric medium ($n_r \cong 1.55$) instead of air? What is this angle for $\Delta n_r = 0.025$?

5.10 A planar GaAs LED exhibits an external power efficiency of 2% when driven with a current of 40 mA. The voltage developed across its terminals is 2.2 V. Estimate the optical power generated within the device assuming transmission into air.

5.11 Using Eq. 5.25, determine the energy at which the spectral distribution of $r_{sp}(\mathcal{E})$ shown in Fig. 5.8 attains a peak value. Also calculate its full width at half-maximum (FWHM).

5.12 Describe the LED structures commonly used for optical-fiber communication and discuss their merits and disadvantages.

5.13 Light-emitting diodes are made of Si and GaAs. The device dimensions and doping levels are identical. Which one do you expect to exhibit a larger modulation bandwidth and why?

5.14 With reference to Eq. 5.29, explain why edge-emitting LEDs promise high-modulation bandwidths.

5.15 Design a simple dc biasing circuit with a 9V battery for an LED that has a maximum forward current of 2 mA at 2 V.

READING LIST

KRESSEL, H., and BUTLER, J. K. *Semiconductor Lasers and Heterojunction LEDs*. Academic Press, New York, 1977.

Vizio EX1016 Page 0143

SEIPPEL, R. G. *Optoelectronics for Technology and Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

SZE, S. M. *Physics of Semiconductor Devices*, 2nd ed. Wiley, New York, 1981.

WILLIAMS, E. W., and HALL, R. *Luminescence and Light Emitting Diode*. Pergamon Press, Oxford, 1978.

WILSON, J., and HAWKES, J. F. B. *Optoelectronics: An Introduction*, 2nd ed. Prentice Hall International, United Kingdom, 1983.

Vizio EX1016 Page 0144