

Optical Properties of Solids

Mark Fox



oxford master series in condensed matter physics
mark fox

14

1	H	Hydrogen 1.00794 1s	IIA
3	Li	Lithium 6.941 1s ² 2s	4 Be
11	Na	Sodium 22.98977 [Ne]3s	12 Mg
19	K	Potassium 39.0983 [Ar]4s	20 Ca
37	Rb	Rubidium 85.4678 [Kr]5s	38 Sr
55	Cs	Cesium 132.90545 [Xe]6s	56 Ba
87	Fr	Francium (223) [Rn]7s	88 Ra

For atomic weight, a number in brackets indicates the mass number of the most stable isotope.

三

2 He		3 Li		4 Be		5 B		6 C		7 N		8 O		9 F		10 Ne	
He	Helium 4.00260 1s ²																
Atomic number	5	Boron 10.811 1s ² 2s ² 2p	Carbon 12.0107 1s ² 2s ² 2p	Nitrogen 14.00374 1s ² 2s ² 2p ³	Oxygen 15.9944 1s ² 2s ² 2p ⁴	Fluorine 18.9980 1s ² 2s ² 2p ⁵	Neon 20.1197 1s ² 2s ² 2p ⁶										
Name	Boron	Carbon	Nitrogen	Oxygen	Fluorine	Neon											
Atomic weight	10.811	12.0107	14.00374	15.9944	18.9980	20.1197											
Electronic structure	[He]2s ² 2p	[He]2s ² 2p	[He]2s ² 2p ³	[He]2s ² 2p ⁴	[He]2s ² 2p ⁵	[He]2s ² 2p ⁶											
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl		18Ar		19Kr		20Xe	
11B		14V		13Al		15P		16S		17Cl							

For atomic weight, a number in brackets indicates the mass number of the most stable isotope.

Actinides +

The Oxford Master Series in Condensed Matter Physics is designed for final year undergraduate and beginning graduate students in physics and related disciplines. It has been driven by a perceived gap in the literature today. While basic undergraduate condensed matter physics texts often show little or no connection with the huge explosion of research in condensed matter physics over the last two decades, more advanced and specialized texts tend to be rather daunting for students. In this series, all topics and their consequences are treated at a simple level, while pointers to recent developments are provided at various stages. The emphasis is on clear physical principles of symmetry, quantum mechanics, and electromagnetism which underlie the whole field. At the same time, the subjects are related to real measurements and to the experimental techniques and devices currently used by physicists in academe and industry.

Books in this series are written as course books, and include ample tutorial material, examples, illustrations, revision points, and problem sets. They can likewise be used as preparation for students starting a doctorate in condensed matter physics and related fields (e.g. in the fields of semiconductor devices, opto-electronic devices, or magnetic materials), or for recent graduates starting research in one of these fields in industry.

M. T. Dove: *Structure and dynamics*

J. Singleton: *Band theory and electronic properties of solids*

A. M. Fox: *Optical properties of solids*

S. J. Blundell: *Magnetism in condensed matter*

J. F. Annett: *Superconductivity*

R. A. L. Jones: *Soft condensed matter*

Optical Properties of Solids

MARK FOX

*Department of Physics and Astronomy
University of Sheffield*

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Athens Auckland Bangkok Bogotá Buenos Aires Cape Town
Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul Karachi
Kolkata Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
Paris São Paulo Shanghai Singapore Taipei Tokyo Toronto Warsaw

with associated companies in Berlin Ibadan

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Oxford University Press, 2001

The moral rights of the author have been asserted

Database right Oxford University Press (maker)

First published 2001

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose this same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloguing in Publication Data
Fox, Mark (Anthony Mark)

Optical properties of solids / Mark Fox.

p. cm. — (Oxford master series in condensed matter physics)

Includes index.

1. Solids—Optical properties. I. Title. II. Series.

QC176.8 O6 F69 2001 530.4'12—dc21 2001036967

ISBN 0 19 850613 9 (hardback)

ISBN 0 19 850612 0 (paperback)

10 9 8 7 6 5 4 3 2 1

Typeset using the author's L^AT_EX files by HK Typesetting Ltd, London
Printed in Great Britain
on acid-free paper by The Bath Press Ltd, Avon

Preface

This book is about the way light interacts with solids. The beautiful colours of gemstones have been valued in all societies, and metals have been used for making mirrors for thousands of years. However, the scientific explanations for these phenomena have only been given in relatively recent times. Nowadays, we build on this understanding and make use of rubies and sapphires in high power solid state lasers. Meanwhile, the arrival of inorganic and organic semiconductors has created the modern optoelectronics industry. The onward march of science and technology therefore keeps this perennial subject alive and active.

The book is designed for final year undergraduates and first year graduate students in physics. At the same time, I hope that some of the topics will be of interest to students and researchers of other disciplines such as engineering or materials science. It evolved from a final year undergraduate course in condensed matter physics given as part of the Master of Physics degree at Oxford University. In preparing the course I became aware that the discussion of optical phenomena in most of the general solid state texts was relatively brief. My aim in writing was therefore to supplement the standard texts and to introduce new subjects that have come to the fore in the last 10–20 years.

Practically all textbooks on this subject are built around a number of core topics such as interband transitions, excitons, free electron reflectivity, and phonon polaritons. This book is no exception. These core topics form the backbone for our understanding of the optical physics, and pave the way for the introduction of more modern topics. Much of this core material is well covered in the standard texts, but it can still benefit from the inclusion of more recent experimental data. This is made possible through the ever-improving purity of optical materials and the now widespread use of laser spectroscopy.

The overall plan of the subject material is summarized in Fig. 1. The flow diagram shows that some of the chapters can be read more or less independently of the others, on the assumption that the introductory material in Chapters 1 and 2 has been fully assimilated. I say ‘more or less’ here because it does not really make sense, for example, to try to understand nonlinear optics without a firm grasp of linear optics. The rest of the chapters have been arranged into groups, with their order following a certain logical progression. For example, knowledge of interband absorption is required to understand quantum wells, and is also needed to explain certain details in the reflectivity spectra of metals. Similarly, molecular materials provide an intuitive introduction to the concept of configuration diagrams, which are required for the understanding of colour centres and luminescent impurities.

The inclusion of recent developments in the subject has been one of the main priorities motivating this work. The chapters on semiconductor quantum

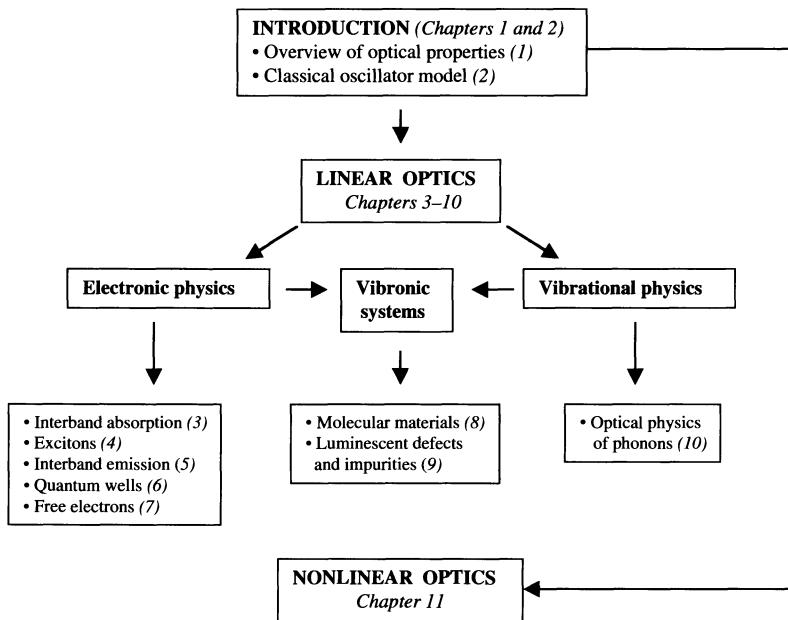


Fig. 1 Scheme of the subject topics covered in this book. The numbers in brackets refer to the chapters where the subject material is developed.

wells, molecular materials, and nonlinear optics will not be found in most of the standard texts. Other new topics such as the Bose–Einstein condensation of excitons are included alongside traditional subject material. Furthermore, it is my deliberate intention to illustrate the physics with up-to-date examples of optical technology. This provides an interesting modern motivation for traditional topics such as colour centres and also helps to emphasize the importance of the solid state devices.

Throughout the book I have understood the term ‘optical’ in a wider sense than its strict meaning referring to the visible spectral region. This has allowed me to include discussions of infrared phenomena such as those due to phonons and free carriers, and also the properties of insulators and metals in the ultraviolet. I have likewise taken the scope of the word ‘solid’ beyond the traditional emphasis on crystalline materials such as metals, semiconductors and insulators. This has allowed me to include both ‘soft condensed matter’ materials, such as polymers, and also glasses, which are not solids in the strict sense.

The process of relating measured optical phenomena to the electronic and vibrational properties of the material under study can proceed in two ways. We can work forwards from known electronic or vibrational physics to predict the results of optical experiments, or we can work backwards from experimental data to the microscopic properties. An example of the first approach is to use the free electron theory to explain why metals reflect light, while an example of the second is to use absorption or emission data to deduce the electron level structure of a crystal. Textbooks such as this one inevitably tend to work forwards from the microscopic properties to the measured data, even though an experimental scientist would probably be working in the other direction.

The book presupposes that the reader has a working knowledge of solid state physics at the level appropriate to a third year undergraduate, such as that found in H.M. Rosenberg’s *The Solid State* (Oxford University Press, third

edition, 1988). This puts the treatment at about the same as, or at a slightly higher level, than that given in the *Introduction to Solid State Physics* by Charles Kittel. The book also necessarily presupposes a reasonable knowledge of electromagnetism and quantum theory. Classical and quantum arguments are used interchangeably throughout, and the reader will need to revise their own favourite texts on these subjects if any of the material is unfamiliar. Three appendices are included to provide a succinct summary of the principal results from band theory, electromagnetism and quantum theory that are presupposed.

The text has been written in a tutorial style, with worked examples in most chapters. A collection of exercises is provided at the end of each chapter, with solutions at the end of the book. The exercises follow the presentation of the material in the chapter, and the more challenging ones are identified with an asterisk.

Acknowledgements

I would like to acknowledge the many people who have helped in various ways in the production of this book. Pride of place goes to Soenke Adlung at Oxford University Press, for instigating and editing the Oxford Masters Series, and to my series co-authors, Stephen Blundell and John Singleton, who have offered many useful ideas on how to approach the project. I would also like to thank Anja Tschoertner and Richard Lawrence at the Press for handling the day-to-day administration of the manuscript and producing the book.

Numerous colleagues have helped to clarify my understanding of certain specialist topics and have made comments on the text. Among these I would like to single out Dr Simon Martin and Dr Paul Lane for their critical reading of the chapter on molecular materials. I am, of course, grateful to my teachers, who enthused me with the subject, and to the students who have taken the course and tried out some of the problems.

I would like to thank Dr Robert Taylor for providing me with unpublished data for use in Figs 5.3, 6.13 and 6.17, Adam Ashmore for taking the data in Figs 5.6 and 5.13, and Drs Gero von Plessen and Andrew Tomlinson for the data presented in Fig. 4.5. I would also like to thank Excerpta Medica Inc. for granting permission to reproduce Figs 1.6, 4.4, 4.7, 7.13, 8.10, 8.16, 8.19, 9.6, and 10.11; the Optical Society of America for Fig. 1.7; the American Physical Society for Figs 3.4, 3.7, 4.3, 7.2, 7.3, 7.7, 7.11, 10.7, and 10.14; Springer-Verlag for Figs 3.9 and 3.12; the Physical Society of Japan for Fig. 4.9; the American Institute of Physics for Figs 5.8, 6.10, 8.9, 8.17, and 11.9; Taylor and Francis Ltd for Figs 6.9 and 6.12; the Institute of Physics, Polish Academy of Sciences, for Fig. 6.16; Verlag der Zeitschrift für Naturforschung for Fig. 8.13; Kluwer Academic/Plenum Publishers for Figs 9.5, 9.15, and 11.7; Academic Press for Figs 9.9 and 10.5; Penwell Corp. for Fig. 9.11; and the IEEE for Fig. 11.10.

Finally, I would like to thank the Royal Society for supporting me as a University Research Fellow during the years in which most of the book has been written.

M.F.

*Sheffield
January 2001*

Contents

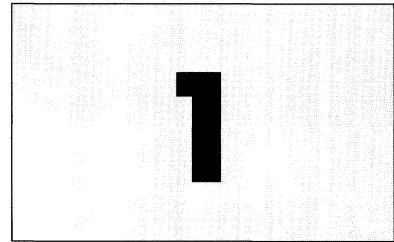
1	Introduction	1
1.1	Classification of optical processes	1
1.2	Optical coefficients	2
1.3	The complex refractive index and dielectric constant	5
1.4	Optical materials	8
1.4.1	Crystalline insulators and semiconductors	8
1.4.2	Glasses	11
1.4.3	Metals	12
1.4.4	Molecular materials	13
1.4.5	Doped glasses and insulators	14
1.5	Characteristic optical physics in the solid state	15
1.5.1	Crystal symmetry	16
1.5.2	Electronic bands	18
1.5.3	Vibronic bands	19
1.5.4	The density of states	19
1.5.5	Delocalized states and collective excitations	19
1.6	Microscopic models	20
2	Classical propagation	25
2.1	Propagation of light in a dense optical medium	25
2.1.1	Atomic oscillators	26
2.1.2	Vibrational oscillators	28
2.1.3	Free electron oscillators	29
2.2	The dipole oscillator model	29
2.2.1	The Lorentz oscillator	29
2.2.2	Multiple resonances	34
2.2.3	Comparison with experimental data	37
2.2.4	Local field corrections	39
2.2.5	The Kramers–Kronig relationships	40
2.3	Dispersion	40
2.4	Optical anisotropy: birefringence	42
3	Interband absorption	49
3.1	Interband transitions	49
3.2	The transition rate for direct absorption	51
3.3	Band edge absorption in direct gap semiconductors	54
3.3.1	The atomic physics of the interband transitions	54
3.3.2	The band structure of a direct gap III–V semiconductor	56
3.3.3	The joint density of states	57

3.3.4	The frequency dependence of the band edge absorption	58
3.3.5	The Franz–Keldysh effect	60
3.3.6	Band edge absorption in a magnetic field	61
3.4	Band edge absorption in indirect gap semiconductors	63
3.5	Interband absorption above the band edge	66
3.6	Measurement of absorption spectra	68
3.7	Semiconductor photodetectors	69
3.7.1	Photodiodes	69
3.7.2	Photoconductive devices	71
3.7.3	Photovoltaic devices	71
4	Excitons	76
4.1	The concept of excitons	76
4.2	Free excitons	77
4.2.1	Binding energy and radius	77
4.2.2	Exciton absorption	79
4.2.3	Experimental data for free excitons in GaAs	80
4.3	Free excitons in external fields	82
4.3.1	Electric fields	82
4.3.2	Magnetic fields	83
4.4	Free excitons at high densities	84
4.5	Frenkel excitons	86
4.5.1	Rare gas crystals	87
4.5.2	Alkali halides	87
4.5.3	Molecular crystals	88
5	Luminescence	92
5.1	Light emission in solids	92
5.2	Interband luminescence	95
5.2.1	Direct gap materials	95
5.2.2	Indirect gap materials	96
5.3	Photoluminescence	98
5.3.1	Excitation and relaxation	98
5.3.2	Low carrier densities	100
5.3.3	Degeneracy	101
5.3.4	Photoluminescence spectroscopy	102
5.4	Electroluminescence	103
5.4.1	General principles of electroluminescent devices	103
5.4.2	Light emitting diodes	106
5.4.3	Diode lasers	107
6	Semiconductor quantum wells	115
6.1	Quantum confined structures	115
6.2	Growth and structure of semiconductor quantum wells	117
6.3	Electronic levels	119
6.3.1	Separation of the variables	119
6.3.2	Infinite potential wells	119
6.3.3	Finite potential wells	121
6.4	Optical absorption and excitons	124

58	6.4.1 Selection rules	125
60	6.4.2 Two-dimensional absorption	127
61	6.4.3 Experimental data	129
63	6.4.4 Excitons in quantum wells	130
66	6.5 The quantum confined Stark effect	131
68	6.6 Optical emission	133
69	6.7 Intersubband transitions	135
69	6.8 Bloch oscillators	135
71	6.9 Quantum dots	136
71	6.9.1 Semiconductor doped glasses	137
71	6.9.2 Self-organized III–V quantum dots	138
76		
76	7 Free electrons	143
77	7.1 Plasma reflectivity	143
77	7.2 Free carrier conductivity	145
79	7.3 Metals	147
80	7.3.1 The Drude model	148
82	7.3.2 Interband transitions in metals	151
82	7.4 Doped semiconductors	154
83	7.4.1 Free carrier reflectivity and absorption	154
84	7.4.2 Impurity absorption	158
86	7.5 Plasmons	160
87		
87	8 Molecular materials	165
88	8.1 Introduction to molecular materials	166
88	8.2 Electronic states in conjugated molecules	167
92	8.3 Optical spectra of molecules	169
92	8.3.1 Electronic–vibrational transitions	170
95	8.3.2 Molecular configuration diagrams	171
95	8.3.3 The Franck–Condon principle	173
96	8.3.4 Experimental spectra	176
98	8.4 Aromatic hydrocarbons	177
98	8.5 Conjugated polymers	179
100	8.6 Organic optoelectronics	181
101		
102	9 Luminescence centres	186
103	9.1 Vibronic absorption and emission	186
103	9.2 Colour centres	189
103	9.3 Paramagnetic impurities in ionic crystals	192
106	9.3.1 The crystal field effect and vibronic coupling	192
107	9.3.2 Rare earth ions	193
115	9.3.3 Transition metal ions	195
115	9.4 Solid state lasers and optical amplifiers	196
117	9.5 Phosphors	199
119		
119	10 Phonons	204
119	10.1 Infrared active phonons	204
119	10.2 Infrared reflectivity and absorption in polar solids	206
121	10.2.1 The classical oscillator model	206
124	10.2.2 The Lyddane–Sachs–Teller relationship	209

10.2.3 Restrahlen	210
10.2.4 Lattice absorption	212
10.3 Polaritons	214
10.4 Polarons	215
10.5 Inelastic light scattering	218
10.5.1 General principles of inelastic light scattering	218
10.5.2 Raman scattering	220
10.5.3 Brillouin scattering	222
10.6 Phonon lifetimes	222
11 Nonlinear optics	227
11.1 The nonlinear susceptibility tensor	227
11.2 The physical origin of optical nonlinearities	230
11.2.1 Non-resonant nonlinearities	230
11.2.2 Resonant nonlinearities	234
11.3 Second-order nonlinearities	236
11.3.1 Nonlinear frequency mixing	237
11.3.2 Crystal symmetry	239
11.3.3 Phase matching	241
11.4 Third-order nonlinear effects	243
11.4.1 Overview of third-order phenomena	243
11.4.2 Isotropic third-order nonlinear media	247
11.4.3 Resonant nonlinearities in semiconductors	249
A Electromagnetism in dielectrics	255
A.1 Electromagnetic fields and Maxwell's equations	255
A.2 Electromagnetic waves	258
B Quantum theory of radiative absorption and emission	263
B.1 Einstein coefficients	263
B.2 Quantum transition rates	266
B.3 Selection rules	269
C Band theory	271
C.1 Metals, semiconductors and insulators	271
C.2 The nearly free electron model	273
C.3 Example band structures	276
D Semiconductor p-n diodes	280
Solutions to exercises	282
Bibliography	290
Symbols	295
Index	297

Introduction



Light interacts with matter in many different ways. Metals are shiny, but water is transparent. Stained glass and gemstones transmit some colours, but absorb others. Other materials such as milk appear white because they scatter the incoming light in all directions.

In the chapters that follow, we will be looking at a whole host of these optical phenomena in a wide range of solid state materials. Before we can begin to do this, we must first describe the way in which the phenomena are classified, and the coefficients that are used to quantify them. We must then introduce the materials that we will be studying, and clarify in general terms how the solid state is different from the gas and liquid phase. This is the subject of the present chapter.

1.1 Classification of optical processes

The wide-ranging optical properties observed in solid state materials can be classified into a small number of general phenomena. The simplest group, namely **reflection**, **propagation** and **transmission**, is illustrated in Fig. 1.1. This shows a light beam incident on an optical medium. Some of the light is reflected from the front surface, while the rest enters the medium and propagates through it. If any of this light reaches the back surface, it can be reflected again, or it can be transmitted through to the other side. The amount of light transmitted is therefore related to the reflectivity at the front and back surfaces and also to the way the light propagates through the medium.

The phenomena that can occur while light propagates through an optical medium are illustrated schematically in Fig. 1.2.

1.1	Classification of optical processes	1
1.2	Optical coefficients	2
1.3	The complex refractive index and dielectric constant	5
1.4	Optical materials	8
1.5	Characteristic optical physics in the solid state	15
1.6	Microscopic models	20

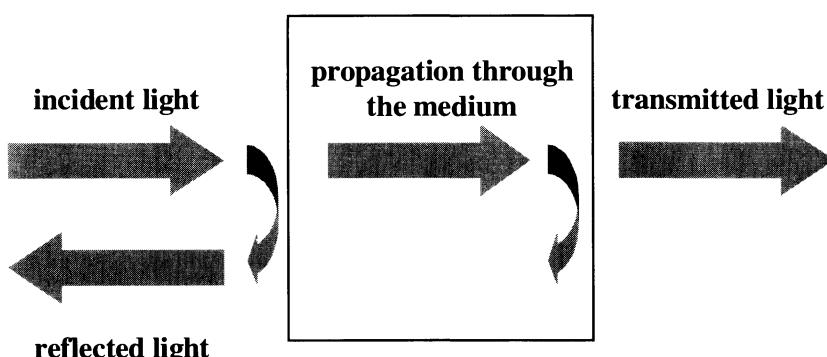


Fig. 1.1 Reflection, propagation and transmission of a light beam incident on an optical medium.

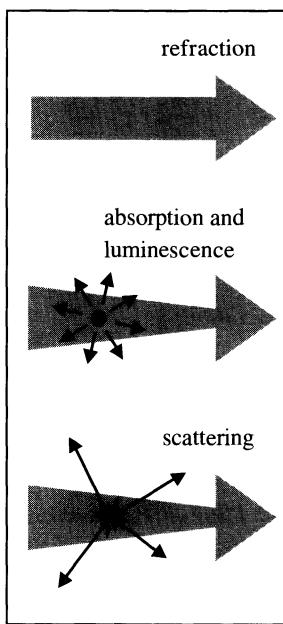


Fig. 1.2 Phenomena that can occur as a light beam propagates through an optical medium. Refraction causes a reduction in the velocity of the wave, while absorption causes attenuation. Luminescence can accompany absorption if the excited atoms re-emit by spontaneous emission. Scattering causes a redirection of the light. The diminishing width of the arrow for the processes of absorption and scattering represents the attenuation of the beam.

Refraction causes the light waves to propagate with a smaller velocity than in free space. This reduction of the velocity leads to the bending of light rays at interfaces described by Snell's law of refraction. Refraction, in itself, does not affect the intensity of the light wave as it propagates.

Absorption occurs during the propagation if the frequency of the light is resonant with the transition frequencies of the atoms in the medium. In this case, the beam will be attenuated as it progresses. The transmission of the medium is clearly related to the absorption, because only unabsorbed light will be transmitted. Selective absorption is responsible for the colouration of many optical materials. Rubies, for example, are red because they absorb blue and green light, but not red.

Luminescence is the general name given to the process of spontaneous emission of light by excited atoms in a solid state material. One of the ways in which the atoms can be promoted into excited states prior to spontaneous emission is by the absorption of light. Luminescence can thus accompany the propagation of light in an absorbing medium. The light is emitted in all directions, and has a different frequency to the incoming beam.

Luminescence does not always have to accompany absorption. It takes a characteristic amount of time for the excited atoms to re-emit by spontaneous emission. This means that it might be possible for the excited atoms to dissipate the excitation energy as heat before the radiative re-emission process occurs. The efficiency of the luminescence process is therefore closely tied up with the dynamics of the de-excitation mechanisms in the atoms.

Scattering is the phenomenon in which the light changes direction and possibly also its frequency after interacting with the medium. The total number of photons is unchanged, but the number going in the forward direction decreases because light is being re-directed in other directions. Scattering therefore has the same attenuating effect as absorption. The scattering is said to be **elastic** if the frequency of the scattered light is unchanged, or **inelastic** if the frequency changes in the process. The difference in the photon energy in an inelastic scattering process has to be taken from the medium if the frequency increases or given to the medium if the frequency decreases.

A number of other phenomena can occur as the light propagates through the medium if the intensity of the beam is very high. These are described by **nonlinear optics**. An example is frequency doubling, in which the frequency of part of a beam is doubled by interaction with the optical medium. These nonlinear effects have only been discovered through the use of lasers. At this stage, we only mention their existence for completeness, and postpone their further discussion until Chapter 11.

1.2 Optical coefficients

The optical phenomena described in the previous section can be quantified by a number of parameters that determine the properties of the medium at the macroscopic level.

The reflection at the surfaces is described by the **coefficient of reflection** or **reflectivity**. This is usually given the symbol R and is defined as the ratio of the reflected power to the power incident on the surface. The **coefficient**

of transmission or transmissivity T is defined likewise as the ratio of the transmitted power to the incident power. If there is no absorption or scattering, then by conservation of energy we must have that:

$$R + T = 1. \quad (1.1)$$

The propagation of the beam through a transparent medium is described by the refractive index n . This is defined as the ratio of the velocity of light in free space c to the velocity of light in the medium v according to:

$$n = \frac{c}{v}. \quad (1.2)$$

The refractive index depends on the frequency of the light beam. This effect is called dispersion, and will be discussed in detail in Section 2.3. In colourless transparent materials such as glass, the dispersion is small in the visible spectral region, and it therefore makes sense to speak of ‘the’ refractive index of the substance in question.

The absorption of light by an optical medium is quantified by its absorption coefficient α . This is defined as the fraction of the power absorbed in a unit length of the medium. If the beam is propagating in the z direction, and the intensity (optical power per unit area) at position z is $I(z)$, then the decrease of the intensity in an incremental slice of thickness dz is given by:

$$dI = -\alpha dz \times I(z). \quad (1.3)$$

This can be integrated to obtain Beer’s law:

$$I(z) = I_0 e^{-\alpha z}, \quad (1.4)$$

where I_0 is the optical intensity at $z = 0$. The absorption coefficient is a strong function of frequency, so that optical materials may absorb one colour but not another.

In the next section we will explain how both the absorption and the refraction can be incorporated into a single quantity called the complex refractive index. Knowledge of this quantity enables us to calculate the reflectivity R , and hence the transmissivity T . This last point follows because the transmissivity of an absorbing medium of thickness l is given by:

$$T = (1 - R_1) e^{-\alpha l} (1 - R_2), \quad (1.5)$$

where R_1 and R_2 are the reflectivities of the front and back surfaces respectively. This formula applies to the transmission of light through an optical medium such as the one shown in Fig. 1.1. The first and third terms on the right hand side of eqn 1.5 account for the transmission of the front and back surfaces respectively, while the middle term gives the exponential decrease in intensity due to the absorption according to Beer’s law. If the front and back surfaces have equal reflectivities R , as will usually be the case, then eqn 1.5 simplifies to:

$$T = (1 - R)^2 e^{-\alpha l}. \quad (1.6)$$

Equation (1.5) ignores the possibility of multiple reflections between the front and back surfaces. These will have to be included if the surfaces are parallel and the reflection coefficients are sufficiently large. We will come across some examples where these effects are important when we consider semiconductor laser diodes in Section 5.4.3 and optical bistability in Section 11.4.3. In many cases, however, the effects are small enough to be neglected, as shown in Exercises 1.8 and 1.9.

The optical density, and hence the absorption coefficient, is usually worked out from the measured transmissivity of the sample. This requires accurate normalization of the reflection losses at the surfaces. (See Exercise 1.10.)

The absorption of an optical medium can also be sometimes quantified in terms of the **optical density** (O.D.). This is sometimes called the **absorbance**, and is defined as:

$$\text{O.D.} = -\log_{10} \left(\frac{I(l)}{I_0} \right), \quad (1.7)$$

where l is the length of the absorbing medium. It is apparent from eqn 1.4 that the optical density is directly related to the absorption coefficient α through:

$$\text{O.D.} = \frac{\alpha l}{\log_e(10)} = 0.434 \alpha l. \quad (1.8)$$

In this book we will quantify the absorption by α instead of the optical density because it is independent of the sample length.

The phenomenon of luminescence was studied extensively by George Stokes in the nineteenth century before the advent of quantum theory. Stokes discovered that the luminescence is down-shifted in frequency relative to the absorption, an effect now known as the **Stokes shift**. Luminescence cannot be described easily by macroscopic classical parameters because spontaneous emission is fundamentally a quantum process (see Appendix B).

The simplest sequence of events that takes place in luminescence is illustrated in Fig. 1.3. The atom jumps to an excited state by absorbing a photon, then relaxes to an intermediate state, and finally re-emits a photon as it drops back to the ground state. The Stokes shift is explained by applying the law of conservation of energy to the process. It is easy to see that the energy of the photon emitted must be less than that of the photon absorbed, and hence that the frequency of the emitted light is less than that of the absorbed light. The magnitude of the Stokes shift is therefore determined by the energy levels of the atoms in the medium.

Scattering is caused by variations of the refractive index of the medium on a length scale smaller than the wavelength of the light. This could be caused by the presence of impurities, defects, or inhomogeneities. Scattering causes attenuation of a light beam in an analogous way to absorption. The intensity decreases exponentially as it propagates into the medium according to:

$$I(z) = I_0 \exp(-N\sigma_s z), \quad (1.9)$$

where N is the number of scattering centres per unit volume, and σ_s is the **scattering cross-section** of the scattering centre. This is identical in form to Beer's law given in eqn 1.4, with $\alpha \equiv N\sigma_s$.

The scattering is described as **Rayleigh scattering** if the size of the scattering centre is very much smaller than the wavelength of the light. In this case, the scattering cross-section will vary with the wavelength λ according to:

$$\sigma_s(\lambda) \propto \frac{1}{\lambda^4}. \quad (1.10)$$

The Rayleigh scattering law implies that inhomogeneous materials tend to scatter short wavelengths more strongly than longer wavelengths.

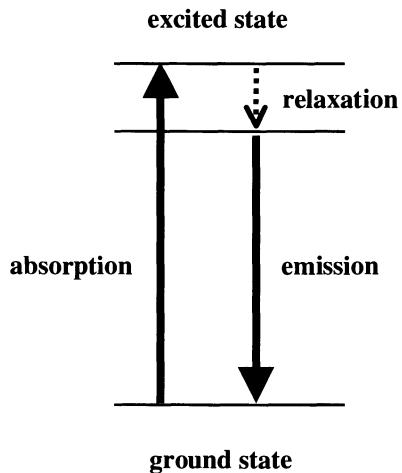


Fig. 1.3 Luminescence process in an atom. The atom jumps to an excited state by absorption of a photon, then relaxes to an intermediate state, before re-emitting a photon by spontaneous emission as it falls back to the ground state. The photon emitted has a smaller energy than the absorbed photon. This reduction in the photon energy is called the Stokes shift.

Example 1.1

The reflectivity of silicon at 633 nm is 35% and the absorption coefficient is $3.8 \times 10^5 \text{ m}^{-1}$. Calculate the transmission and optical density of a sample with a thickness of $10 \mu\text{m}$.

Solution

The transmission is given by eqn 1.6 with $R = 0.35$ and $\alpha l = (3.8 \times 10^5) \times (10 \times 10^{-6}) = 3.8$. This gives:

$$T = (1 - 0.35)^2 \cdot \exp(-3.8) = 0.0095.$$

The optical density is given by eqn 1.8:

$$\text{O.D.} = 0.434 \times 3.8 = 1.65.$$

1.3 The complex refractive index and dielectric constant

In the previous section we mentioned that the absorption and refraction of a medium can be described by a single quantity called the **complex refractive index**. This is usually given the symbol \tilde{n} and is defined through the equation:

$$\tilde{n} = n + i\kappa. \quad (1.11)$$

The real part of \tilde{n} , namely n , is the same as the normal refractive index defined in eqn. 1.2. The imaginary part of \tilde{n} , namely κ , is called the **extinction coefficient**. As we will see below, κ is directly related to the absorption coefficient α of the medium.

The relationship between α and κ can be derived by considering the propagation of plane electromagnetic waves through a medium with a complex refractive index. If the wave is propagating in the z direction, the spatial and time dependence of the electric field is given by (see eqn A.32 in Appendix A):

$$\mathcal{E}(z, t) = \mathcal{E}_0 e^{i(kz - \omega t)}, \quad (1.12)$$

where k is the wave vector of the light and ω is the angular frequency. $|\mathcal{E}_0|$ is the amplitude at $z = 0$. In a non-absorbing medium of refractive index n , the wavelength of the light is reduced by a factor n compared to the free space wavelength λ . k and ω are therefore related to each other through:

$$k = \frac{2\pi}{(\lambda/n)} = \frac{n\omega}{c}. \quad (1.13)$$

This can be generalized to the case of an absorbing medium by allowing the refractive index to be complex:

$$k = \tilde{n} \frac{\omega}{c} = (n + i\kappa) \frac{\omega}{c}. \quad (1.14)$$

On substituting eqn 1.14 into eqn 1.12, we obtain:

$$\begin{aligned}\mathcal{E}(z, t) &= \mathcal{E}_0 e^{i(\omega\tilde{n}z/c - \omega t)} \\ &= \mathcal{E}_0 e^{-\kappa\omega z/c} e^{i(\omega n z/c - \omega t)}.\end{aligned}\quad (1.15)$$

This shows that a non-zero extinction coefficient leads to an exponential decay of the wave in the medium. At the same time, the real part of \tilde{n} still determines the phase velocity of the wave front, as in the standard definition of the refractive index given in eqn 1.2.

The optical intensity of a light wave is proportional to the square of the electric field, namely $I \propto \mathcal{E}\mathcal{E}^*$ (c.f. eqn A.40). We can therefore deduce from eqn 1.15 that the intensity falls off exponentially in the medium with a decay constant equal to $2 \times (\kappa\omega/c)$. On comparing this to Beer's law given in eqn 1.4 we conclude that:

$$\alpha = \frac{2\kappa\omega}{c} = \frac{4\pi\kappa}{\lambda}, \quad (1.16)$$

where λ is the free space wavelength of the light. This shows us that κ is directly proportional to the absorption coefficient.

We can relate the refractive index of a medium to its relative dielectric constant ϵ_r by using the standard result derived from Maxwell's equations (cf. eqn A.31 in Appendix A):

$$n = \sqrt{\epsilon_r}. \quad (1.17)$$

This shows us that if n is complex, then ϵ_r must also be complex. We therefore define the **complex relative dielectric constant** $\tilde{\epsilon}_r$ according to:

$$\tilde{\epsilon}_r = \epsilon_1 + i\epsilon_2. \quad (1.18)$$

By analogy with eqn 1.17, we see that \tilde{n} and $\tilde{\epsilon}_r$ are related to each other through:

$$\tilde{n}^2 = \tilde{\epsilon}_r \quad (1.19)$$

We can now work out explicit relationships between the real and imaginary parts of \tilde{n} and $\tilde{\epsilon}_r$ by combining eqns 1.11, 1.18 and 1.19. These are:

$$\epsilon_1 = n^2 - \kappa^2 \quad (1.20)$$

$$\epsilon_2 = 2n\kappa, \quad (1.21)$$

and

$$n = \frac{1}{\sqrt{2}} \left(\epsilon_1 + (\epsilon_1^2 + \epsilon_2^2)^{\frac{1}{2}} \right)^{\frac{1}{2}} \quad (1.22)$$

$$\kappa = \frac{1}{\sqrt{2}} \left(-\epsilon_1 + (\epsilon_1^2 + \epsilon_2^2)^{\frac{1}{2}} \right)^{\frac{1}{2}}. \quad (1.23)$$

This analysis shows us that \tilde{n} and $\tilde{\epsilon}_r$ are not independent variables: if we know ϵ_1 and ϵ_2 we can calculate n and κ , and *vice versa*. Note that if the medium is only weakly absorbing, then we can assume that κ is very small, so that eqns 1.22 and 1.23 simplify to:

$$n = \sqrt{\epsilon_1} \quad (1.24)$$

$$\kappa = \frac{\epsilon_2}{2n}. \quad (1.25)$$

These equations show us that the refractive index is basically determined by the real part of the dielectric constant, while the absorption is mainly determined by the imaginary part. This generalization is obviously not valid if the medium has a very large absorption coefficient.

The microscopic models that we will be developing throughout the book usually enable us to calculate $\tilde{\epsilon}_r$ rather than \tilde{n} . The measurable optical properties can then be obtained by converting ϵ_1 and ϵ_2 to n and κ through eqns 1.22 and 1.23. The refractive index is given directly by n , while the absorption coefficient can be worked out from κ using eqn 1.16. The reflectivity depends on both n and κ and is given by

$$R = \left| \frac{\tilde{n} - 1}{\tilde{n} + 1} \right|^2 = \frac{(n - 1)^2 + \kappa^2}{(n + 1)^2 + \kappa^2}. \quad (1.26)$$

This formula is derived in eqn A.50. It gives the coefficient of reflection between the medium and the air (or vacuum) at normal incidence.

In a transparent material such as glass in the visible region of the spectrum, the absorption coefficient is very small. Equations 1.16 and 1.21 then tell us that κ and ϵ_2 are negligible, and hence that both \tilde{n} and $\tilde{\epsilon}_r$ may be taken as real numbers. This is why tables of the properties of transparent optical materials generally list only the real parts of the refractive index and dielectric constant. On the other hand, if there is significant absorption, then we will need to know both the real and imaginary parts of \tilde{n} and $\tilde{\epsilon}_r$.

In the remainder of this book we will take it as explicitly assumed that both the refractive index and the dielectric constant are complex quantities. We will therefore drop the tilde notation on n and ϵ_r from now on, except where it is explicitly needed to avoid ambiguity. It will usually be obvious from the context whether we are dealing with real or complex quantities.

Example 1.2

The complex refractive index of germanium at 400 nm is given by $\tilde{n} = 4.141 + i 2.215$. Calculate for germanium at 400 nm: (a) the phase velocity of light, (b) the absorption coefficient, and (c) the reflectivity.

Solution

(a) The velocity of light is given by eqn 1.2, where n is the real part of \tilde{n} . Hence we obtain:

$$v = \frac{c}{n} = \frac{2.998 \times 10^8}{4.141} \text{ m s}^{-1} = 7.24 \times 10^7 \text{ m s}^{-1}.$$

(b) The absorption coefficient is given by eqn 1.16. By inserting $\kappa = 2.215$ and $\lambda = 400 \text{ nm}$, we obtain:

$$\alpha = \frac{4\pi \times 2.215}{400 \times 10^{-9}} \text{ m}^{-1} = 6.96 \times 10^7 \text{ m}^{-1}.$$

(c) The reflectivity is given by eqn 1.26. Inserting $n = 4.141$ and $\kappa = 2.215$ into this, we obtain:

$$R = \frac{(4.141 - 1)^2 + 2.215^2}{(4.141 + 1)^2 + 2.215^2} = 47.1 \text{ \%}.$$

We will see in Chapter 10 that the restrahlen absorption is caused by the interaction between the light and the optical phonons.

Example 1.3

Salt (NaCl) absorbs very strongly at infrared wavelengths in the ‘restrahlen’ band. The complex dielectric constant at $60\text{ }\mu\text{m}$ is given by $\tilde{\epsilon}_r = -16.8 + i 91.4$. Calculate the absorption coefficient and the reflectivity at this wavelength.

Solution

We must first work out the complex refractive index using eqns 1.22 and 1.23. This gives:

$$n = \frac{1}{\sqrt{2}} \left(-16.8 + ((-16.8)^2 + 91.4^2)^{\frac{1}{2}} \right)^{\frac{1}{2}} = 6.17$$

and

$$\kappa = \frac{1}{\sqrt{2}} \left(+16.8 + ((-16.8)^2 + 91.4^2)^{\frac{1}{2}} \right)^{\frac{1}{2}} = 7.41.$$

We then insert these values into eqns 1.16 and 1.26 to obtain the required results:

$$\alpha = \frac{4\pi \times 7.41}{60 \times 10^{-6}} \text{ m}^{-1} = 1.55 \times 10^6 \text{ m}^{-1},$$

and

$$R = \frac{(6.17 - 1)^2 + 7.41^2}{(6.17 + 1)^2 + 7.41^2} = 76.8 \text{ \%}.$$

1.4 Optical materials

We will be studying the optical properties of many different types of solid state materials throughout this book. The materials can be loosely classified into five general categories:

- Crystalline insulators and semiconductors
- Glasses
- Metals
- Molecular materials
- Doped glasses and insulators.

Before delving into the details, we give here a brief overview of the main optical properties of these materials. This will serve as an introduction to the optical physics that will be covered in the following chapters.

1.4.1 Crystalline insulators and semiconductors

Figure 1.4(a) shows the transmission spectrum of crystalline sapphire (Al_2O_3) from the infrared to the ultraviolet spectral region. The spectrum for sapphire shows the main features observed in all insulators, although of course the details will vary considerably from material to material. The principal optical properties can be summarized as follows:

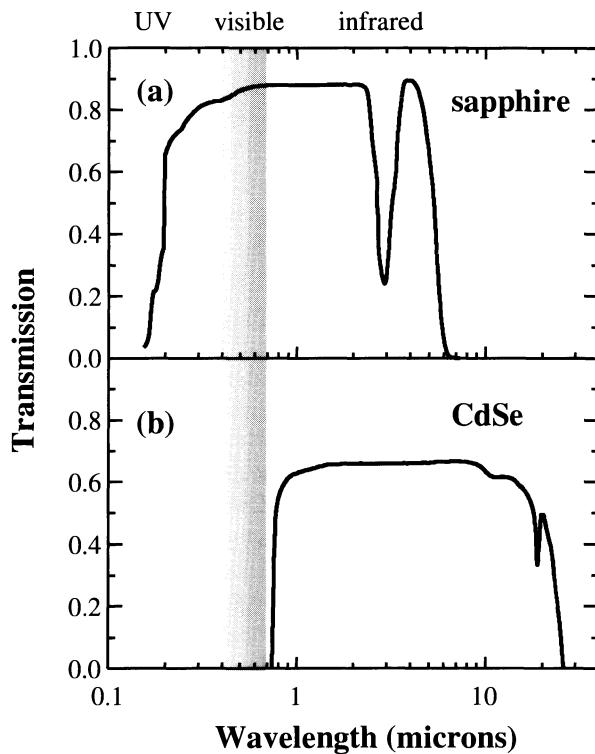


Fig. 1.4 (a) Transmission spectrum of a sapphire (Al_2O_3) crystal of thickness 3 mm. (b) Transmission spectrum of a CdSe crystal of thickness 1.67 mm. After [1].

- (1) Sapphire has a high transmission in the wavelength range $0.2\text{--}6 \mu\text{m}$. This defines the **transparency range** of the crystal. The transparency region of sapphire includes the whole of the visible spectrum, which explains why it appears colourless and transparent to the human eye.
- (2) Within the transparency range the absorption coefficient is very small, and the refractive index may be taken to be real with no imaginary component. The value of the refractive index is approximately constant, and is equal to 1.77 in sapphire.
- (3) The transmission coefficient in the transparency range is determined by the reflectivity of the surfaces through eqn 1.6. The reflectivity in turn is determined by the refractive index through eqn 1.26. For sapphire with $n = 1.77$, this gives $R = 0.077$. Hence we find $T = (1 - R)^2 = 0.85$.
- (4) The dip in the transmission in the infrared around $3 \mu\text{m}$, and the sharp drop in the transmission for $\lambda > 6 \mu\text{m}$, is caused by vibrational absorption. This absorption mechanism is analogous to the infrared absorption due to vibrations in polar molecules. The vibrational excitations of a crystal lattice are called phonon modes, and so the vibrational absorption in a solid is usually called phonon absorption or lattice absorption. This absorption mechanism will be discussed in Chapter 10.
- (5) The transmission drops sharply in the ultraviolet spectral region for $\lambda < 0.2 \mu\text{m}$ due to absorption by bound electrons. The onset of the absorption is called the **fundamental absorption edge**. The wavelength of the fundamental edge is determined by the band gap of the insulator. The

Sapphire gemstones tend to be blue. This is caused by the presence of chromium, titanium and iron impurities in the Al_2O_3 crystal. Pure synthetic Al_2O_3 crystals are colourless.

Sapphire actually transmits in the far infrared spectral region when the frequency is well below that of the optical phonons.

explanation of the absorption spectra due to bound electrons needs band theory, and will be discussed in Chapters 3 and 4.

Point (1) is perhaps the most obvious aspect of the optical properties of insulators: they all tend to be colourless and transparent in the visible spectral region. If they are coloured, this is most likely caused by the presence of impurities, as will be explained in Section 1.4.5 below. This transparency is slightly deceptive. The insulators do absorb very strongly in the ultraviolet and in the infrared, but this is hidden from the human eye. The transparent region between the infrared and ultraviolet absorption bands is particularly useful for making optical windows and lenses. The approximate transparency range and refractive index of a number of common crystalline insulators are listed in Table 1.1.

The crystallinity of the materials gives rise to a number of properties relating to the underlying symmetry of the lattice. This point will be expanded further in Section 1.5.1. One immediate consequence is that some of the materials listed in Table 1.1 are birefringent. The optical properties are anisotropic, and the value of the refractive index depends on the direction of propagation of the light relative to the crystallographic axes. The phenomenon of birefringence will be described in more detail in Section 2.4.

The optical properties of semiconductors are conceptually similar to those of insulators, except that the electronic and vibrational transitions occur at longer wavelengths. By way of example, Fig. 1.4(b) shows the transmission spectrum of the II–VI compound semiconductor CdSe over the same wavelength range as for the sapphire crystal. Just as with sapphire, we have a transparency range which is limited by electronic absorption at short wavelengths and lattice absorption at long wavelengths. The maximum transmission is around 60% which is again mainly limited by the surface reflectivities. The short wavelength edge occurs beyond 700 nm, which means that the whole of the transparency range lies outside the visible spectrum. Hence no visible light is transmitted through the crystal, and it has a dark metallic appearance to the eye.

Table 1.2 lists the transparency range and refractive index of several semiconductors. The data show that the lower limit of the transmission range coincides closely with the wavelength of the fundamental band gap. This happens because the band gap determines the lowest energy for interband transitions, as will be explained in Chapter 3. Note that the refractive index increases as the band gap wavelength gets larger.

The upper limit of the transmission range is determined by the lattice absorption, as for insulators, and also by free carrier absorption. Free carriers are present in semiconductors at room temperature through the thermal excitation of electrons across the band gap or due to the presence of impurities. This causes infrared absorption, as will be explained in Section 7.4. Insulators have very small free carrier densities due to their large band gaps.

One very important aspect of the optical properties of semiconductors is that a subset of them, namely those with direct band gaps, luminesce strongly when electrons are promoted to the conduction band. This is the physical basis for the light-emitting devices used in the optoelectronics industry. The physical processes behind the luminescence will be explained in Chapter 5. The main point is that the wavelength of the luminescence coincides with the band gap

The very high transparency of diamond in the infrared is noteworthy. This is caused by the fact that diamond is a purely covalent crystal, which means that its optical phonons cannot interact directly with light waves. This point will be discussed further in Chapter 10.

Table 1.1 Approximate transparency range and refractive index n of a number of crystalline insulators. n is measured at 546 nm. Values of n are given both for the o-ray and e-ray of birefringent materials. After [1] and [2].

Crystal	Transparency range (μm)	n
Al_2O_3 (sapphire)	0.2–6	1.771 (o) 1.763 (e)
BaF_2	0.2–12	1.476
Diamond	0.25–> 80	2.424
KBr	0.3–30	1.564
KCl	0.21–25	1.493
KI	0.3–40	1.673
MgF_2	0.12–8	1.379 (o) 1.390 (e)
NaCl	0.21–20	1.55
NaF	0.19–15	1.326
SiO_2 (quartz)	0.2–3	1.546 (o) 1.555 (e)
TiO_2 (rutile)	0.45–5	2.652 (o) 2.958 (e)

of the semiconductor. In Chapter 6 we will see how quantum size effects in low-dimensional semiconductors can be used to shift the effective band gap to higher energy. This is a highly desirable feature, because it provides a way to ‘tune’ the emission wavelength by controlled variation of the parameters during the crystal growth.

1.4.2 Glasses

Glasses are extremely important optical materials. They have been used for centuries in prisms and lenses for optical instruments, in addition to their common usage in windows and glassware. In more recent times they have found new applications in optical fibre technology. With the exception of stained glasses, they are usually made to be transparent in the visible spectrum. They are not crystalline solids, and therefore do not exhibit the optical anisotropy that is characteristic of some crystals.

Most types of glasses are made by fusing sand (silica: SiO_2) with other chemicals. Pure fused silica is an insulator, and shows all the characteristic features of insulators discussed in the previous section. It is transparent in the visible region, but absorbs in the ultraviolet due to the electronic transitions of the SiO_2 molecules, and in the infrared due to vibrational absorption. The transparency range thus goes from around 200 nm in the ultraviolet to beyond 2000 nm in the infrared.

The properties of fused silica will be described in more detail in Section 2.2.3. Fused silica is used extensively in the fibre optics industry, as the principal material from which many fibres are made. It has been refined to such an extent that the absorption and scattering losses are so small that light can travel many kilometres down the fibre before being fully attenuated.

The refractive index of silica in the transparency range is tabulated against the wavelength in Table 1.3. This variation of the refractive index with wavelength is called dispersion. Note that it is not a very large effect: n changes by less than 1 % over the whole visible spectral region. Note also that the dispersion is largest at the shortest wavelengths near the fundamental absorption edge. Dispersion is present in all optical materials, as will be explained in Section 2.3.

Chemicals are commonly added to silica during the fusion process to produce a whole range of other types of glasses. The presence of these additives can alter the refractive index and the transmission range. Table 1.4.2 lists the composition of a number of common glasses together with their refractive index and ultraviolet transmission. It is apparent that the additives have the effect of increasing the refractive index, at the expense of increasing the ultraviolet absorption. A high refractive index is desirable for cut-glass products, since it increases the reflectivity (see Exercise 1.2), and hence gives the glassware a more shiny appearance.

Stained glass and colour glass filters are made by adding semiconductors with band gaps in the visible spectral region during the fusion process. The properties of these coloured glasses will be discussed further in Section 1.4.5 below.

Table 1.2 Approximate transparency range, band gap wavelength λ_g , and refractive index n of a number of common semiconductors. n is measured at 10 μm . After [1], [2] and [3].

Crystal	Transparency range (μm)	λ_g (μm)	n
Ge	1.8–23	1.8	4.00
Si	1.2–15	1.1	3.42
GaAs	1.0–20	0.87	3.16
CdTe	0.9–14	0.83	2.67
CdSe	0.75–24	0.71	2.50
ZnSe	0.45–20	0.44	2.41
ZnS	0.4–14	0.33	2.20

Table 1.3 Refractive index of synthetic fused silica versus wavelength. After [2].

Wavelength (nm)	Refractive index
213.9	1.53430
239.9	1.51336
275.3	1.49591
334.2	1.47977
404.7	1.46962
467.8	1.46429
508.6	1.46186
546.1	1.46008
632.8	1.45702
706.5	1.45515
780.0	1.45367
1060	1.44968
1395	1.44583
1530	1.44427
1970	1.43853
2325	1.43293

Table 1.4 Composition, refractive index and ultraviolet transmission of common glasses. The letters after the names give the abbreviations used to identify the glass type. The composition figures are the percentage by mass. The refractive index is measured at 546.1 nm, and the transmission is for a 1 cm plate at 310 nm. After [1], [4].

Name	SiO ₂	B ₂ O ₃	Al ₂ O ₃	Na ₂ O	K ₂ O	CaO	BaO	PbO	P ₂ O ₅	<i>n</i>	<i>T</i>
Fused silica	100									1.460	0.91
Crown (K)	74			9	11	6				1.513	0.4
Borosilicate crown (BK)	70	10		8	8	1	3			1.519	0.35
Phosphate crown (PK)		3	10		12	5		70		1.527	0.46
Light flint (LF)	53			5	8		34			1.585	0.008
Flint (F)	47			2	7		44			1.607	–
Dense flint (SF)	33				5		62			1.746	–

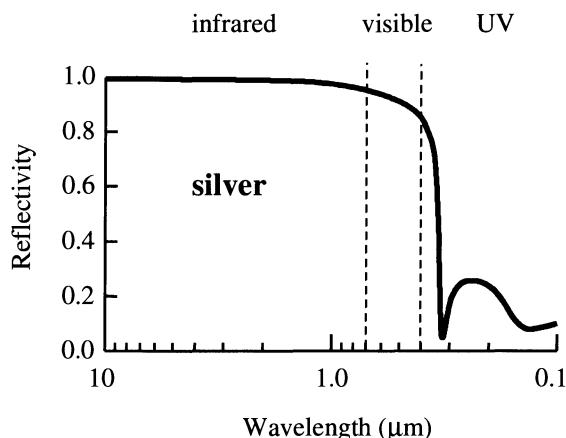


Fig. 1.5 Reflectivity of silver from the infrared to the ultraviolet. After [4].

1.4.3 Metals

The characteristic optical feature of metals is that they are shiny. This is why metals like silver and aluminium have been used for making mirrors for centuries. The shiny appearance is a consequence of their very high reflection coefficients. We will see in Chapter 7 that the high reflectivity is caused by the interaction of the light with the free electrons that are present in the metal.

Figure 1.5 shows the reflectivity of silver from the infrared spectral region to the ultraviolet. We see that the reflectivity is very close to 100 % in the infrared, and stays above 80 % throughout the whole visible spectral region. The reflectivity then drops sharply in the ultraviolet. This general behaviour is observed in all metals. There is strong reflection for all frequencies below a characteristic cut-off frequency called the plasma frequency. The plasma frequency corresponds to a wavelength in the ultraviolet spectral region, and so metals reflect infrared and visible wavelengths, but transmit ultraviolet wavelengths. This effect is called the ultraviolet transmission of metals.

Some metals have characteristic colours. Copper, for example, has a pinkish colour, while gold is yellowish. These colours are caused by interband electronic transitions that occur in addition to the free carrier effects that cause the reflection. This point will be explained in Section 7.3.2 of Chapter 7.

1.4.4 Molecular materials

The term ‘molecular material’ could in principle cover the solid phase of any molecule. However, the crystalline phase of inorganic molecules such as NaCl or GaAs are classified as insulators or semiconductors in this book, while simple organic molecules such as methane (CH_4) tend to be gases or liquids at room temperature. We therefore restrict our attention here to large organic molecules.

Some organic compounds form crystals in the condensed phase, but many others are amorphous. The solids are held together by the relatively weak van der Waals interactions between the molecules, which are themselves held together by strong covalent bonds. The optical properties of the solid therefore tend to be very similar to those of the individual molecules.

Organic compounds can be generally classified into either saturated or conjugated systems. This classification depends on the type of bonding in the molecule, and will be explained in more detail in Chapter 8.

In saturated compounds, the valence electrons are incorporated into strong, localized bonds between neighbouring atoms. This means that all the electrons are tightly held in their bonds, and can only respond at high frequencies in the ultraviolet spectral range. Saturated compounds are therefore usually colourless and do not absorb in the visible region. Their properties are generally similar to those of the glasses discussed in Section 1.4.2 above: they absorb in the infrared and ultraviolet due to vibrational and electronic transitions respectively, and are transparent in the visible. Plastics such as poly-methylmethacrylate (commonly known as ‘perspex’ or ‘plexiglass’) or poly-ethylene (polythene) are typical examples.

Conjugated molecules, by contrast, have much more interesting optical properties. The electrons from the p -like atomic states of the carbon atoms form large delocalized orbitals called π orbitals which spread out across the whole molecule. The standard example of a conjugated molecule is benzene (C_6H_6), in which the π electrons form a ring-like orbital above and below the plane of the carbon and hydrogen atoms. Further examples include the other aromatic hydrocarbons, dye molecules, and conjugated polymers.

π electrons are less tightly bound than the electrons in saturated molecules, and are optically-active at lower frequencies. In benzene the absorption edge is in the ultraviolet at 260 nm, but with other molecules the transition energy is shifted down to visible frequencies. The molecules with visible absorption also tend to emit strongly at visible frequencies. This makes them of high technological interest for applications as light-emitting devices. These are the solid state counterparts of the organic dyes that have been used in liquid lasers for several decades.

The optical processes that occur in π conjugated materials will be described in Chapter 8. By way of example, Fig. 1.6 shows the absorption spectrum of the technologically important polyfluorene-based polymer called ‘F8’. Thin film samples of this material are typically prepared by spin coating the molecules onto a glass slide. The data in Fig. 1.6 show that the polymer is transparent throughout most of the visible spectral region, but absorbs strongly at ultraviolet wavelengths. The broad absorption band which peaks at 380 nm is caused by vibrational-electronic transitions to the first singlet excited state of the

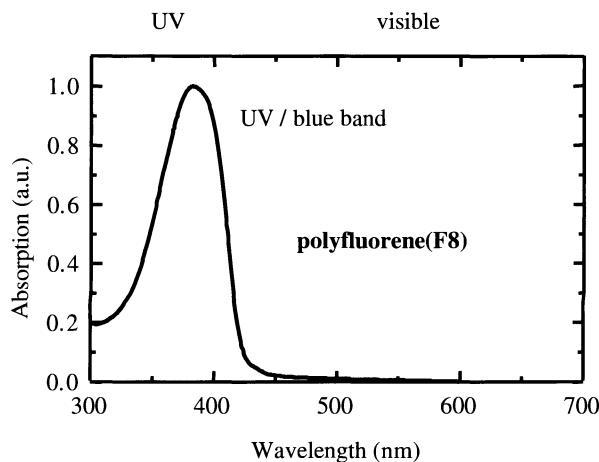


Fig. 1.6 Absorption spectrum of the polyfluorene-based polymer F8 [poly(9,9-dioctylfluorene)]. After [5], copyright 2001 Exerpta Medica Inc., reprinted with permission.

molecule. This band extends slightly into the blue spectral region, and gives the material a pale yellow colour.

Conjugated polymers such as F8 luminesce strongly when electrons are promoted into the excited states of the molecule. The luminescence is Stokes shifted to lower energy compared to the absorption, and typically occurs in the middle of the visible spectral region. An attractive feature of these organic materials is that the emission wavelength can be ‘tuned’ by small alterations to the chemical structure of the molecular units within the polymers. We will see in Section 8.6 how this property has been used to develop organic light-emitting devices to cover the full range of the visible spectral region.

1.4.5 Doped glasses and insulators

We have already mentioned in Section 1.4.2 above that colour glass filters and stained glass are made by adding appropriately chosen semiconductors to silica during the fusion process. This is a typical example of how a colourless material such as fused silica can take on new properties by controlled doping with optically active substances.

The colour of a colour glass filter can be controlled in two different ways.

- (1) The most obvious way is by variation of the composition of the dopant. For example, the glass might be doped with the alloy semiconductor $Cd_xZn_{1-x}Se$ during the fusion process, with the value of x determined by the $ZnSe : CdSe$ ratio in the original melt. The band gap of the alloy can be ‘tuned’ through the visible spectrum region by varying x , and this determines the short wavelength transmission cut-off for the filter.
- (2) The size of the semiconductor crystallites within the glass can be very small, and this can also have an effect on the colour produced. Normally, the optical properties of a material are independent of the size of the crystal, but this ceases to be the case if the dimensions are comparable to the electron wavelength. The ‘quantum size effect’ increases the energy of the electrons and hence shifts the effective band gap to higher energy. This point will be explained further in Section 6.9 of Chapter 6.

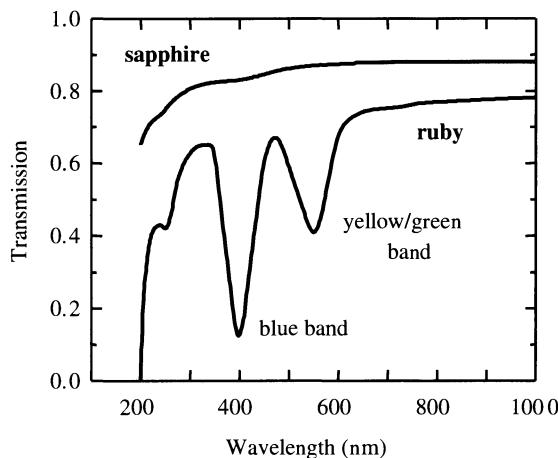


Fig. 1.7 Transmission spectrum of ruby (Al_2O_3 with 0.05 % Cr^{3+}) compared to sapphire (pure Al_2O_3). The thicknesses of the two crystals were 6.1 mm and 3.0 mm respectively. After [6], reprinted with permission.

The principle of doping optically active atoms into colourless hosts is employed extensively in the crystals used for solid state lasers. A typical example is the ruby crystal. Rubies consist of Cr^{3+} ions doped into Al_2O_3 (sapphire). In the natural crystals, the Cr^{3+} ions are present as impurities, but in synthetic crystals, the dopants are deliberately introduced in controlled quantities during the crystal growth process.

Figure 1.7 compares the transmission spectra of synthetic ruby (Al_2O_3 with 0.05 % Cr^{3+}) to that of synthetic sapphire (pure Al_2O_3). It is seen that the presence of the chromium ions produces two strong absorption bands, one in the blue spectral region and the other in the green/yellow region. These two absorption bands give rubies their characteristic red colour. The other obvious difference between the two transmission curves is that the overall transmission of the ruby is lower. This is caused in part by the increased scattering of light by the impurities in the crystal.

The optical properties of crystals like ruby will be covered in Chapter 9. We will see there that the broadening of the discrete transition lines of the isolated dopant ions into absorption bands is caused by vibronic coupling between the valence electrons of the dopant and the phonons in the host crystal. We will also see how the centre wavelength of the bands is determined by the crystal field effect, that is, the interaction between the dopant ions and electric field of the host crystal. These properties are very important in the design of solid state lasers and phosphors.

1.5 Characteristic optical physics in the solid state

The previous section has given a brief overview of the optical properties of several different classes of solid state materials. It is natural to ask whether any of these properties are exclusive to the solid state. In other words, how do the optical properties of a solid differ from those of its constituent atoms or molecules? This question is essentially the same as asking what the difference is between solid state and atomic or molecular physics.

The answer clearly depends on the type of material that we are considering. In some materials there will be a whole range of new effects associated with the solid state, while with others, the differences may not be so great. Molecular materials are an example of the second type. We would expect the absorption spectra of a solid film and that of an equivalent dilute solution to be very similar. This happens because the forces between the molecules in the condensed phase are relatively weak compared to the forces within the molecule itself. The appeal of the solid state in this case is the high number density of molecules that are present, and the possibility of incorporating them into solid state electronic devices.

With many other materials, however, there will be substantial differences between the condensed phase and the gaseous or liquid state. It is obviously not possible to give a full catalogue of these effects in an introductory chapter such as this one. Instead, we will highlight here five aspects that make the physics of the solid state interesting and different, namely

- Crystal symmetry
- Electronic bands
- Vibronic bands
- The density of states
- Delocalized states and collective excitations.

There are many others, of course, but these themes occur over and over again and are therefore worth considering briefly in themselves before we start going into the details.

1.5.1 Crystal symmetry

Most of the materials that we will be studying occur as crystals. Crystals have long range **translational order**, and can be categorized into 32 classes according to their **point group symmetry**. The point group symmetry refers to the group of symmetry operations that leaves the crystal invariant. Examples of these include rotations about particular axes, reflections about planes, and inversion about points in the unit cell. Some crystal classes such as the cubic ones possess a very high degree of symmetry. Others have much lower symmetry.

The link between the measurable properties and the point group symmetry of a crystal can be made through **Neumann's principle**. This states that:

Any macroscopic physical property must have at least the symmetry of the crystal structure.

For example, if a crystal has four-fold rotational symmetry about a particular axis, then we must get the same result in any experiment we might perform in the four equivalent orientations.

It is instructive to compare the properties of a crystal to those of the atoms from which it has been formed. A gas of atoms has no translational order. Therefore we expect to find new effects in the solid state that reflect its translational symmetry. The formation of electronic bands and delocalized states discussed in Sections 1.5.2 and 1.5.5 below are examples of this. At the same time, the point group symmetry of a crystal is lower than that of the individual

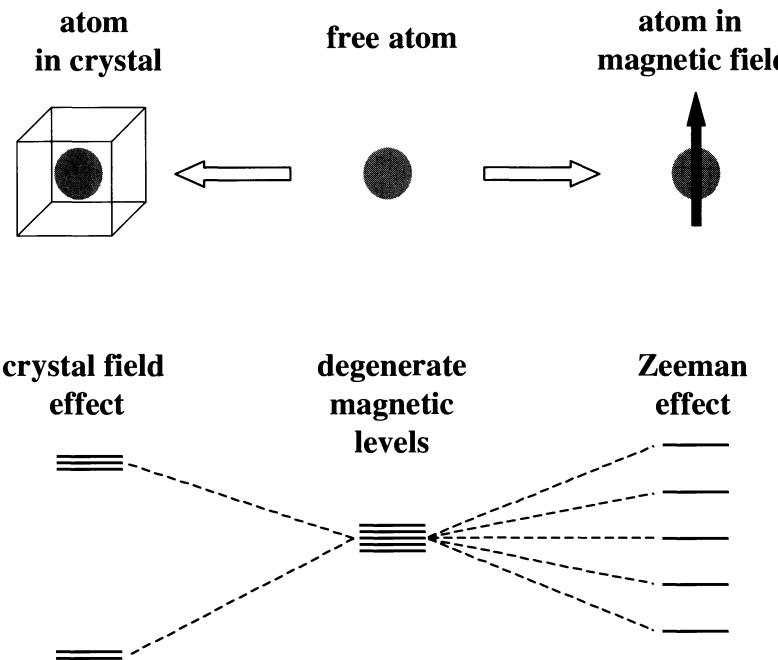


Fig. 1.8 Splitting of the magnetic levels of a free atom by the crystal field effect. In the free atoms, the magnetic levels are degenerate. We must apply a magnetic field to split them by the Zeeman effect. However, the magnetic levels can be split even without applying an external magnetic field in a crystal. The details of the way the levels split are determined by the symmetry class of the crystal.

atoms, which have the highest possible symmetry due to their spherical invariance. We therefore expect to find other effects in the solid state that relate to the lowering of the symmetry on going from free atoms to the particular point group of the crystal class. Two specific examples of this are discussed briefly here, namely **optical anisotropy** and the **lifting of degeneracies**.

A crystal is said to be anisotropic if its properties are not the same in all directions. Anisotropy is only found in the solid state, because gases and liquids do not have any preferred directions. The degree of anisotropy found in a crystal depends strongly on the point group symmetry that it possesses. In cubic crystals, for example, the optical properties must be the same along the x , y and z axes because they are physically indistinguishable. On the other hand, in a uniaxial crystal, the properties along the optic axis will be different from those along the axes at right angles to it. The optical anisotropy is manifested by the property of birefringence which is discussed in Section 2.4. It is also important for the description of the nonlinear optical coefficients of crystals discussed in Chapter 11.

The lifting of degeneracies by reduction of the symmetry is a well-known effect in atomic physics. Free atoms are spherically symmetric and have no preferred directions. The symmetry can be broken by applying an external magnetic or electric field which creates a preferred axis along the field direction. This can lead to the lifting of certain level degeneracies that are present in the free atoms. The Zeeman effect, for example, describes the splitting of degenerate magnetic levels when a magnetic field is applied. If the same atom is introduced into a crystal, it will find itself in an environment with a point group symmetry determined by the lattice. This symmetry is lower than that of the free atom, and therefore some level degeneracies can be lifted.

This point is illustrated schematically in Fig. 1.8, which shows how the magnetic levels of a free atom can be split by the crystal field effect in an analogous way to the Zeeman effect. The splitting is caused by the interaction of the orbitals of the atoms with the electric fields of the crystalline environment. The details do not concern us here. The important point is that the splittings are determined by the symmetry class of the crystal and do not require an external field. Optical transitions between these crystal-field split levels often occur in the visible spectral region, and cause the material to have very interesting properties that are not found in the free atoms. These effects will be explored in more detail in Chapter 9.

Before closing this section on crystal symmetry, it is worth pointing out that many important solid state materials do not possess long range translational symmetry. Glass is an obvious example. Other examples include thin molecular films such as light-emitting polymers sputtered onto substrates, and amorphous silicon. The optical properties of these materials may be very similar to those of their constituent atoms or molecules. Their importance is usually related to the convenience of the solid phase rather than to new optical properties that relate to the solid state physics.

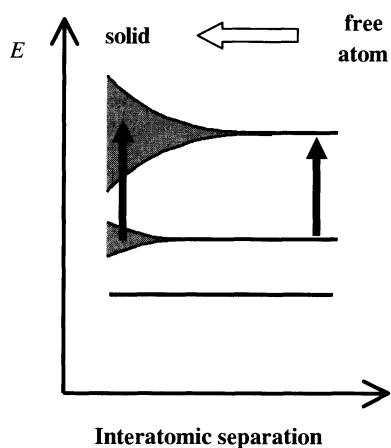


Fig. 1.9 Schematic diagram of the formation of electronic bands in a solid from the condensation of free atoms. As the atoms are brought closer together to form the solid, their outer orbitals begin to overlap with each other. These overlapping orbitals interact strongly, and broad bands are formed. The inner core orbitals do not overlap and so remain discrete even in the solid state. Optical transitions between the bands can occur, and this causes strong absorption over a continuous range of frequencies rather than discrete lines.

1.5.2 Electronic bands

The atoms in a solid are packed very close to each other, with the interatomic separation approximately equal to the size of the atoms. Hence the outer orbitals of the atoms overlap and interact strongly with each other. This broadens the discrete levels of the free atoms into bands, as illustrated schematically in Fig. 1.9.

The electron states within the bands are delocalized and possess the translational invariance of the crystal. **Bloch's theorem** states that the wave functions should be written in the form:

$$\psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}), \quad (1.27)$$

where $u_{\mathbf{k}}(\mathbf{r})$ is a function that has the periodicity of the lattice. The Bloch states described by eqn 1.27 are modulated plane waves. Each electronic band has a different envelope function $u_{\mathbf{k}}(\mathbf{r})$ which retains some of the atomic character of the states from which the band was derived.

Optical transitions can occur between the electronic bands if they are allowed by the selection rules. This ‘interband’ absorption is possible over a continuous range of photon energies determined by the lower and upper energy limits of the bands. This contrasts with the absorption spectra of free atoms, which consist of discrete lines. The observation of broad bands of absorption rather than discrete lines is one of the characteristic features of the solid state.

Interband transitions will be discussed at length in a number of chapters in this book, most notably Chapters 3 and 5. The absorption strength is usually very high because of the very large density of absorbing atoms in the solid. This means that we can produce sizeable optical effects in very thin samples, allowing us to make the compact optical devices that form the basis of the modern optoelectronics industry.

1.5.3 Vibronic bands

The electronic states of the atoms or molecules in a solid may be strongly coupled to the vibrational modes of the crystal through the vibronic interaction. A typical example of where this effect occurs is the doped insulator crystals introduced in Section 1.4.5. The vibronic coupling broadens the discrete electronic states of the isolated dopant atoms into bands. This has the effect of broadening the discrete absorption and emission lines of the atoms into continuous bands. These vibronic effects will be described in more detail in Chapter 9.

It is important to realize that the reason for the formation of the vibronic bands is different to that for the electronic bands considered in the previous section. In the case of vibronic bands, the continuum of states arises from the coupling of discrete electronic states to a continuous spectrum of vibrational modes. This contrasts with the electronic bands, where the continuum arises from interactions between electronic states of neighbouring atoms.

Vibronic effects are also observed in molecular materials. This is an interesting case which highlights the difference between the solid state and the liquid or gaseous phase. The absorption spectra of simple free molecules also show vibrational–electronic bands, but the transition frequencies are discrete because both the electronic energies and the vibrational energies are discrete. In molecular solids, by contrast, the vibrational frequencies are continuous, and this causes continuous absorption and emission spectra.

1.5.4 The density of states

The concept of the **density of states** is an inevitable corollary of band formation in solids. The electronic and vibrational states of free molecules and atoms have discrete energies, but this is not the case in a solid: both the electronic states and the phonon modes have a continuous range of energies. This continuum of states leads to continuous absorption and emission bands, as has already been stressed in the previous two sections.

The number of states within a given energy range of a band is conveniently expressed in terms of the density of states function $g(E)$. This is defined as:

$$\text{Number of states in the range } E \rightarrow (E + dE) = g(E) dE . \quad (1.28)$$

$g(E)$ is worked out in practice by first calculating the density of states in momentum space $g(k)$, and then using the relationship between $g(E)$ and $g(k)$, namely:

$$g(E) = g(k) \frac{dk}{dE} . \quad (1.29)$$

This can be evaluated from knowledge of the E - k relationship for the electrons or phonons. Knowledge of $g(E)$ is crucial for calculating the absorption and emission spectra due to interband transitions and also for calculating the shape of vibronic bands.

1.5.5 Delocalized states and collective excitations

The fact that the atoms in a solid are very close together means that it is possible for the electron states to spread over many atoms. The wave functions

of these delocalized states possess the underlying translational symmetry of the crystal. The Bloch waves described by eqn 1.27 are a typical example. The delocalized electron waves move freely throughout the whole crystal and interact with each other in a way that is not possible in atoms. The delocalization also allows collective excitations of the whole crystal rather than individual atoms. Two examples that we will consider in this book are the excitons formed from delocalized electrons and holes in a semiconductor, and the plasmons formed from free electrons in metals and doped semiconductors. These collective excitations may be observed in optical spectra, and have no obvious counterpart in the spectra of free atoms. These excitonic effects will be discussed in Chapter 4, while plasmons are covered in Section 7.5.

Other wave-like excitations of the crystal are delocalized in the same way as the electrons. In the case of the lattice vibrations, the delocalized excitations are described by the phonon modes. We have already mentioned above that the phonon frequencies are continuous, which contrasts with the discrete vibrational frequencies of molecules. Some optical effects related to phonons have direct analogies with the vibrational phenomena observed in isolated molecules but others are peculiar to the solid state. Examples of the former are Raman scattering and infrared absorption. Examples of the latter include the phonon-assisted interband transitions in semiconductors with indirect band gaps (cf. Section 3.4), and the broadening of the discrete levels of impurity atoms into continuous vibronic bands by interactions with phonons as discussed in Chapter 9.

The delocalized states of a crystal are described by quantum numbers such as \mathbf{k} and \mathbf{q} which have the dimensions of inverse length. These quantum numbers follow from the translational invariance, and are therefore a fundamental manifestation of the crystal symmetry. To all intents and purposes, the quantum numbers like \mathbf{k} and \mathbf{q} behave like the wave vectors of the excitations, and they will be treated as such whenever we encounter them in derivations. However, it should be borne in mind that this is really a consequence of the deep underlying symmetry which is unique to the solid state.

1.6 Microscopic models

In the following chapters we will be developing many microscopic models to explain the optical phenomena that are observed in the solid state. The types of models will obviously vary considerably, but they can all be classified into one of the following three general categories:

- Classical
- Semiclassical
- Fully quantum.

These approaches get progressively more difficult, and so we usually apply them in the order listed above.

In the classical approach we treat both the medium and the light according to classical physics. The dipole oscillator model described in Chapter 2 is a typical example. This model is the basic starting point for understanding the general optical properties of a medium, and in particular for describing the

main effects due to free electrons (Chapter 7) and phonons (Chapter 10). We will also use it as a starting point for the discussion of nonlinear optics in Chapter 11. It would be a mistake to undervalue the classical approach in this modern day and age. The value of more sophisticated models will only be appreciated fully once the classical physics has been properly understood.

In semiclassical models we apply quantum mechanics to the atoms, but treat the light as a classical electromagnetic wave. The treatment of interband absorption in Chapter 3 is a typical example. The absorption coefficient is calculated using Fermi's golden rule, which requires knowledge of the wave functions of the quantized levels of the atoms, but treats the light-matter interaction as that between a quantized atom and a classical electric field wave. This semiclassical approach is used extensively throughout the book. Appendix B summarizes the main results that will be needed.

The final approach is the full quantum treatment. This is the realm of **quantum optics**, where both the atoms and the light are treated quantum mechanically. We use this approach implicitly whenever we refer to the light as a beam of photons and draw Feynman diagrams to represent the interaction processes that are occurring. This might give the impression that the explanations we are giving are fully quantum because we speak in terms of photons interacting with atoms. However, in the equations used to describe the process quantitatively, the light is treated classically and only the atoms are quantized. The quantitative description is therefore only semiclassical. The use of the fully quantum approach at the quantitative level is beyond the scope of this present book.

Chapter summary

- The propagation of light through a medium is quantified by the complex refractive index \tilde{n} . The real part of \tilde{n} determines the velocity of light in the medium, while the imaginary part determines the absorption coefficient. Beer's law (eqn 1.4) shows that the intensity of light in an absorbing medium decays exponentially.
- Reflection occurs at the interface between two optical materials with different refractive indices. The coefficient of reflectivity can be calculated from the complex refractive index using eqn 1.26.
- The transmission of a sample is determined by the reflectivities of the surfaces and the absorption coefficient through eqn 1.6.
- The complex refractive index is related to the complex dielectric constant through eqn 1.19. The relationships between the real and imaginary parts of \tilde{n} and $\tilde{\epsilon}_r$ are given in eqns 1.20–1.25.
- Luminescent materials re-emit light by spontaneous emission after absorbing photons. The frequency shift between the emission and absorption is called the Stokes shift.
- Scattering causes beam attenuation in accordance with Beer's law. The scattering is called elastic if the frequency is unchanged, and inelastic otherwise.

- The optical spectra of solid state materials usually consist of broad bands rather than sharp lines. The bands arise either from electronic interactions between neighbouring atoms or from vibronic coupling to the phonon modes.
- Insulators and glasses have vibrational absorption at infrared wavelengths and electronic absorption in the ultraviolet spectral region. They are transparent and colourless in the visible spectral region between these two absorption bands. In semiconductors and molecular materials the electronic absorption occurs at lower frequencies in the near infrared or visible spectral region.
- The free carriers present in metals make them highly reflective in the infrared and visible spectral regions. The colouration of some metals is caused by electronic interband absorption.
- The addition of optically active dopants to a colourless host crystal or glass produces the characteristic colours of stained glasses and gemstones.
- Crystals have both translational symmetry and point group symmetry. The consequences of the point group symmetry for the optical properties are determined by Neumann's principle.

Further reading

A good general discussion of the optical properties of materials can be found in Hecht (1998). A more advanced treatment may be found in Born and Wolf (1999). The introduction to the optical properties of various materials given in Section 1.4 will be expanded in subsequent chapters, where suitable further reading will be suggested.

The relationship between the optical properties and the complex refractive index and dielectric constant is discussed in most texts on electromagnetism, for example, Bleaney and Bleaney (1976), or Lorrain, Corson and Lorrain (2000). This material is also covered in Born and Wolf (1999).

A classic discussion of the effects of the point group symmetry on the physical properties of crystals is given in Nye (1957).

References

- [1] Driscoll, W.G. and Vaughan, W. (1978). *Handbook of optics*. McGraw-Hill, New York.
- [2] Kaye, G.W.C and Laby, T.H. (1986). *Tables of physical and chemical constants* (15th edn). Longman Scientific, Harlow, Essex.
- [3] Madelung, O. (1996). *Semiconductors, basic data* (2nd edn). Springer-Verlag, Berlin.
- [4] Lide, D.R. (1996). *CRC handbook of chemistry and physics* (77th edn). CRC Press, Boca Raton.

- [5] Buckley, A.R., Rahn, M.D., Hill, J., Cabanillas-Gonzales, J., Fox, A.M. and Bradley, D.D.C. (2001). *Chem. Phys. Lett.*, **339**, 331.
- [6] McCarthy, D.E. (1967). *Applied Optics*, **6**, 1896.

Exercises

- (1.1) Crown glass has a refractive index of 1.51 in the visible spectral region. Calculate the reflectivity of the air–glass interface, and the transmission of a typical glass window.
- (1.2) Use the data in Table 1.4.2 to calculate the ratio of the reflectivities of fused silica and dense flint glass.
- (1.3) The complex dielectric constant of the semiconductor cadmium telluride is given by $\tilde{\epsilon}_r = 8.92 + i 2.29$ at 500 nm. Calculate for CdTe at this wavelength: the phase velocity of light, the absorption coefficient and the reflectivity.
- (1.4) The detectors used in optical fibre networks operating at 850 nm are usually made from silicon, which has an absorption coefficient of $1.3 \times 10^5 \text{ m}^{-1}$ at this wavelength. The detectors have coatings on the front surface that makes the reflectivity at the design wavelength negligibly small. Calculate the thickness of the active region of a photodiode designed to absorb 90 % of the light.
- (1.5) GaAs has a refractive index of 3.68 and an absorption coefficient of $1.3 \times 10^6 \text{ m}^{-1}$ at 800 nm. Calculate the transmission coefficient and optical density of a $2 \mu\text{m}$ thick GaAs sample.
- (1.6) Sea water has a refractive index of 1.33 and absorbs 99.8 % of red light of wavelength 700 nm in a depth of 10 m. What is its complex dielectric constant at this wavelength?
- (1.7) How would you expect the absorption coefficient of a yellow colour glass filter to vary with wavelength?
- (1.8) A beam of light is incident on a parallel-sided plate of thickness l as shown in Fig. 1.10. The reflectivity of the front and back surfaces is R and the absorption coefficient is α .
- Show that the intensity of the beam exiting the sample after having been reflected from the back surface once is smaller than that of the beam that has suffered no reflections by a factor $R^2 e^{-2\alpha l}$.
 - Calculate this ratio for a transparent glass window with a refractive index of 1.5.
 - Repeat part (ii), but calculate the ratio of the electric fields of the beams rather than their intensities.

- (iv) Explain why the ratio of the fields might be important rather than the ratio of the intensities.

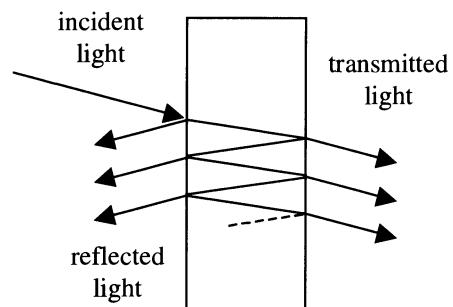


Fig. 1.10 Multiple reflections in a parallel-sided plate.

- (1.9) Repeat Exercise 1.8(ii) and (iii) for the GaAs plate considered in Exercise 1.5.
- (1.10) Show that the optical density (O.D.) of a sample is related to its transmission T and reflectivity R through:
- $$\text{O.D.} = -\log_{10}(T) + 2\log_{10}(1 - R).$$
- Hence explain how you would determine the optical density by making two transmission measurements, one at wavelength λ where the material absorbs, and the other at a wavelength λ' where the material is transparent.
- (1.11) The refractive index of fused silica is 1.45248 at 850 nm and 1.44427 at 1500 nm. Calculate the difference in time taken for short light pulses at 850 nm and 1500 nm to propagate down a silica optical fibre of length 1 km.
- (1.12) The complex dielectric constant of a metal at infrared frequencies is given by
- $$\tilde{\epsilon}_r = \epsilon_r + i \frac{\sigma}{\epsilon_0 \omega},$$
- where ϵ_r is the static relative dielectric constant, σ is the electrical conductivity, and ω is the angular frequency. (See eqn A.45 in Appendix A with $\mu_r = 1$.) Estimate the reflectivity of a silver mirror at a wavelength of $100 \mu\text{m}$. Assume that $\epsilon_2 \gg \epsilon_1$, and that the conductivity of silver is $6.6 \times 10^7 \Omega^{-1}\text{m}$.

- (1.13) Estimate the distance over which the light intensity falls by a factor of 2 in a gold film at a wavelength of $100\text{ }\mu\text{m}$. The electrical conductivity of gold is $4.9 \times 10^7\text{ }\Omega^{-1}\text{ m}$. Make the same assumptions as in the previous question.
- (1.14) The data shown in Fig. 1.5 indicates that the reflectivity of silver is close to zero at around 320 nm. What is the approximate value of the complex dielectric constant at this wavelength?
- (1.15) A neodymium laser crystal absorbs photons at 850 nm and luminesces at 1064 nm. The efficiency of the luminescence process is quantified in terms of the radiative quantum efficiency η_R , which is defined as the fraction of the atoms that emit a photon after absorbing a photon.
- Calculate the amount of energy dissipated as heat in each emission process.
 - If the total power absorbed at 850 nm is 10 W, calculate the power emitted at 1064 nm if $\eta_R = 100\text{ \%}$. How much power is dissipated as heat in the crystal?
 - Repeat part (ii) for a crystal with $\eta_R = 50\text{ \%}$.
- (1.16) A photon of wavelength 514 nm is scattered inelastically from an NaCl crystal by exciting a phonon of frequency $7.92 \times 10^{12}\text{ Hz}$. By applying the law of conservation of energy to the scattering process, calculate the wavelength of the scattered photon.
- (1.17) A certain optical fibre transmits 10 % of the light coupled into it at 850 nm. Calculate the transmission of the same fibre at 1550 nm, on the assumption that the dominant loss is Rayleigh scattering from inhomogeneities in the fibre. Hence explain why telecommunications companies use a wavelength of 1550 nm for their long distance optical fibre networks instead of the wavelength of 850 nm used for local area networks.
- (1.18) Calculate the distance over which the intensity falls to 50 % of its original value in a medium which contains 10^{16} m^{-3} scattering centres with $\sigma_s = 2 \times 10^{-17}\text{ m}^2$. Calculate the equivalent distance at half the wavelength, on the assumption that the Rayleigh scattering law applies.
- (1.19) Explain why ice is birefringent, but water is not.

Classical propagation

2

The propagation of light through an optical medium was discussed in general terms in Sections 1.1–1.3 of Chapter 1. We saw there that the propagation is characterized by two parameters, namely the refractive index and the absorption coefficient. In this chapter we will investigate the classical theory of optical propagation, in which the light is treated as electromagnetic waves and the atoms or molecules are modelled as classical dipole oscillators. We will see that this model gives a good general overview of the optical properties, and enables us to calculate the frequency dependence of the complex dielectric constant. This gives us the frequency dependence of the absorption coefficient and refractive index, and hence enables us to explain the phenomenon of dispersion. We will also see that the model provides the framework for describing the effects due to optical anisotropy such as birefringence.

The treatment given here presupposes a working knowledge of the electromagnetic properties of dielectrics. A summary of the main results that we will use is given in Appendix A. The model will be revisited in subsequent chapters when we consider the optical properties of free electrons in Chapter 7, and when we discuss lattice vibrations in Chapter 10. The model is also the starting point for the treatment of nonlinear optical effects in Chapter 11.

2.1	Propagation of light in a dense optical medium	25
2.2	The dipole oscillator model	29
2.3	Dispersion	40
2.4	Optical anisotropy: birefringence	42

2.1 Propagation of light in a dense optical medium

The classical model of light propagation was developed at the end of the nineteenth century following Maxwell's theory of electromagnetic waves and the introduction of the concept of the dipole oscillator. In this section we will give a qualitative discussion of the physical assumptions of this model, leaving the quantitative calculation to the next section.

The model assumes that there are several different types of oscillators within a medium, each with their own characteristic resonant frequency. At optical frequencies the most important contribution is from the oscillations of the **bound electrons** within the atoms, and so we begin this section by considering atomic oscillators. We then go on to introduce the idea of **vibrational oscillators**, which resonate at lower frequencies in the infrared spectral region, and finally mention **free electron oscillators**, which are responsible for the principal optical properties of metals.

2.1.1 Atomic oscillators

The concept of the dipole oscillator was introduced soon after Maxwell's electromagnetic theory. It was shown theoretically that an oscillating electric dipole would emit electromagnetic waves, and this was confirmed in 1887 when Heinrich Hertz succeeded in generating and detecting radio waves in the laboratory. He used an oscillatory discharge across a spark gap as the source and a wire loop as the aerial of the detector. This was an elegant confirmation of the validity of Maxwell's electromagnetic theory, and the beginning of radio telecommunications.

The idea of considering atoms as oscillating dipoles was originally proposed by Henrick Antoon Lorentz in 1878, thus preceding Hertz's demonstration by several years. It was known that atoms emit and absorb at discrete frequencies, and Lorentz's model provided a simple explanation for these observations in terms of the newly discovered electromagnetic theories.

The oscillator model of the atom is illustrated schematically in Fig. 2.1. It is assumed that the electron is held in a stable orbit with respect to the nucleus, and the spring represents the restoring force for small displacements from the equilibrium. The negatively charged electron and the positively charged nucleus form an electric dipole with a magnitude proportional to their separation. Lorentz, of course, could not have known about electrons and nuclei, because they were not discovered until 1897 and 1911 by J.J. Thomson and Ernest Rutherford respectively. Lorentz simply postulated the existence of dipoles without knowing their origin.

The natural resonant frequency ω_0 of the atomic dipoles is determined by their mass and the magnitude of the restoring force experienced for small displacements. The appropriate mass is the **reduced mass** given by:

$$\frac{1}{\mu} = \frac{1}{m_0} + \frac{1}{m_N}, \quad (2.1)$$

where m_0 and m_N are the masses of the electron and nucleus respectively. Since $m_N \gg m_0$, we may safely take $\mu \approx m_0$ here. The restoring force is quantified in terms of a spring constant K_s , which is chosen so that ω_0 coincides with one of the natural frequencies of the atoms (see Exercise 2.1):

$$\omega_0 = \sqrt{\frac{K_s}{\mu}}. \quad (2.2)$$

We have to suppose that there are several dipoles within every atom, to account for the fact that a given atom has many transition frequencies. These are known from the absorption and emission spectra, and the frequencies occur in the near-infrared, visible and ultraviolet spectral regions (10^{14} – 10^{15} Hz).

We can understand the connection between the atomic dipoles and the emission spectra by considering the oscillations of the dipole shown in Fig. 2.2. An electric dipole consists of a positive charge $+q$ at position \mathbf{r}_+ and a negative charge $-q$ at \mathbf{r}_- . The electric dipole moment is defined as

$$\mathbf{p} = q(\mathbf{r}_+ - \mathbf{r}_-). \quad (2.3)$$

Hence the positive nucleus and negative electron form a dipole with magnitude equal to $e|\mathbf{r}_N - \mathbf{r}_e|$.

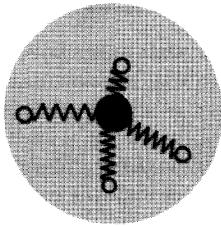


Fig. 2.1 Classical model of the bound electrons in an atom. The electrons are represented by the open circles, while the black circle at the centre of the atom represents the nucleus. The electrons are held to the heavy nucleus by springs which represent the restoring forces due to the binding between them. Each atom has a series of characteristic resonant frequencies which we now know to correspond with the quantized transition energies.

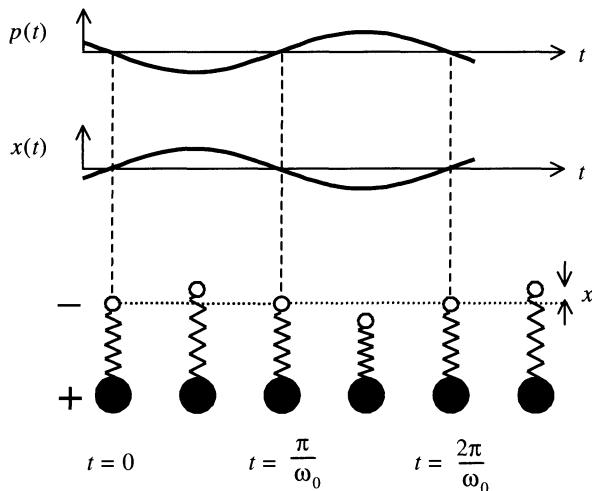


Fig. 2.2 Oscillations of a classical dipole consisting of a heavy positive charge and a light negative charge bound together by a spring. $x(t)$ is the time-dependent displacement of the negative charge from its equilibrium position. The natural vibrations of the dipole about the equilibrium length at frequency ω_0 generate a time dependent dipole moment $p(t)$ as indicated in the top half of the figure.

During the oscillations of the atomic dipole, the nucleus remains more or less stationary due to its heavy mass, while the electron oscillates backwards and forwards at frequency ω_0 . Hence the oscillations produce a time varying dipole in addition to any permanent dipole the atom may have. The magnitude of the time varying dipole is given by:

$$p(t) = -ex(t), \quad (2.4)$$

where $x(t)$ is the time varying displacement of the electron from its equilibrium position. This connection between the electron displacement and the time dependent atomic dipole is illustrated in the top half of Fig. 2.2. The oscillating dipole radiates electromagnetic waves at frequency ω_0 , in accordance with the theory of classical Hertzian dipoles. Hence the atom is expected to radiate light at its resonant frequency whenever sufficient energy is imparted to it to excite the oscillations.

We can also use the dipole model to understand how the atom interacts with an external electromagnetic wave at frequency ω . The AC electric field exerts forces on the electron and the nucleus and drives oscillations of the system at frequency ω . If ω coincides with one of the natural frequencies of the atom, then we have a resonance phenomenon. This induces very large amplitude oscillations, and transfers energy from the external wave to the atom. The atom can therefore absorb energy from the light wave if $\omega = \omega_0$. The absorption strength is characterized by the absorption coefficient α , and the intensity of the wave will decay exponentially according to Beer's law (eqn 1.4).

We now know from quantum theory that what actually happens during absorption is that the atom jumps to an excited state by absorbing a photon. This can only occur if $\hbar\omega = E_2 - E_1$, where E_1 and E_2 are the quantized energies of the initial and final states. Once it has been excited, the atom can return to the ground state by a series of radiationless transitions, in which case the energy from the absorbed photon is ultimately converted into heat. Alternatively, it can luminesce by re-emitting a photon at some later time. The re-radiated photons are incoherent with each other and are emitted in all directions rather than in the specific direction of the incoming wave. Hence there is a net decrease in

We assume here that the forces exerted by the electric fields are very small compared to the binding forces that hold the electrons to the nucleus. This approximation may not be valid if we are using a very powerful laser beam to excite the medium. If this is the case, then we are working in the regime of nonlinear optics. These effects are considered in Chapter 11.

the energy flow in the beam direction, which is equivalent to absorption.

If ω does not coincide with any of the resonant frequencies, then the atoms will not absorb, and the medium will be transparent. In this situation the light wave drives non-resonant oscillations of the atoms at its own frequency ω . The oscillations of the atoms follow those of the driving wave, but with a phase lag. The phase lag is a standard feature of forced oscillators and is caused by damping. (See Exercise 2.2.) The oscillating atoms all re-radiate instantaneously, but the phase lag acquired in the process accumulates through the medium and retards the propagation of the wave front. This implies that the propagation velocity is smaller than in free space. The reduction of the velocity in the medium is characterized by the refractive index defined in eqn 1.2.

The slowing of the wave due to the non-resonant interactions can be considered as a repeated scattering process. The scattering is both coherent and elastic, and each atom behaves like a Huygens point source. The scattered light interferes constructively in the forward direction, and destructively in all other directions, so that the direction of the beam is unchanged by the repetitive scattering process. However, each scattering event introduces a phase lag which causes a slowing of the propagation of the phase front through the medium.

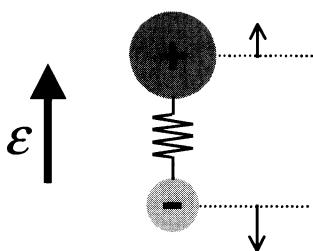


Fig. 2.3 Classical model of a polar molecule. The atoms are positively and negatively charged, and can vibrate about their equilibrium separation. These vibrations produce an oscillating electric dipole which will radiate electromagnetic waves at the resonant frequency. Alternatively, the molecule will interact with the electric field \mathbf{E} of a light wave through the forces exerted on the charged atoms.

2.1.2 Vibrational oscillators

An optical medium may contain other types of dipole oscillators in addition to those originating from the bound electrons within the atoms. If the medium is ionic, it will contain oppositely charged ions. Vibrations of these charged atoms from their equilibrium positions within the crystal lattice will produce an oscillating dipole moment, in exactly the same way as the oscillations of the electrons within the individual atoms that we considered above. Therefore, we must also consider the optical effects due to these vibrational oscillators when we consider the interaction of light with an ionic optical medium.

The optical effects of vibrational oscillators are well known in molecular physics. Figure 2.3 gives a schematic illustration of a classical polar molecule. This consists of two charged atoms bound together in a stable configuration, with the spring representing the molecular bond between them. The charged atoms can vibrate about their equilibrium positions and induce an oscillating electric dipole in an analogous way to the bound electrons in the atoms. We see immediately from eqn 2.2 that the vibrations will occur at lower frequencies because the reduced mass is larger. The vibrations therefore occur at infrared frequencies with $\omega/2\pi \sim 10^{12}\text{--}10^{13}$ Hz. These molecular vibrations are associated with strong absorption lines in the infrared spectral region.

The interaction between the vibrations of the molecule and the light wave occurs through the forces exerted on the atoms by the electric field. It is obvious that this can only happen if the atoms are charged. This is why we specified that the molecule was **polar** in the preceding paragraph. A polar molecule is one in which the electron charge cloud that forms the bond sits closer to one of the atoms than to the other. Ionic molecules like the alkali halides (e.g. Na^+Cl^-) clearly fall into this category, while purely covalent ones such as the elemental molecules (e.g. O_2) do not. Many other molecules fall somewhere between these two limits. Water (H_2O) is a well known example. Oxygen has

a greater electron affinity than hydrogen, and so the valence electrons in the O–H bond sit closer to the oxygen atoms. The two hydrogen atoms therefore possess a small positive charge which is balanced by a negative charge of twice the magnitude on the oxygen atom.

In a crystalline solid formed from the condensation of polar molecules, the atoms are arranged in an alternating sequence of positive and negative ions. The ions can vibrate about their equilibrium positions, and this produces oscillating dipole waves. These oscillations are associated with **lattice vibrations**, and they occur at frequencies in the infrared spectral region. We will consider the optical properties related to the lattice vibrations in detail in Chapter 10. We will see there that the light–matter interaction is associated with the excitation of **phonons**, which are quantized lattice waves. At this stage, we simply note that the lattice vibrations of a polar crystal give rise to strong optical effects in the infrared spectral region. These effects occur in addition to those due to the bound electrons of the atoms that comprise the crystal. In practice we can treat these two types of dipoles separately because the resonances are sharp and they occur at very different frequencies. Therefore the resonant effects of the bound electrons are negligible at the frequencies of the lattice vibrations, and *vice versa*. This point will be considered in more detail in Section 2.2.2.

2.1.3 Free electron oscillators

The electronic and vibrational dipoles considered above are both examples of bound oscillators. Metals and doped semiconductors, by contrast, contain significant numbers of **free electrons**. As the name implies, these are electrons that are not bound to any atoms, and therefore do not experience any restoring forces when they are displaced. This implies that the spring constant in eqn 2.2 is zero, and hence that the natural resonant frequency $\omega_0 = 0$.

The free electron model of metals is attributed to Paul Drude, and so the application of the dipole oscillator model to free electron systems is generally called the **Drude–Lorentz model**. The dipole oscillator model is perfectly valid, except that we must set $\omega_0 = 0$ throughout. The optical properties of free electron systems will be discussed in Chapter 7.

2.2 The dipole oscillator model

In the previous section we introduced the general assumptions of the dipole oscillator model. We now want to use the model to calculate the frequency dependence of the refractive index and absorption coefficient. This will provide a simple explanation for the dispersion of the refractive index in optical materials, and will also illustrate a very general point that the phenomena of absorption and refraction are related to each other.

2.2.1 The Lorentz oscillator

We consider the interaction between a light wave and an atom with a single resonant frequency ω_0 due to the bound electrons, as given by eqn 2.2. We

We know from experimental observations that atoms must have many natural resonant frequencies to account for the multiplicity of lines in the absorption and emission spectra. However, the salient features of the physical behaviour are well illustrated by a singly resonant system, and the inclusion of multiple resonances complicates the discussion without adding much to the physical understanding at this stage. We therefore postpone the discussion of the effects of multiple resonances to subsection 2.2.2 below.

model the displacement of the atomic dipoles as damped harmonic oscillators. The inclusion of damping is a consequence of the fact that the oscillating dipoles can lose their energy by collisional processes. In solids, this would typically occur through an interaction with a phonon which has been thermally excited in the crystal. As we will see, the damping term has the effect of reducing the peak absorption coefficient and broadening the absorption line.

The electric field of the light wave induces forced oscillations of the atomic dipole through the driving forces exerted on the electrons. We make the assumption that $m_N \gg m_0$ here so that we can ignore the motion of the nucleus. The displacement x of the electron is governed by an equation of motion of the form:

$$m_0 \frac{d^2x}{dt^2} + m_0 \gamma \frac{dx}{dt} + m_0 \omega_0^2 x = -e\mathcal{E}, \quad (2.5)$$

where γ is the damping rate, e is the magnitude of the electric charge of the electron, and \mathcal{E} is the electric field of the light wave. The terms on the left hand side represent the acceleration, the damping and the restoring force respectively. The damping is modelled by a frictional force which is proportional to the velocity and impedes the motion. The term on the right hand side represents the driving force due to the AC electric field of the light wave.

We consider the interaction of the atom with a monochromatic light wave of angular frequency ω . The time dependence of the electric field is given by

$$\mathcal{E}(t) = \mathcal{E}_0 \cos(\omega t + \Phi) = \mathcal{E}_0 \Re e \left(\exp(-i\omega t - \Phi) \right), \quad (2.6)$$

where \mathcal{E}_0 is the amplitude and Φ is the phase of the light. In order to keep consistency with the sign convention introduced later, we have chosen to take the negative frequency part of the complex exponential.

The AC electric field will drive oscillations at its own frequency ω . We therefore substitute eqn 2.6 into eqn 2.5 and look for solutions of the form:

$$x(t) = X_0 \Re e \left(\exp(-i\omega t - \Phi') \right), \quad (2.7)$$

where X_0 and Φ' are the amplitude and phase of the oscillations. We can incorporate the phase factors of eqns 2.6 and 2.7 into the amplitudes by allowing both \mathcal{E}_0 and X_0 to be complex numbers. We then substitute $\mathcal{E}(t) = \mathcal{E}_0 e^{-i\omega t}$ into eqn 2.5, and look for solutions of the form $x(t) = X_0 e^{-i\omega t}$. This gives:

$$-m_0 \omega^2 X_0 e^{-i\omega t} - i m_0 \gamma \omega X_0 e^{-i\omega t} + m_0 \omega_0^2 X_0 e^{-i\omega t} = -e \mathcal{E}_0 e^{-i\omega t}, \quad (2.8)$$

which implies that:

$$X_0 = \frac{-e \mathcal{E}_0 / m_0}{\omega_0^2 - \omega^2 - i \gamma \omega}. \quad (2.9)$$

The displacement of the electrons from their equilibrium position produces a time varying dipole moment $p(t)$, as shown in Fig. 2.2. The magnitude of the dipole is given by eqn 2.4. This gives a resonant contribution to the macroscopic polarization (dipole moment per unit volume) of the medium. If

N is the number of atoms per unit volume, the resonant polarization is given by:

$$\begin{aligned} P_{\text{resonant}} &= Np \\ &= -Nex \\ &= \frac{Ne^2}{m_0} \frac{1}{(\omega_0^2 - \omega^2 - i\gamma\omega)} \mathbf{\mathcal{E}}. \end{aligned} \quad (2.10)$$

A quick inspection of eqn 2.10 shows that the magnitude of P_{resonant} is small unless the frequency is close to ω_0 . This is another general property of forced oscillations: the response is small unless the frequency is close to resonance with the natural frequency of the oscillator.

Equation 2.10 can be used to obtain the complex relative dielectric constant ϵ_r . The electric displacement \mathbf{D} of the medium is related to the electric field $\mathbf{\mathcal{E}}$ and polarization \mathbf{P} through:

$$\mathbf{D} = \epsilon_0 \mathbf{\mathcal{E}} + \mathbf{P}, \quad (2.11)$$

where the bold font indicates vector quantities (see eqn A.2 in Appendix A). We are interested in the optical response at frequencies close to ω_0 , and so we split the polarization into a non-resonant background term and the resonant term arising from the driven response of the oscillator. We therefore write:

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \mathbf{\mathcal{E}} + \mathbf{P}_{\text{background}} + \mathbf{P}_{\text{resonant}} \\ &= \epsilon_0 \mathbf{\mathcal{E}} + \epsilon_0 \chi \mathbf{\mathcal{E}} + \mathbf{P}_{\text{resonant}}. \end{aligned} \quad (2.12)$$

To simplify the mathematics, we will assume that the material is isotropic, in which case the relative dielectric constant is defined through the relationship:

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{\mathcal{E}}. \quad (2.13)$$

We then combine eqns 2.10–2.13 to obtain:

$$\epsilon_r(\omega) = 1 + \chi + \frac{Ne^2}{\epsilon_0 m_0} \frac{1}{(\omega_0^2 - \omega^2 - i\gamma\omega)}. \quad (2.14)$$

This can be split into its real and imaginary parts according to eqn 1.18 to give:

$$\epsilon_1(\omega) = 1 + \chi + \frac{Ne^2}{\epsilon_0 m_0} \frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} \quad (2.15)$$

$$\epsilon_2(\omega) = \frac{Ne^2}{\epsilon_0 m_0} \frac{\gamma\omega}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}. \quad (2.16)$$

These formulae can be simplified further if we are working at frequencies close to resonance, where $\omega \approx \omega_0 \gg \gamma$. This allows us to approximate $(\omega_0^2 - \omega^2)$ by $2\omega_0\Delta\omega$, where $\Delta\omega = (\omega - \omega_0)$ is the detuning from ω_0 . We then notice that the low and high frequency limits of $\epsilon_r(\omega)$ are given by

$$\epsilon_r(0) \equiv \epsilon_{\text{st}} = 1 + \chi + \frac{Ne^2}{\epsilon_0 m_0 \omega_0^2}, \quad (2.17)$$

The electric susceptibility χ in eqn 2.12 accounts for all other contributions to the polarizability of the atoms. We will discuss the physical meaning of the ‘non-resonant polarization’ in subsection 2.2.2 below.

The treatment of non-isotropic materials only introduces unnecessary complications at this stage, and will be covered briefly in Section 2.4.

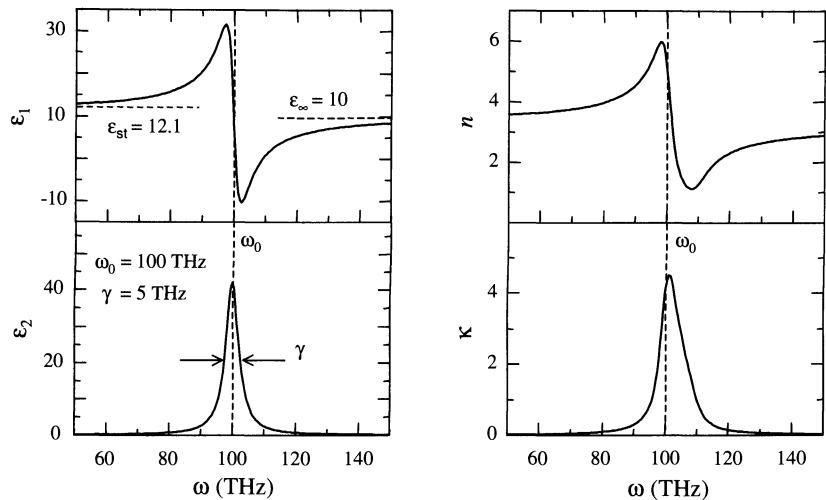


Fig. 2.4 Frequency dependence of the real and imaginary parts of the complex dielectric constant of a dipole oscillator at frequencies close to resonance. The graphs are calculated for an oscillator with $\omega_0 = 100$ THz, $\gamma = 5$ THz, $\epsilon_{\text{st}} = 12.1$ and $\epsilon_\infty = 10$. ($1 \text{ THz} = 10^{12} \text{ Hz}$) Also shown is the real and imaginary part of the refractive index calculated from the dielectric constant.

and

$$\epsilon_r(\infty) \equiv \epsilon_\infty = 1 + \chi \quad (2.18)$$

respectively. The subscript on ϵ_{st} stands for ‘static’, since it represents the dielectric response to static electric fields. With this notation we find that:

$$(\epsilon_{\text{st}} - \epsilon_\infty) = \frac{Ne^2}{\epsilon_0 m_0 \omega_0^2}. \quad (2.19)$$

We finally rewrite eqns 2.15 and 2.16 in the following form valid at frequencies close to resonance:

$$\epsilon_1(\Delta\omega) = \epsilon_\infty - (\epsilon_{\text{st}} - \epsilon_\infty) \frac{2\omega_0 \Delta\omega}{4(\Delta\omega)^2 + \gamma^2}, \quad (2.20)$$

$$\epsilon_2(\Delta\omega) = (\epsilon_{\text{st}} - \epsilon_\infty) \frac{\gamma \omega_0}{4(\Delta\omega)^2 + \gamma^2}. \quad (2.21)$$

These equations describe a sharp atomic absorption line centred at ω_0 with full width at half maximum equal to γ .

Figure 2.4 shows the frequency dependence of ϵ_1 and ϵ_2 predicted by eqns 2.20–2.21 for an oscillator with $\omega_0 = 10^{14}$ Hz, $\gamma = 5 \times 10^{12}$ Hz, $\epsilon_{\text{st}} = 12.1$ and $\epsilon_\infty = 10$. These numbers are fairly typical of the infrared absorption lines in an ionic crystal. We see that ϵ_2 is a strongly peaked function of ω with a maximum value at ω_0 and a full width at half maximum equal to γ . The frequency dependence of ϵ_1 is more complicated. As we approach ω_0 from below, ϵ_1 gradually rises from the low frequency value of ϵ_{st} , and reaches a peak at $\omega_0 - \gamma/2$. (See Example 2.1.) It then falls sharply, passing through a minimum at $\omega_0 + \gamma/2$ before rising again to the high frequency limit of ϵ_∞ . Note that the frequency scale over which these effects occur is determined by γ for both ϵ_1 and ϵ_2 . This shows that the damping of the oscillator causes line broadening. The frequency dependence determined of ϵ_1 and ϵ_2 shown in Fig. 2.4 is called Lorentzian after the originator of the dipole model.

In an experiment we actually measure the refractive index n and the absorption coefficient α . The measurement of α then determines the extinction

coefficient κ through eqn 1.16. Figure 2.4 shows the values of n and κ calculated from ϵ_1 and ϵ_2 using eqns 1.22 and 1.23. We see that n approximately follows the frequency dependence of $\sqrt{\epsilon_1(\omega)}$, while κ more or less follows $\epsilon_2(\omega)$. The correspondence $n \leftrightarrow \sqrt{\epsilon_1}$ and $\kappa \leftrightarrow \epsilon_2$ would be exact if κ were much smaller than n (cf. eqns 1.24 and 1.25). This is what generally happens in gases in which the low density of atoms makes the total absorption small. In the example shown in Fig. 2.4 the correspondence is only approximate because the absorption is very strong near ω_0 , so that we cannot always assume $n \gg \kappa$. Nevertheless, the basic behaviour shows that the absorption peaks at a frequency very close to ω_0 and has a width of about γ , while the refractive index shows positive and negative excursions below and above ω_0 . This is the typical behaviour expected of an atomic absorption line.

One interesting aspect of the Lorentz oscillator is that it affects the refractive index over a much larger frequency range than the absorption. This point is clearly shown in the graphs given in Fig. 2.4. The absorption is a strongly peaked function of ω and falls off as $(\Delta\omega)^{-2}$ as we tune away from resonance. Thus there is no significant absorption if we tune sufficiently far from resonance. On the other hand, the frequency dependence of the refractive index varies as $|\Delta\omega|^{-1}$ for large $|\Delta\omega|$. This follows from eqn 2.20 with the approximation $n = \sqrt{\epsilon_1}$, which is valid for large $|\Delta\omega|$ when ϵ_2 is very small.

Example 2.1

The full width at half maximum of the strongest hyperfine component of the sodium D₂ line at 589.0 nm is 100 MHz. A beam of light passes through a gas of sodium with an atom density of $1 \times 10^{17} \text{ m}^{-3}$. Calculate: (i) The peak absorption coefficient due to this absorption line. (ii) The frequency at which the resonant contribution to the refractive index is at a maximum. (iii) The peak value of the resonant contribution to the refractive index.

Solution

(i) We are dealing with a low density gas of atoms, and so the approximations given in eqns 1.24 and 1.25 will be valid. This means that the absorption will directly follow the frequency dependence of $\epsilon_2(\omega)$, and the peak absorption will occur precisely at the line centre. The peak extinction coefficient can be worked out from eqns 2.16 and 1.25. This gives:

$$\kappa(\omega_0) = \frac{\epsilon_2(\omega_0)}{2n} = \frac{Ne^2}{2n\epsilon_0 m_0} \frac{1}{\gamma\omega_0}.$$

We do not know what n is, but because we are dealing with a gas, it will only be very slightly different from unity. This point is confirmed in part (iii) of the question. We therefore take $n = 1$ here, and insert $N = 1 \times 10^{17} \text{ m}^{-3}$, $\gamma = 2\pi \times 100 \text{ MHz}$ and $\omega_0 = 2\pi c/\lambda = 3.20 \times 10^{15} \text{ Hz}$, to find that $\kappa(\omega_0) = 7.90 \times 10^{-5}$. This confirms that $n \gg \kappa$, and hence that it is valid to use eqn 1.25. We then work out the absorption coefficient from Eq. 1.16, which gives:

$$\alpha_{\max} \equiv \alpha(\omega_0) = \frac{4\pi\kappa(\omega_0)}{\lambda} = 1.7 \times 10^3 \text{ m}^{-1}.$$

The absorption coefficient measured in an experiment would actually be smaller than the value calculated here by about a factor of 3. This discrepancy is caused by the fact that we are assuming that the oscillator strength of the transition is unity. This point is discussed further in section 2.2.2 below.

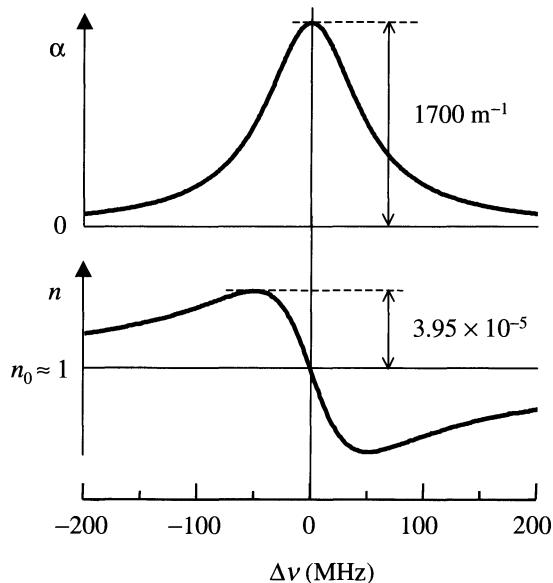


Fig. 2.5 Absorption coefficient and refractive index of sodium gas in the vicinity of the strongest hyperfine component of the D₂ line, on the assumption that the oscillator strength of the transition is unity, and that the atom density is $1 \times 10^{17} \text{ m}^{-3}$. See Example 2.1 for the details. n_0 represents the off-resonant refractive index, which is approximately equal to unity.

(ii) We know from Fig. 2.4 that there will be a peak in the refractive index just below ω_0 . Equation 1.24 tells us that $n(\omega) = \sqrt{\epsilon_1(\omega)}$, and hence that the local maximum of n will occur at the same frequency as the maximum in ϵ_1 . Since the peak occurs near ω_0 , it will be valid to use eqn 2.20. The local maximum occurs when:

$$\frac{d\epsilon_1(\omega)}{d\omega} \equiv \frac{d\epsilon_1(\Delta\omega)}{d\Delta\omega} \propto \frac{4(\Delta\omega)^2 - \gamma^2}{[4(\Delta\omega)^2 + \gamma^2]^2} = 0.$$

This gives $\Delta\omega = \pm\gamma/2$. We see from Fig. 2.4 that $\Delta\omega = -\gamma/2$ corresponds to the local maximum, while $\Delta\omega = +\gamma/2$ corresponds to the local minimum. Therefore the peak in the refractive index occurs 50 MHz below the line centre.

(iii) From part (ii) we know that the local maximum in the refractive index occurs when $\Delta\omega = -\gamma/2$. We see from eqns 1.24 and 2.20 that the refractive index at this frequency is given by:

$$n_{\max} = \sqrt{\epsilon_1} = \left(\epsilon_\infty + \frac{Ne^2}{2\epsilon_0 m_0 \omega_0 \gamma} \right)^{\frac{1}{2}} = n_0 \left(1 + \frac{7.90 \times 10^{-5}}{n_0^2} \right)^{\frac{1}{2}},$$

where $n_0 = \sqrt{\epsilon_\infty}$ is the off-resonant refractive index. We are dealing with a low density gas, and so it is justified to take $n_0 \approx 1$ here. This implies that the peak value of the resonant contribution to the refractive index is 3.95×10^{-5} .

The full frequency dependence of the absorption and refractive index near this absorption line is plotted in Fig. 2.5.

2.2.2 Multiple resonances

In general, an optical medium will have many characteristic resonant frequencies. We already discussed in Section 2.1 how we expect to observe separate

resonances due to the lattice vibrations and to the oscillations of the bound electrons within the atoms. Furthermore, a particular medium may have many resonances of each type. We can treat these multiple resonances without difficulty in our model provided they occur at distinct frequencies.

In writing eqn 2.12 we split the polarization of the medium into a resonant part and a non-resonant part. We then discussed the resonant part in detail, without specifying very accurately what we meant by the non-resonant term. We simply stated that \mathbf{P} was proportional to $\boldsymbol{\epsilon}$ through the susceptibility χ . In reality, the non-resonant polarization of the medium must originate from the polarizability of the atoms in exactly the same way as the resonant part. Equation 2.19 tells us that the dielectric constant decreases each time we go through an absorption line. The contributions that enter the background electric susceptibility χ in eqn 2.12 thus arise from the polarization due to all the other oscillators at higher frequencies.

We can understand this point better by making it more quantitative. The contribution to the polarization of a particular oscillator is given by eqn 2.10. In a medium with many electronic oscillators of different frequencies, the total polarization will therefore be given by

$$\mathbf{P} = \left(\frac{Ne^2}{m_0} \sum_j \frac{1}{(\omega_j^2 - \omega^2 - i\gamma_j\omega)} \right) \boldsymbol{\epsilon}, \quad (2.22)$$

where ω_j and γ_j are the frequency and damping terms of a particular resonance line. We then substitute this into eqn 2.11, and recall the definition of ϵ_r given in eqn 2.13. This gives:

$$\epsilon_r(\omega) = 1 + \frac{Ne^2}{\epsilon_0 m_0} \sum_j \frac{1}{(\omega_j^2 - \omega^2 - i\gamma_j\omega)}. \quad (2.23)$$

This equation takes account of all the transitions in the medium and can be used to calculate the full frequency dependence of the dielectric constant.

The refractive index and absorption coefficient calculated from eqn 2.23 are plotted against frequency in Fig. 2.6. The figure has been calculated for a hypothetical solid with three well-separated resonances with ω_j equal to 4×10^{13} Hz, 4×10^{15} Hz and 1×10^{17} Hz respectively. The width of each absorption line has been set to 10 % of the centre frequency by appropriate choice of the damping term. The resonance in the infrared is included to represent the vibrational absorption. In a real solid, we would have to adapt the model appropriately to account for the different reduced mass and effective charge of the vibrational oscillator.

We can understand this figure by starting at the highest frequencies and gradually working our way down to the lower frequencies. At the very highest frequencies, the electrons are incapable of responding to the driving field. The medium therefore has no polarization, and the dielectric constant is unity. As we reduce the frequency, we first run into the transitions of the inner electrons in the X-ray/vacuum-ultraviolet spectral region, and then the transitions of the outer electrons in the ultraviolet and visible. We then have a region with no transitions until we finally reach the vibrational frequencies in the infrared. Each time we go through one of these resonances, we see the characteristic

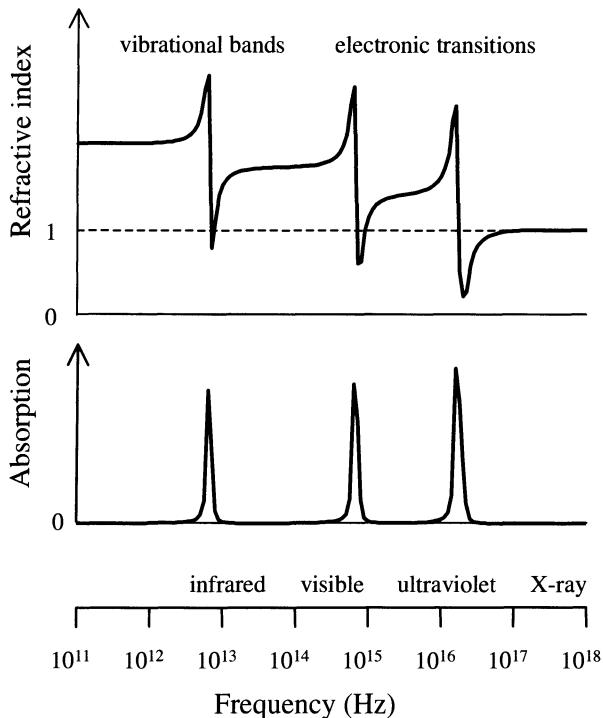


Fig. 2.6 Schematic diagram of the frequency dependence of the refractive index and absorption of a hypothetical solid from the infrared to the X-ray spectral region. The solid is assumed to have three resonant frequencies with $\omega_j = 4 \times 10^{13} \text{ Hz}$, $4 \times 10^{15} \text{ Hz}$ and $1 \times 10^{17} \text{ Hz}$ respectively. The width of each absorption line has been set to 10 % of the centre frequency by appropriate choice of the γ_j 's.

frequency dependence of the Lorentz oscillator, with a peak in the absorption spectrum and a ‘wiggle’ in the refractive index. In between the resonances the medium is transparent: the absorption coefficient is zero and the refractive index is almost constant.

The value of the refractive index in the transparent regions gradually increases as we go through more and more resonance lines on decreasing the frequency. This increase of the refractive index is caused by the fact that $\epsilon_{st} > \epsilon_\infty$ (cf. eqn 2.19), which implies that n is larger below an absorption line than above it. By reference to Fig. 2.6, we now see that we have to understand ‘static’ and ‘ ∞ ’ as relative to a particular resonance. The variation of n with frequency due to the resonances is the origin of the dispersion found in optical materials even when they are transparent. This point will be discussed further in Section 2.3 below.

The dipole oscillator model predicts that each oscillator contributes a term given by eqn 2.10. This leads to a series of absorption lines of the same strength. However, experimental data shows that the absorption strength actually varies considerably between different atomic transitions. With the benefit of hindsight, we know that this is caused by the variation of the quantum mechanical transition probability. (See Appendix B.) In classical physics, however, there is no explanation, and we just assign a phenomenological oscillator strength f_j to each transition, rewriting eqn 2.23 as:

$$\epsilon_r(\omega) = 1 + \frac{Ne^2}{\epsilon_0 m_0} \sum_j \frac{f_j}{(\omega_{0j}^2 - \omega^2 - i\gamma_j\omega)} . \quad (2.24)$$

An astute reader will have noticed that the peak absorption coefficient for the three transition lines shown in Fig. 2.6 decreases slightly with decreasing frequency. This happens because n is larger at the lower frequencies. The transitions all have the same peak ϵ_2 , but we can see from eqn 1.21 that κ must be slightly smaller if n is larger.

It can be shown from quantum mechanics that we must have $\sum_j f_j = 1$ for each electron. Since the classical model predicts $f_j = 1$ for each oscillator, we then interpret this by saying that a particular electron is involved in several transitions at the same time, and the absorption strength is being divided between these transitions.

2.2.3 Comparison with experimental data

The schematic behaviour shown in Fig. 2.6 can be compared to experimental data on a typical solid state material. Figure 2.7 shows the frequency dependence of the refractive index and extinction coefficient of fused silica (SiO_2) glass from the infrared to the X-ray spectral region. The general characteristics indicated by Fig. 2.6 are clearly observed, with strong absorption in the infrared and ultraviolet, and a broad region of low absorption in between. The data confirms that $n \gg \kappa$ except near the peaks of the absorption. This means that the approximation whereby we associate the frequency dependence of n with that of ϵ_1 , and that of κ with ϵ_2 (eqns 1.24 and 1.25), is valid at most frequencies.

The general behaviour shown in Fig. 2.7 is typical of optical materials which are transparent in the visible spectral region. We already noted in Sections 1.4.1 and 1.4.2 that the transmission range of colourless materials is determined by

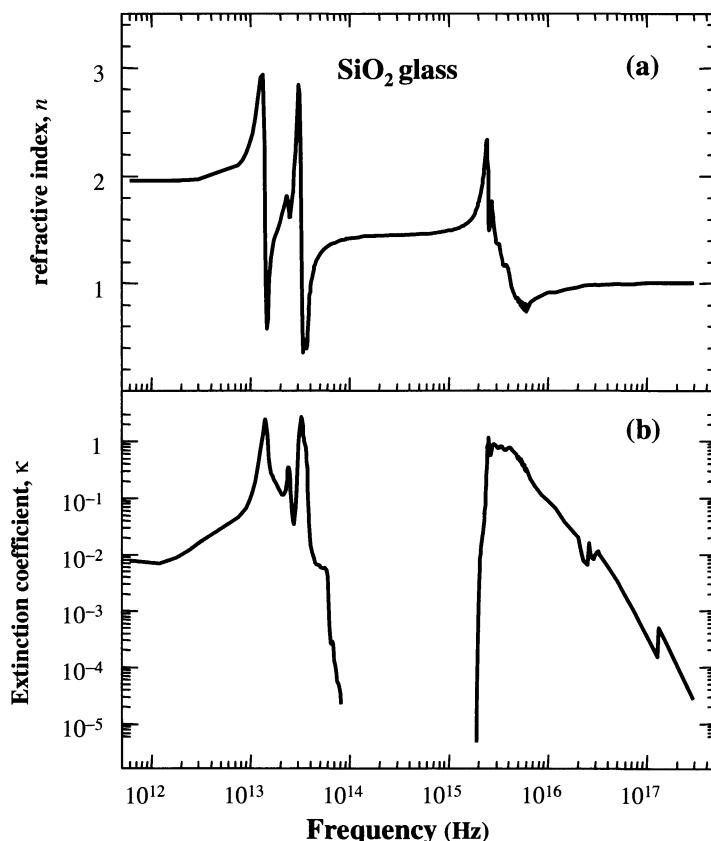


Fig. 2.7 (a) Refractive index and (b) extinction coefficient of fused silica (SiO_2) glass from the infrared to the x-ray spectral region. After [1].

the electronic absorption in the ultraviolet and the vibrational absorption in the infrared. This is demonstrated by the transmission data for sapphire shown in Fig. 1.4(a).

Silica is a glass, and hence does not have a regular crystal lattice. The infrared absorption is therefore caused by excitation of vibrational quanta in the SiO_2 molecules themselves. Two distinct peaks are observed at 1.4×10^{13} Hz ($21\ \mu\text{m}$) and 3.3×10^{13} Hz ($9.1\ \mu\text{m}$) respectively. These correspond to different vibrational modes of the molecule. The detailed modelling of these absorption bands by the oscillator model will be discussed in Chapter 10.

The ultraviolet absorption in silica is caused by interband electronic transitions. SiO_2 has a fundamental band gap of about 10 eV, and interband transitions are possible whenever the photon energy exceeds this value. Hence we observe an absorption threshold in the ultraviolet at 2×10^{15} Hz (150 nm). The interband absorption peaks at around 3×10^{15} Hz with an extremely high absorption coefficient of $\sim 10^8\ \text{m}^{-1}$, and then gradually falls off to higher frequency. Subsidiary peaks are observed at $\sim 3 \times 10^{16}$ Hz and 1.3×10^{17} Hz. These are caused by transitions of the inner core electrons of the silicon and oxygen atoms. The fact that the electronic absorption consists of a continuous band rather than a discrete line makes it hard to model accurately as a Lorentz oscillator. We will discuss the quantum theory of the interband absorption in Chapter 3.

The refractive index of glass has resonances in the infrared and the ultraviolet which correspond to the interband and vibrational absorption. In the far infrared region below the vibrational resonance, the refractive index is ~ 2 , while in the hard ultraviolet and X-ray region it approaches unity. In the transparency region between the vibrational and interband absorption, the refractive index has a value of ~ 1.5 . Closer inspection of Fig. 2.7 shows that the refractive index actually increases with frequency in this transparency region, rising from a value of 1.40 at 8×10^{13} Hz ($3.5\ \mu\text{m}$) to 1.55 at 1.5×10^{15} Hz (200 nm). This dispersion originates from the low frequency wings of the ultraviolet absorption and the high frequency wings of the infrared absorption, and will be discussed in more detail in Section 2.3 below.

The data in Fig. 2.7 show that the refractive index falls below unity at a number of frequencies. This implies that the phase velocity of the light is greater than c , which might seem to imply a contradiction with relativity. However, this overlooks the fact that a signal must be transmitted as a wave packet rather than as a monochromatic wave. In a dispersive medium, a wave packet will propagate at the group velocity v_g given by:

$$v_g = \frac{d\omega}{dk}, \quad (2.25)$$

rather than at the phase velocity $v = \omega/k = c/n$. The relationship between v_g and v is:

$$v_g = v \left(1 - \frac{k}{n} \frac{dn}{dk} \right). \quad (2.26)$$

The derivation of this result is left as an exercise to the reader. (See Exercise 2.7.) We will see in Section 2.3 that dn/dk is positive in most materials at optical frequencies. This then implies that v_g is always less than v , and if we were to try to transmit a signal in a spectral region where $v > c$, we

It is apparent from Fig. 2.6 that dn/dk will be negative at some frequencies close to one of the resonance lines. Equation 2.26 then implies that $v_g > v$, and so we could again run into a problem with relativity. However, the medium is highly absorbing in these frequency regions, and this means that the signal travels with yet another velocity called the signal velocity. This is always less than c .

would always find that v_g is less than c . The proof of this for a simple Lorentz oscillator is considered in Exercise 2.8.

2.2.4 Local field corrections

The calculation of the dielectric constant given in eqn 2.24 is valid in a rarefied gas with a low density of atoms. However, in a dense optical medium such as a solid, there is another factor that we must consider. The individual atomic dipoles respond to the local field that they experience. This may not necessarily be the same as the external field, because the dipoles themselves generate electric fields which will be felt by all the other dipoles. The actual local field experienced by an atom therefore takes the form:

$$\boldsymbol{\epsilon}_{\text{local}} = \boldsymbol{\epsilon} + \boldsymbol{\epsilon}_{\text{other dipoles}}, \quad (2.27)$$

where $\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_{\text{other dipoles}}$ represent the fields due to the external field and the other dipoles respectively. We should have been using $\boldsymbol{\epsilon}_{\text{local}}$ instead of $\boldsymbol{\epsilon}$ all along throughout the calculation given in Sections 2.2.1 and 2.2.2.

The calculation of the correction field due to the other dipoles in the medium is actually a rather complicated one. An approximate solution due to Lorentz can be derived if we assume that all the dipoles are parallel to the applied field and are arranged on a cubic lattice. The calculation works by separating the contribution from the nearby dipoles and that from the rest of the sample, as indicated in Fig. 2.8. The division is effected by an imaginary spherical surface with a radius large enough to make it sensible to average the material outside it. The problem is then reduced to summing the field of the dipoles inside the sphere at the one in the middle, and then calculating the effect of a uniformly polarized dielectric outside the sphere. The final result is:

$$\boldsymbol{\epsilon}_{\text{other dipoles}} = \frac{\mathbf{P}}{3\epsilon_0}, \quad (2.28)$$

where \mathbf{P} is the polarization of the dielectric outside the sphere. The derivation of this result is the subject of Exercise 2.9. By using the result of eqn 2.28 in eqn 2.27 we find that:

$$\boldsymbol{\epsilon}_{\text{local}} = \boldsymbol{\epsilon} + \frac{\mathbf{P}}{3\epsilon_0}. \quad (2.29)$$

The macroscopic polarization \mathbf{P} will be given by

$$\mathbf{P} = N\epsilon_0\chi_a\boldsymbol{\epsilon}_{\text{local}} \quad (2.30)$$

where χ_a is the electric susceptibility per atom. χ_a is defined by:

$$\mathbf{p} = \epsilon_0\chi_a\boldsymbol{\epsilon}_{\text{local}}, \quad (2.31)$$

\mathbf{p} being the induced dipole moment per atom. This is analogous to the usual definition of the macroscopic susceptibility given in eqn A.1, except that it is now applied to individual atoms interacting with the local field. We can see from eqn 2.10 that χ_a is given by

$$\chi_a = \frac{e^2}{\epsilon_0 m_0} \frac{1}{(\omega_0^2 - \omega^2 - i\gamma\omega)}, \quad (2.32)$$

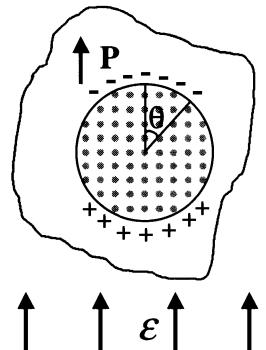


Fig. 2.8 Model used to calculate the local field by the Lorentz correction. An imaginary spherical surface drawn around a particular atom divides the medium into nearby dipoles and distant dipoles. The field at the centre of the sphere due to the nearby dipoles is summed exactly, while the field due to the distant dipoles is calculated by treating the material outside the sphere as a uniformly polarized dielectric.

if there is just a single resonance. This is modified to

$$\chi_a = \frac{e^2}{\epsilon_0 m_0} \sum_j \frac{f_j}{(\omega_j^2 - \omega^2 - i\gamma_j\omega)}, \quad (2.33)$$

if there are multiple resonances (cf. eqn 2.24).

We can combine eqns 2.29 and 2.30 with eqns 2.11 and 2.13 by writing

$$\mathbf{P} = N\epsilon_0\chi_a \left(\boldsymbol{\epsilon} + \frac{\mathbf{P}}{3\epsilon_0} \right) = (\epsilon_r - 1)\epsilon_0\boldsymbol{\epsilon}. \quad (2.34)$$

We put all this together to find that:

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{N\chi_a}{3}. \quad (2.35)$$

This result is known as the **Clausius–Mossotti relationship**. The relationship works well in gases and liquids. It is also valid for those crystals in which the Lorentz correction given in eqn 2.29 gives an accurate account of the local field effects, namely cubic crystals.

2.2.5 The Kramers–Kronig relationships

The discussion of the dipole oscillator shows that the refractive index and the absorption coefficient are not independent parameters but are related to each other. This is a consequence of the fact that they are derived from the real and imaginary parts of a single parameter, namely the complex refractive index. If we invoke the law of causality (that an effect may not precede its cause) and apply complex number analysis, we can derive general relationships between the real and imaginary parts of the refractive index. These are known as the Kramers–Kronig relationships and may be stated as follows:

$$n(\omega) = 1 + \frac{1}{\pi} \text{P} \int_{-\infty}^{\infty} \frac{\kappa(\omega')}{\omega' - \omega} d\omega' \quad (2.36)$$

$$\kappa(\omega) = -\frac{1}{\pi} \text{P} \int_{-\infty}^{\infty} \frac{n(\omega') - 1}{\omega' - \omega} d\omega', \quad (2.37)$$

where P indicates that we take the principal part of the integral.

The Kramers–Kronig relationships allow us to calculate n from κ , and *vice versa*. This can be very useful in practice, because it would allow us, for example, to measure the frequency dependence of the optical absorption and then calculate the dispersion without needing to make a separate measurement of n .

2.3 Dispersion

Figure 2.9 plots the refractive index data from Fig. 2.7 in more detail. The data show that the refractive index increases with frequency in the near infrared and visible spectral regions. We have seen in Section 2.2.3 that this dispersion originates mainly from the interband absorption in the ultraviolet. At visible

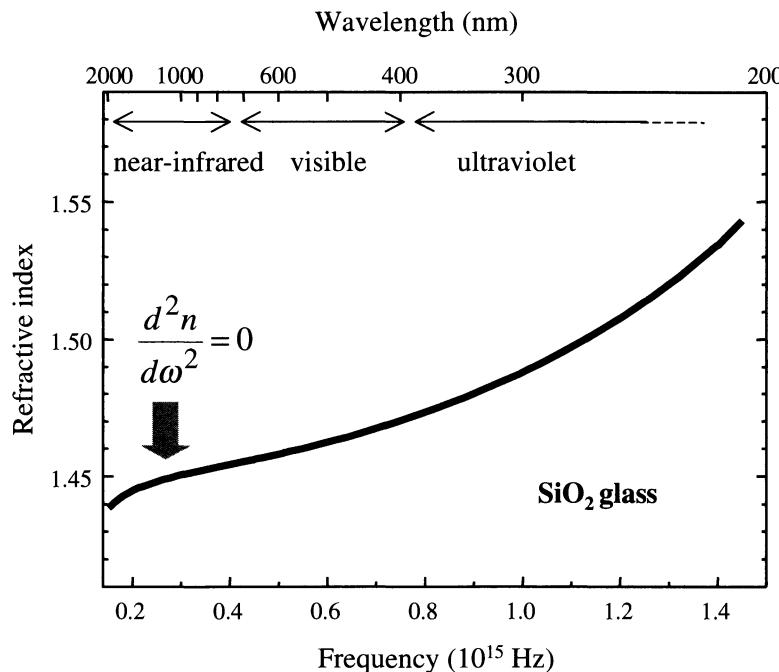


Fig. 2.9 Refractive index of SiO_2 glass in the near infrared, visible and ultraviolet spectral regions. After [1].

frequencies the absorption from these transitions is negligible and the glass is transparent. However, the ultraviolet absorption still affects the refractive index through the extreme wings of the Lorentzian line. In the near infrared, the dispersion is also affected by the high frequency wings of the vibrational absorption at lower frequency.

A material in which the refractive index increases with frequency is said to have **normal** dispersion, while one in which the contrary occurs is said to have **anomalous** dispersion. A number of empirical formulae to describe the normal dispersion of glasses have been developed over the years. (See Exercise 2.12.)

The dispersion of the refractive index of glasses such as silica can be used to separate different wavelengths of light with a prism, as shown in Fig. 2.10. The blue light is refracted more because of the higher index of refraction, and is therefore deviated through a larger angle by the prism. (See Exercise 2.13.) This effect is used in prism spectrometers.

One of the effects of dispersion is that light of different frequencies takes a different amount of time to propagate through a material. (See Exercise 1.11, for example.) A pulse of light of duration t_p must necessarily contain a spread of frequencies given approximately by

$$\Delta\nu \approx \frac{1}{t_p} \quad (2.38)$$

in order to satisfy the ‘uncertainty principle’ $\Delta\nu\Delta t \sim 1$. Dispersion will therefore cause the pulse to broaden in time as it propagates through the medium. This can become a serious problem when attempting to transmit very short pulses through a long length of an optical material, for example in a high speed optical fibre telecommunications system.

The use of the words ‘normal’ and ‘anomalous’ is somewhat misleading here. The dipole oscillator model shows us that all materials have anomalous dispersion at some frequencies. The phraseology was adopted before measurements of the refractive index had been made over a wide frequency range and the origin of dispersion had been properly understood.

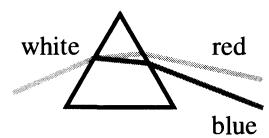


Fig. 2.10 Separation of white light into different colours by dispersion in a glass prism.

We mentioned in Section 2.2.3 that a pulse of light travels with the group velocity v_g . The important parameter for pulse spreading due to dispersion is therefore the **group velocity dispersion** (GVD) (see Exercise 2.14):

$$\text{GVD} = \frac{d^2\omega}{dk^2} \propto \frac{d^2n}{d\omega^2} \propto \frac{d^2n}{d\lambda^2}. \quad (2.39)$$

The Lorentz model indicates that the GVD is positive below an absorption line and negative above it. Applying this to the data in Fig. 2.9, we see negative GVD in the infrared due to the vibrational absorption and positive GVD in the visible due to the interband absorption in the ultraviolet. These two effects cancel at a wavelength in the near infrared which is identified in Fig. 2.9. This region of zero GVD occurs around $1.3 \mu\text{m}$ in silica optical fibres. Short pulses can be transmitted down the fibre with negligible temporal broadening at this wavelength, and so it is one of the preferred wavelengths for optical fibre communication systems.

2.4 Optical anisotropy: birefringence

The atoms in a solid are locked into a crystalline lattice with well defined axes. In general, we cannot assume that the optical properties along the different crystalline axes are equivalent. For example, the separation of the atoms might not be the same in all directions. This would lead to different vibrational frequencies, and hence a change in the refractive index between the relevant directions. This optical anisotropy contrasts with gases and liquids which are isotropic because the atoms have no preferred directions in the absence of external perturbations such as applied magnetic or electric fields.

Optical anisotropy gives rise to the phenomenon of **birefringence**. We can describe the properties of a birefringent crystal by generalizing the relationship between the polarization and the applied electric field. If the electric field is applied along an arbitrary direction relative to the crystalline axes, we must write a tensor equation to relate \mathbf{P} to \mathbf{E} :

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \quad (2.40)$$

where χ represents the susceptibility tensor. Written explicitly in terms of the components, we have:

$$\begin{pmatrix} P_x \\ P_y \\ P_z \end{pmatrix} = \epsilon_0 \begin{pmatrix} \chi_{11} & \chi_{12} & \chi_{13} \\ \chi_{21} & \chi_{22} & \chi_{23} \\ \chi_{31} & \chi_{32} & \chi_{33} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}. \quad (2.41)$$

We can simplify this by choosing the cartesian coordinates x , y , and z to correspond to the principal crystalline axes. In this case, the off-diagonal components are zero, and the susceptibility tensor takes the form:

$$\chi = \begin{pmatrix} \chi_{11} & 0 & 0 \\ 0 & \chi_{22} & 0 \\ 0 & 0 & \chi_{33} \end{pmatrix}. \quad (2.42)$$

The relationships between the components are determined by the crystal symmetry.

Table 2.1 Refractive indices of some common uniaxial crystals at 589.3 nm. After [2].

Crystal	Chemical structure	Symmetry class	type	n_o	n_e
Ice	H_2O	trigonal	positive	1.309	1.313
Quartz	SiO_2	trigonal	positive	1.544	1.553
Beryl	$Be_3Al_2(SiO_3)_6$	hexagonal	negative	1.581	1.575
Sodium nitrate	$NaNO_3$	trigonal	negative	1.584	1.336
Calcite	$CaCO_3$	trigonal	negative	1.658	1.486
Tourmaline	complex silicate	trigonal	negative	1.669	1.638
Sapphire	Al_2O_3	trigonal	negative	1.768	1.760
Zircon	$ZrSiO_4$	tetragonal	positive	1.923	1.968
Rutile	TiO_2	tetragonal	positive	2.616	2.903

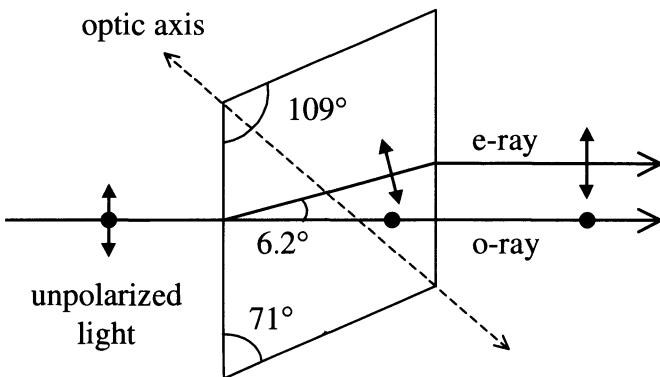


Fig. 2.11 Double refraction in a natural calcite crystal. The shape of the crystal and the orientation of the optic axis is determined by the cleavage planes of calcite. An unpolarized incident light ray is split into two spatially separated orthogonally polarized rays. The • symbol for the o-ray indicates that it is polarized with its field pointing out of the page.

- In cubic crystals, the x , y and z axes are indistinguishable. They therefore have $\chi_{11} = \chi_{22} = \chi_{33}$, and their optical properties are isotropic.
- Crystals with tetragonal, hexagonal or trigonal (rhombohedral) symmetry are called **uniaxial** crystals. These crystals possess a single **optic axis**, which is usually taken as the z axis. In hexagonal crystals, for example, the optic axis is defined by the direction normal to the plane of the hexagons. The optical properties are the same along the x and y directions, but not along the z direction. This implies that $\chi_{11} = \chi_{22} \neq \chi_{33}$. Some examples of uniaxial crystals are listed in Table 2.1.
- Crystals with orthorhombic, monoclinic or triclinic symmetry are called **biaxial** crystals. They have two optic axes, and all three diagonal components of the susceptibility tensor are different. Mica is an important example of a biaxial crystal, since it is widely used for making optical wave plates.

One very striking demonstration of optical anisotropy is the phenomenon of **double refraction**. In this effect an unpolarized light ray is separated into two rays which emerge displaced from each other, as shown in Fig. 2.11. These two rays are called ‘ordinary’ and ‘extraordinary’, and are orthogonally polarized to each other.

The phenomenon of double refraction can be explained by assuming that the crystal has different refractive indices for the orthogonal polarizations of the ordinary and extraordinary rays. These two refractive indices are usually labelled n_o and n_e respectively. Consider the propagation of a beam of unpo-

Crystals with cubic symmetry are only isotropic as regards their *linear* optical properties. We will see in Chapter 11 that cubic crystals can actually have anisotropic *nonlinear* optical properties.

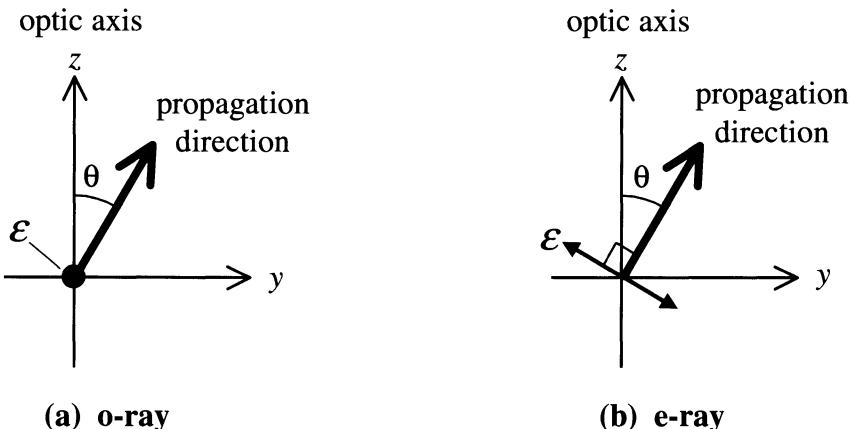


Fig. 2.12 Electric field vector of ray propagating in a uniaxial crystal with its optic axis along the z direction. The ray makes an angle of θ with respect to the optic axis. The x and y axes are chosen so that the beam is propagating in the y, z plane. The polarization can be resolved into: (a) a component along the x axis and (b) a component at an angle of $90^\circ - \theta$ to the optic axis. (a) is the o-ray and (b) is the e-ray.

polarized light which enters a uniaxial crystal at an angle θ to the optic axis, which is taken to lie along the z axis. The optical properties are isotropic in the x, y plane, and so we can choose the axes so that the beam is propagating in the y, z plane without loss of generality, as shown in Fig. 2.12. This allows us to split the polarization of the light into two orthogonal components, one of which is polarized along the x axis, and the other polarized at an angle of $(90^\circ - \theta)$ to the optic axis. The former is the o(ordinary)-ray, and the latter is the e(xtraordinary)-ray. Now the refractive index will be different for light which is polarized along the z axis or in the x, y plane. Therefore the o-ray experiences a different refractive index to the e-ray, and will thus be refracted differently: hence double refraction. On the other hand, if the beam propagates along the optic axis so that $\theta = 0$, the \mathbf{E} -vector of the light will always fall in the x, y plane. In this case, no double refraction will be observed because the x and y directions are equivalent and there is no e-ray.

Double refraction was first observed in natural uniaxial crystals such as calcite ('Iceland Spar') and quartz. Table 2.1 lists the refractive indices for the o- and e-rays of calcite and quartz, together with those of several other uniaxial crystals. The birefringent crystals are classified as being either positive or negative depending on whether n_e is greater or smaller than n_o .

Further discussion of the detailed effects of birefringence can be found in most optics textbooks. The purpose of introducing birefringence here is to give an example of how the phenomenon of optical anisotropy arises from the underlying symmetry of the crystal structure. This is a very standard example of an optical effect that occurs in crystalline solids and is not found in gases or liquids.

Example 2.2

The optic axis of a uniaxial crystal lies along the z axis. The refractive index for light polarized in the z direction is n_e , while that for light polarized in the x, y plane is n_o . Write down the dielectric constant tensor defined through the tensor relationship

$$D = \epsilon_0 \epsilon_r \epsilon,$$

Solution

We make use of eqns 2.11 and 2.40 to write:

$$\begin{aligned}\mathbf{D} &= \epsilon_0 \mathbf{E} + \mathbf{P} \\ &= \epsilon_0 \mathbf{E} + \epsilon_0 \chi \mathbf{E} \\ &= \epsilon_0 (1 + \chi) \mathbf{E} \equiv \epsilon_0 \epsilon_r \mathbf{E}.\end{aligned}\quad (2.43)$$

Hence we see that:

$$\epsilon_r = 1 + \chi \quad (2.44)$$

The susceptibility tensor is given by eqn 2.42, and hence the dielectric constant tensor will take the form:

$$\epsilon_r = \begin{pmatrix} 1 + \chi_{11} & 0 & 0 \\ 0 & 1 + \chi_{22} & 0 \\ 0 & 0 & 1 + \chi_{33} \end{pmatrix}. \quad (2.45)$$

In a uniaxial crystal with the optic axis along the z direction, we must have $\chi_{11} = \chi_{22} \neq \chi_{33}$.

We now further assume that the crystal is transparent, so that the dielectric constant is just equal to the square of the refractive index (cf. eqns 1.24 and 1.25 with $\kappa = 0$). If we had a linearly polarized light beam with the electric field directed along the x or y directions, we would measure a refractive index of n_o . This tells us that

$$1 + \chi_{11} = 1 + \chi_{22} = n_o^2.$$

On the other hand, if \mathbf{E} is along the z axis, we would measure a refractive index of n_e , which implies that

$$1 + \chi_{33} = n_e^2.$$

Therefore the dielectric constant tensor must be:

$$\epsilon_r = \begin{pmatrix} n_o^2 & 0 & 0 \\ 0 & n_o^2 & 0 \\ 0 & 0 & n_e^2 \end{pmatrix}. \quad (2.46)$$

Chapter summary

- The classical model of a solid treats the atoms and molecules as oscillating electric dipoles with characteristic resonant frequencies. The resonances due to the bound electrons occur in the near infrared, visible and ultraviolet spectral regions (10^{14} – 10^{15} Hz), while those associated with vibrations occur in the infrared (10^{12} – 10^{13} Hz). Free electrons can be treated in the dipole oscillator model by assuming that the natural resonant frequency is 0.

- The medium absorbs light when the frequency coincides with one of its resonant frequencies. In non-resonant conditions, the medium is transparent, but the velocity of light is reduced by multiple coherent elastic scattering.
- The absorption coefficient of an individual dipole oscillator has a Lorentzian line shape (cf. eqn 2.21). The spectral width of the absorption line is equal to the damping constant γ . The peak absorption is proportional to $1/\gamma$.
- The refractive index of a dipole oscillator increases as the frequency approaches the resonant frequency, then drops sharply in the absorbing region, and then increases again at higher frequencies. The off-resonant refractive index decreases each time we go through an absorption line.
- The dielectric constant of a medium with multiple resonant frequencies is given by eqn 2.24. The refractive index and absorption coefficient can be calculated from the real and imaginary parts of ϵ_r .
- The dipole oscillator model demonstrates that the absorption and refraction of an optical medium are fundamentally related to each other. This interrelationship is made explicit through the Kramers–Kronig formulae.
- Dispersion in the refractive index originates from the wings of the resonances at transition frequencies. The dispersion is called normal when the refractive index increases with frequency. Group velocity dispersion causes temporal broadening of short pulses.
- Optical anisotropy leads to birefringence. The anisotropy is described through the electric susceptibility tensor or the dielectric constant tensor.

Further reading

The subject matter of this chapter is covered, to a greater or lesser extent, in most electromagnetism and optics textbooks. See, for example: Bleaney and Bleaney (1976), Born and Wolf (1999), Hecht (1998) or Klein and Furtak (1986).

An excellent collection of optical data on a wide range of solid state materials can be found in Palik (1985).

For a fuller description of birefringence, see: Hecht (1998), Born and Wolf (1999) or Klein and Furtak (1986).

References

- [1] Palik, E.D. (1985). *Handbook of the optical constants of solids*. Academic Press, San Diego.
- [2] Driscoll, W.G. and Vaughan, W. (1978). *Handbook of optics*. McGraw-Hill, New York.

Exercises

- (2.1) Write down the equations of motion for the frictionless displacements x_1 and x_2 of two masses, m_1 and m_2 , connected together by a light spring with a spring constant K_s . Hence show that the angular frequency for small oscillations is equal to $(K_s/\mu)^{1/2}$ where $\mu^{-1} = m_1^{-1} + m_2^{-1}$.
- (2.2) A damped oscillator with mass m , natural frequency ω_0 , and damping constant γ is being driven by a force of amplitude F_0 and frequency ω . The equation of motion for the displacement x of the oscillator is:

$$m \frac{d^2x}{dt^2} + m\gamma \frac{dx}{dt} + m\omega_0^2 x = F_0 \cos \omega t.$$

What is the phase of x relative to the phase of the driving force?

- (2.3) A sapphire crystal doped with titanium absorbs strongly around 500 nm. Calculate the difference in the refractive index of the doped crystal above and below the 500 nm absorption band if the density of absorbing atoms is $1 \times 10^{25} \text{ m}^{-3}$. The refractive index of undoped sapphire is 1.77.
- (2.4) The laser crystal $\text{Ni}^{2+}:\text{MgF}_2$ has a broad absorption band in the blue which peaks at 405 nm and has a full width at half maximum of $8.2 \times 10^{13} \text{ Hz}$. The oscillator strength of the transition is 9×10^{-5} . Estimate the maximum absorption coefficient in a crystal with $2 \times 10^{26} \text{ m}^{-3}$ absorbing atoms per unit volume. The refractive index of the crystal is 1.39.
- (2.5) Show that the absorption coefficient of a Lorentz oscillator at the line centre does not depend on the value of ω_0 .
- (2.6) Figure 2.13 shows the refractive index of NaCl in the infrared spectral region. The data can be modelled approximately by assuming that the resonance feature is caused by the vibrations of the completely ionic Na^+Cl^- molecules. The atomic weights of sodium and chlorine are 23 and 35.5 respectively. Use the data to estimate:
- (i) The static dielectric constant of NaCl.
 - (ii) The natural oscillation frequency of the vibrations.
 - (iii) The restoring force for a unit displacement of the oscillator.
 - (iv) The density of NaCl molecules per unit volume.
 - (v) The damping constant γ for the vibrations.
 - (vi) The peak absorption coefficient.

* Exercises marked with an asterisk are more challenging.

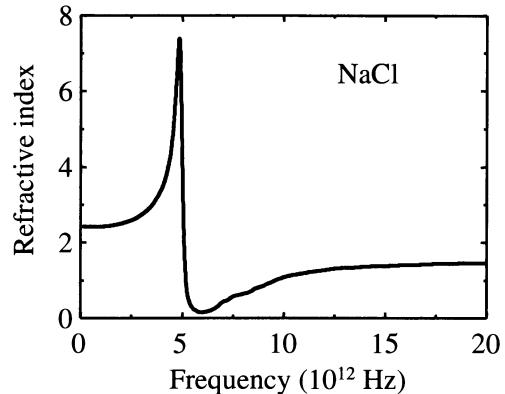


Fig. 2.13 Infrared refractive index of NaCl. After [1].

- (2.7) Derive eqn 2.26.
- (2.8)* Consider a simple Lorentz oscillator with a single undamped resonance. The dielectric constant will be given by eqn 2.14 with χ and γ both zero. This gives:
- $$\epsilon_r(\omega) = 1 + \frac{Ne^2}{\epsilon_0 m_0} \frac{1}{(\omega_0^2 - \omega^2)}.$$
- Prove that the group velocity is always less than c .
- (2.9)* Consider a dielectric sample placed in a uniform electric field pointing in the z direction as shown in Fig. 2.8. Assume that the atoms are arranged on a cubic lattice and the dipoles are all pointing along the external field direction.
- (i) Let us first consider the field generated by the dipoles within the spherical surface. By using the standard formula for the electric field generated by an electric dipole, show that the field at the centre of the sphere is given by
- $$\mathcal{E}_{\text{sphere}} = \frac{1}{4\pi\epsilon_0} \sum_j p_j \frac{3z_j^2 - r_j^2}{r_j^5}$$
- where the summation runs over all the dipoles within the surface except the one at the centre, and p_j is the dipole moment of the atom at the j th lattice site.
- (ii) Show that $\mathcal{E}_{\text{sphere}} = 0$ in a homogenous medium where all the p_j 's are the same.

- (iii) Now consider the uniformly polarized dielectric material outside the spherical hole. Let \mathbf{P} be the macroscopic polarization of the medium, which is assumed to be parallel to the external field. Show that the surface charge density on the sphere at an angle θ from the z axis is equal to $-P \cos \theta$. Hence show that the material outside the spherical surface generates a field at the centre of the sphere equal to $-\mathbf{P}/3\epsilon_0$.
- (2.10) Under what conditions does the Clausius–Mossotti relationship given by eqn 2.35 reduce to the usual relationship between the dielectric constant and the electric susceptibility given in eqn A.4?
- (2.11) The relative dielectric constant of N_2 gas at standard temperature and pressure is 1.000588. Calculate χ_a for the N_2 molecule. Show that the electric field strength required to generate a dipole equivalent to displacing the electron by 1 Å (10^{-10} m) is of a similar magnitude to the electric field between a proton and an electron separated by the same distance.
- (2.12) (a) Sellmeier derived the following equation for the wavelength dependence of the refractive index in 1871:
- $$n^2 = 1 + \sum_j \frac{A_j \lambda^2}{(\lambda^2 - \lambda_j^2)}.$$
- Show that this equation is equivalent to eqn 2.24 in regions of transparency far from any absorption lines. State the values of A_j and λ_j .
- (b) Assume that the dispersion is dominated by the closest resonance, so that we only need to include one term (say the one with $j = 1$) in the summation of Sellmeier's equation. Assume that λ_1^2/λ^2 is small, and expand Sellmeier's equation to derive the earlier dispersion formula determined empirically by Cauchy:
- $$n = C_1 + \frac{C_2}{\lambda^2} + \frac{C_3}{\lambda^4} + \dots$$
- State the values of C_1 , C_2 and C_3 in terms of A_1 and λ_1 .
- (2.13) The refractive index of crown glass is 1.5553 at 402.6 nm and 1.5352 at 706.5 nm.
- (i) Determine the coefficients C_1 and C_2 in Cauchy's formula given in the previous question, on the assumption that the term in C_3 is negligible.
- (ii) Estimate the refractive index for blue light at 450 nm and for red light at 650 nm.
- (iii) White light strikes a crown glass prism with an apex angle of 60°, as shown in Fig. 2.10. The angle of incidence with the first surface is 45°. Calculate the difference in the angle between the light at 450 nm and 650 nm at the exit surface of the prism.
- (2.14) Show that the temporal broadening of a short pulse by a dispersive medium of length L is given approximately by:
- $$\Delta\tau = \frac{L}{c} \left(\lambda^2 \frac{d^2 n}{d\lambda^2} \right) \frac{\Delta\lambda}{\lambda}$$
- where λ is the vacuum wavelength and $\Delta\lambda$ is the spectral width of the pulse. Estimate $\Delta\tau$ for an ultrashort laser pulse with a temporal width of 1 ps in 1 m of optical fibre at 1550 nm, where $\lambda^2 d^2 n/d\lambda^2 = -0.01$.
- (2.15) Consider the propagation of a wave with polarization vector components (x, y, z) , where $x^2 + y^2 + z^2 = 1$, in a birefringent medium. The dielectric constant experienced by the wave is conveniently described by the index ellipsoid:
- $$\frac{x^2}{\epsilon_{11}/\epsilon_0} + \frac{y^2}{\epsilon_{22}/\epsilon_0} + \frac{z^2}{\epsilon_{33}/\epsilon_0} = 1,$$
- where the ϵ_{ij} are the components of the dielectric constant tensor defined in eqn 2.45. The use of the index ellipsoid can be justified by considering the direction of the energy flow through the crystal: see Born and Wolf (1999). Use the index ellipsoid to show that the refractive index for the e-ray propagating at an angle θ to the optic axis of a uniaxial crystal as shown in Fig. 2.12(b) is given by:
- $$\frac{1}{n(\theta)^2} = \frac{\sin^2 \theta}{n_e^2} + \frac{\cos^2 \theta}{n_o^2},$$
- where n_e and n_o are defined in Example 2.2.
- (2.16) A uniaxial birefringent crystal made from quartz has $n_o = 1.5443$ and $n_e = 1.5534$. A wave plate is made by cutting the crystal so that the optic axis is parallel to the surfaces of the plate. The crystal will function as a quarter wave plate if the phase difference between the o- and e-rays is 90°, turning light polarized at 45° to the optic axis into circularly polarized light. Calculate the thickness of the crystal if it behaves as a quarter wave plate at 500 nm.
- (2.17) Look up the crystal structure of the following materials to determine whether they are birefringent or not: (a) NaCl, (b) diamond, (c) graphite (in the infrared, where it transmits), (d) ZnS (wurtzite structure), (e) ZnS (zinc blende structure), (f) solid argon at 4 K, (g) sulphur. Specify which, if any, of the birefringent materials are biaxial.

Interband absorption

3

We noted in Section 1.4.1 that semiconductors and insulators have a fundamental absorption edge in the near-infrared, visible or ultraviolet spectral region. The absorption edge is caused by the onset of optical transitions across the fundamental band gap of the material. This naturally leads us to investigate the physical processes that occur when electrons are excited between the bands of a solid by making optical transitions. This process is called **interband absorption**, and is the subject of the present chapter. The opposite process of **interband luminescence**, in which electrons drop from excited state bands by emitting photons, is discussed in Chapter 5.

Interband transitions are observed in all solids. Our objective here is to understand how the absorption spectrum of a given material is related to its band structure, and in particular to the density of states for the transition. We will postpone the discussion of excitonic effects on the absorption spectra to Chapter 4. We will concentrate on crystalline semiconductors, which illustrate the main points very clearly. The principles that we will find can easily be adapted to other materials, as required. This is done, for example, in Section 7.3.2, where we consider the effects of interband transitions on the reflectivity spectra of metals.

The understanding of interband absorption is based on applying the quantum mechanical treatment of the light-matter interaction to the band states of solids. This presupposes a working knowledge of both quantum mechanics and band theory. A summary of the main results required for the chapter is given in Appendices B and C. The reader is recommended to refer to the quantum mechanics and solid state physics texts listed in these appendices if any of the material is unfamiliar.

3.1 Interband transitions

The energy level diagram of isolated atoms consists of a series of states with discrete energies. Optical transitions between these levels give rise to sharp lines in the absorption and emission spectra. We have to use quantum mechanics to calculate the transition energies and the oscillator strengths. Once we have done this, we can obtain a good understanding of the frequency dependence of the refractive index and absorption coefficient by applying the classical oscillator model described in the previous chapter.

The optical transitions of solids are more complicated to deal with. Some of the properties that apply to the individual atoms carry over, but new physics arises as a result of the formation of bands with their delocalized states. The classical model has difficulty dealing with continuous absorption bands rather

3.1	Interband transitions	49
3.2	The transition rate for direct absorption	51
3.3	Band edge absorption in direct gap semiconductors	54
3.4	Band edge absorption in indirect gap semiconductors	63
3.5	Interband absorption above the band edge	66
3.6	Measurement of absorption spectra	68
3.7	Semiconductor photodetectors	69

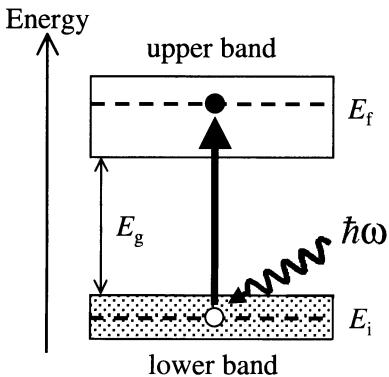


Fig. 3.1 Interband optical absorption between an initial state of energy E_i in an occupied lower band and a final state at energy E_f in an empty upper band. The energy difference between the two bands is E_g .

We will consider the effect of the attractive force between the negative electron and the positive hole in Chapter 4. At this stage we will ignore these effects and concentrate on investigating the general features of interband absorption.

than discrete lines, and we must develop new techniques to describe the frequency dependence of the optical properties. We can only expect the classical oscillator model to work with any accuracy when the frequency is far away from the absorption transitions between the bands.

Figure 3.1 shows a highly simplified energy diagram of two separated bands in a solid. The gap in energy between the bands is called the band gap E_g . Interband optical transitions will be possible between these bands if the selection rules allow them. During the transition an electron jumps from the band at lower energy to the one above it by absorbing a photon. This can only happen if there is an electron in the initial state in the lower band. Furthermore, the Pauli exclusion principle demands that the final state in the upper band must be empty. A typical example of a situation where this applies is the transitions across the fundamental band gap of a semiconductor or insulator. In this case, a photon excites an electron from the filled valence band to the empty conduction band.

By applying the law of conservation of energy to the interband transition shown in Fig. 3.1 we can see that:

$$E_f = E_i + \hbar\omega. \quad (3.1)$$

where E_i is the energy of the electron in the lower band, E_f is the energy of the final state in the upper band, and $\hbar\omega$ is the photon energy. Since there is a continuous range of energy states within the upper and lower bands, the interband transitions will be possible over a continuous range of frequencies. The range of frequencies is determined by the upper and lower energy limits of the bands.

It is apparent from Fig. 3.1 that the minimum value of $(E_f - E_i)$ is E_g . This implies that the absorption shows a threshold behaviour: interband transitions will not be possible unless $\hbar\omega > E_g$. Interband transitions therefore give rise to a continuous absorption spectrum from the low energy threshold at E_g to an upper value set by the extreme limits of the participating bands. This contrasts with the absorption spectrum of isolated atoms which consist of discrete lines.

The excitation of the electron leaves the initial state at energy E_i in the lower band unoccupied. This is equivalent to the creation of a hole in the initial state. The interband absorption process therefore creates a hole in the initial state and an electron in the final state, and may be considered as the creation of an electron-hole pair.

In the sections that follow, we will study how the interband absorption rate depends on the band structure of the solid. At this stage we just make one general distinction based on whether the band gap is direct or indirect. This point is illustrated in Fig. 3.2. Figure 3.2(a) shows the $E-k$ diagram of a solid with a direct band gap, while Fig. 3.2(b) shows the equivalent diagram for an indirect gap material. The distinction concerns the relative positions of the conduction band minimum and the valence band maximum in the Brillouin zone. In a direct gap material, both occur at the zone centre where $k = 0$. In an indirect gap material, however, the conduction band minimum does not occur at $k = 0$, but rather at some other value of k which is usually at the zone edge or close to it.

The distinction between the nature of the band gap has very important consequences for the optical properties. We will see in Section 3.2 that conservation

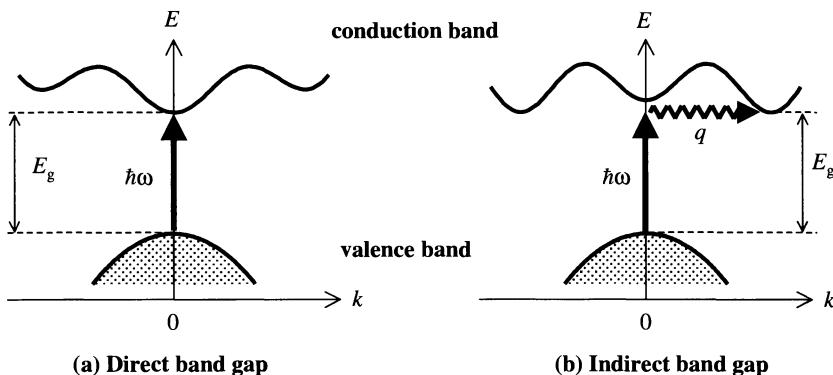


Fig. 3.2 Interband transitions in solids: (a) direct band gap, (b) indirect band gap. The vertical arrow represents the photon absorption process, while the wiggly arrow in part (b) represents the absorption or emission of a phonon.

of momentum implies that the electron wave vector does not change significantly during a photon absorption process. We therefore represent photon absorption processes by vertical lines on E - k diagrams. It is immediately apparent from Fig. 3.2(b) that the electron wave vector must change significantly in jumping from the valence band to the bottom of the conduction band if the band gap is indirect. It is not possible to make this jump by absorption of a photon alone: the transition must involve a phonon to conserve momentum. This contrasts with a direct gap material in which the process may take place without any phonons being involved.

Indirect absorption plays a very significant role in technologically important materials such as silicon. The treatment of indirect absorption is more complicated than direct absorption because of the role of the phonons. We will therefore begin our discussion of interband transitions by restricting our attention to direct processes. Interband absorption processes in indirect gap materials will be considered in Section 3.4.

3.2 The transition rate for direct absorption

The optical absorption coefficient α is determined by the quantum mechanical transition rate $W_{i \rightarrow f}$ for exciting an electron in an initial quantum state ψ_i to a final state ψ_f by absorption of a photon of angular frequency ω . Our task is therefore to calculate $W_{i \rightarrow f}$, and hence to derive the frequency dependence of α . As discussed in Appendix B, the transition rate is given by Fermi's golden rule:

$$W_{i \rightarrow f} = \frac{2\pi}{\hbar} |M|^2 g(\hbar\omega). \quad (3.2)$$

The transition rate thus depends on two factors:

- the matrix element M ,
 - the density of states $g(\hbar\omega)$.

In the discussion below, we consider the matrix element first, and then consider $g(\hbar\omega)$ afterwards.

The bracket symbol $\langle f | H' | i \rangle$ is an example of Dirac notation. The ‘ket’ $|i\rangle$ represents the wave function ψ_i , while the ‘bra’ $\langle f |$ represents ψ_f^* . The closed ‘bra-cket’ with the perturbation in the middle indicates that we evaluate the expectation value written out explicitly in the second line of eqn 3.3.

The matrix element describes the effect of the external perturbation caused by the light wave on the electrons. It is given by:

$$\begin{aligned} M &= \langle f | H' | i \rangle \\ &= \int \psi_f^*(\mathbf{r}) H'(\mathbf{r}) \psi_i(\mathbf{r}) d^3\mathbf{r}, \end{aligned} \quad (3.3)$$

where H' is the perturbation associated with the light wave, and \mathbf{r} is the position vector of the electron. We adopt here the semiclassical approach in which we treat the electrons quantum mechanically, but the photons are described by electromagnetic waves.

In classical electromagnetism, the presence of a perturbing electric field $\boldsymbol{\mathcal{E}}$ causes a shift in the energy of a charged particle equal to $-\mathbf{p} \cdot \boldsymbol{\mathcal{E}}$, where \mathbf{p} is the dipole moment of the particle. The appropriate quantum perturbation to describe the electric dipole interaction between the light and the electron is therefore:

$$H' = -\mathbf{p}_e \cdot \boldsymbol{\mathcal{E}}_{\text{photon}}, \quad (3.4)$$

where \mathbf{p}_e is the electron dipole moment and is equal to $-er$. This form for the perturbation is justified more rigorously in Section B.2 of Appendix B.

The light wave is described by plane waves of the form

$$\boldsymbol{\mathcal{E}}_{\text{photon}}(\mathbf{r}) = \boldsymbol{\mathcal{E}}_0 e^{\pm i\mathbf{k} \cdot \mathbf{r}}, \quad (3.5)$$

where the sign in the phase depends on the direction of propagation of the wave. The perturbation is thus:

$$H'(\mathbf{r}) = e\boldsymbol{\mathcal{E}}_0 \cdot \mathbf{r} e^{\pm i\mathbf{k} \cdot \mathbf{r}}. \quad (3.6)$$

The electron states in a crystalline solid are described by Bloch functions. This allows us to write the wave functions as a product of a plane wave and an envelope function that has the periodicity of the crystal lattice, as discussed in Section 1.5.2 and Appendix C. We therefore write:

$$\psi_i(\mathbf{r}) = \frac{1}{\sqrt{V}} u_i(\mathbf{r}) e^{i\mathbf{k}_i \cdot \mathbf{r}} \quad (3.7)$$

$$\psi_f(\mathbf{r}) = \frac{1}{\sqrt{V}} u_f(\mathbf{r}) e^{i\mathbf{k}_f \cdot \mathbf{r}}. \quad (3.8)$$

where u_i and u_f are the appropriate envelope functions for the initial and final bands respectively, and V is the normalization volume. \mathbf{k}_i and \mathbf{k}_f are the wave vectors of the initial and final electron states.

On substituting the perturbation of eqn 3.6 and the wave functions of eqns 3.7 and 3.8 into eqn 3.3, we obtain:

$$M = \frac{e}{V} \int u_f^*(\mathbf{r}) e^{-i\mathbf{k}_f \cdot \mathbf{r}} (\boldsymbol{\mathcal{E}}_0 \cdot \mathbf{r} e^{\pm i\mathbf{k} \cdot \mathbf{r}}) u_i(\mathbf{r}) e^{i\mathbf{k}_i \cdot \mathbf{r}} d^3\mathbf{r}, \quad (3.9)$$

where the limits of the integration are over the whole crystal. This integral can be simplified by invoking conservation of momentum and Bloch’s theorem. Conservation of momentum demands that the change in crystal momentum of the electron must equal the momentum of the photon, that is:

$$\hbar\mathbf{k}_f - \hbar\mathbf{k}_i = \pm\hbar\mathbf{k}. \quad (3.10)$$

This is equivalent to requiring that the phase factor in eqn 3.9 must be zero. If the phase factor is not zero, the different unit cells within the crystal will be out of phase with each other and the integral will sum to zero. Bloch's theorem requires that u_i and u_f are periodic functions with the same periodicity as the lattice.

These two considerations imply that we can separate the integral over the whole crystal into a sum over identical unit cells, because the unit cells are equivalent and in phase. We thus obtain:

$$|M| \propto \int_{\text{unit cell}} u_i^*(\mathbf{r}) x u_f(\mathbf{r}) d^3\mathbf{r}, \quad (3.11)$$

where we have defined our axes in such a way that the light is polarized along the x axis. This matrix element represents the electric dipole moment of the transition. Its evaluation requires knowledge of the envelope functions u_i and u_f . These functions are derived from the atomic orbitals of the constituent atoms, and so each material has to be considered separately.

The conservation of momentum condition embodied in eqn 3.10 can be simplified further by considering the magnitude of the wave vectors of the electrons and photons. The wave vector of the photon is $2\pi/\lambda$, where λ is the wavelength of the light. Optical frequency photons therefore have k values of about 10^7 m^{-1} . The wave vectors of the electrons, however, are much larger. This is because the electron wave vector is related to the size of the Brillouin zone, which is equal to π/a , where a is the unit cell dimension. Since $a \sim 10^{-10} \text{ m}$, the photon wave vector is much smaller than the size of a Brillouin zone. Therefore we may neglect the photon momentum in eqn 3.10 in comparison to the electron momentum and write:

$$\mathbf{k}_f = \mathbf{k}_i. \quad (3.12)$$

A direct optical transition therefore leads to a negligible change in the wave vector of the electron. This is why we represent the absorption processes by vertical arrows in the electron E - k diagrams such as the ones shown in Fig. 3.2.

The $g(\hbar\omega)$ factor that appears in eqn 3.2 is the **joint density of states** evaluated at the photon energy. As explained in Section 1.5.4, the density of states function describes the distribution of the states within the bands. The *joint* density of states accounts for the fact that both the initial and final electron states lie within continuous bands. For electrons within a band, the density of states per unit energy range $g(E)$ is obtained from:

$$g(E) dE = 2 g(k) dk, \quad (3.13)$$

where $g(k)$ is the density of states in momentum space. The extra factor of 2 here compared to eqn 1.29 allows for the fact that there are two electron spin states for each allowed k -state. This gives:

$$g(E) = \frac{2g(k)}{dE/dk}, \quad (3.14)$$

where dE/dk is the gradient of the E - k dispersion curve in the band diagram. $g(k)$ itself is worked out by calculating the number of k -states in the incremental volume between shells in k -space of radius k and $k + dk$. This is

equal to the number of states per unit volume of k -space, namely $1/(2\pi)^3$ (see Exercise 3.1), multiplied by the incremental volume $4\pi k^2 dk$. Hence $g(k)$ is given by the standard formula:

$$\begin{aligned} g(k)dk &= \frac{1}{(2\pi)^3} 4\pi k^2 dk \\ \Rightarrow g(k) &= \frac{k^2}{2\pi^2}. \end{aligned} \quad (3.15)$$

We can then work out $g(E)$ by using eqn 3.14 if we know the relationship between E and k from the band structure of the material. For electrons in a parabolic band with effective mass m^* , $g(E)$ is given by (see Exercise 3.2):

$$g(E) = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}}. \quad (3.16)$$

This is just the standard formula for free electrons but with the free electron mass m_0 replaced by m^* .

The joint density of states factor is finally obtained by evaluating $g(E)$ at E_i and E_f when they are related to $\hbar\omega$ through the details of the band structure. We will see how to do this in the case of parabolic bands in Section 3.3.3. The density of atoms in a solid is very large, and so the density of states factor will be high and the transition rate correspondingly large. It is common to find values of α in the range 10^6 – 10^8 m^{-1} for the direct absorption coefficient in a solid.

The results derived here are of general applicability for interband transitions in crystalline solids. To proceed further, we need detailed knowledge of the band structure. In the next section we will use these general results to derive the frequency dependence of the absorption near the band edge of a direct gap semiconductor.

3.3 Band edge absorption in direct gap semiconductors

The basic process for an optical transition across the fundamental band gap of a direct gap semiconductor is shown in Fig. 3.2(a). An electron is excited from the valence band to the conduction band by absorption of a photon. The transition rate is evaluated by working out the matrix element and the density of states, as discussed in the preceding section. These factors are considered separately below.

3.3.1 The atomic physics of the interband transitions

The matrix element that we need to evaluate is given in eqn 3.11. This allows us to calculate the probability for electric dipole transitions if we know the atomic character of the envelope wave functions $u_i(\mathbf{r})$ and $u_f(\mathbf{r})$. The full treatment of this problem employs group theory to determine the character of the bands involved. This approach is beyond the scope of our present discussion, and at this level we will just offer a few intuitive arguments.

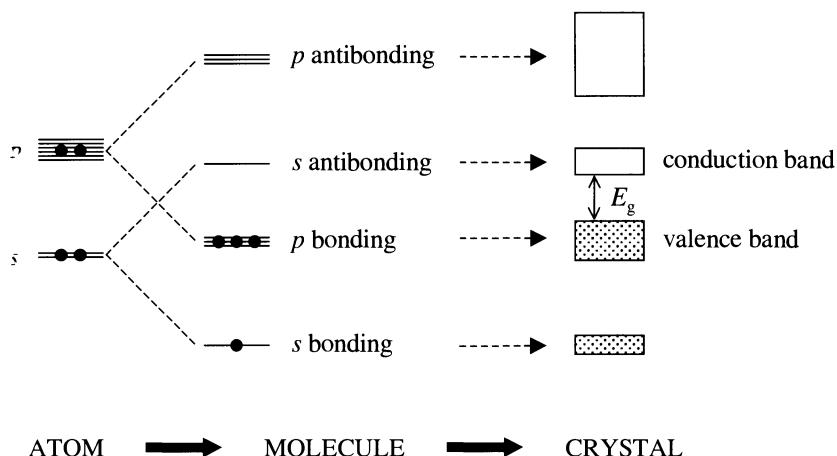


Fig. 3.3 Schematic diagram of the electron levels in a covalent crystal made from four-valent atoms such as germanium or binary compounds such as gallium arsenide. The s and p states of the atoms hybridize to form bonding and antibonding molecular orbitals, which then evolve into the conduction and valence bands of the semiconductor.

The semiconductors that we will be considering all have four valence electrons. This is obvious in the case of the elemental semiconductors such as silicon and germanium, which come from group IV of the periodic table. It is also true, however, for the binary compounds made from elements symmetrically displaced from group IV of the periodic table. The covalent bond in these compounds is made by sharing the electrons in such a way that each atom ends up with four electrons. For example, the bond in the III–V compounds is formed by sharing the five valence electrons from the group V element with the three from the group III element, giving a total of eight electrons for every two atoms. It is energetically favourable to do this because it is then possible to form very stable covalent crystals with a structure similar to diamond. Similar arguments apply to the II–VI semiconductor compounds.

The valence electrons of a four-valent atom are derived from the s and p orbitals. For example, the electronic configuration of germanium is $4s^24p^2$. In the crystalline phase the adjacent atoms share the valence electrons with each other in a covalent bond. Figure 3.3 shows schematically the evolution of the s and p -like atomic states, through the s and p bonding and antibonding orbitals of the molecule, to the valence and conduction bands of the crystalline solid. The level ordering shown is appropriate for most III–V and II–VI semiconductors, as well as germanium. The level ordering in silicon is different: see Exercise 3.9.

The evolution of the levels shown in Fig. 3.3 makes it apparent that the top of the valence band has a p -like atomic character, while the bottom of the conduction band is s -like. This is because the four valence electrons occupy the four bonding orbitals, which then evolve into the valence band. The top of the valence band is derived from the p bonding orbitals, while the bottom of the conduction band originates from the s antibonding orbitals. Therefore optical transitions from the valence band to the conduction band are from p -like states to s -like states. We know from the selection rules for dipole transitions that $p \rightarrow s$ transitions are allowed. (See Exercise 3.3 and Section B.3 in Appendix B.) Hence we conclude that the transitions between the valence band and the conduction band of a semiconductor with a level ordering such as the one shown in Fig. 3.3 are electric-dipole allowed.

The conclusion of this discussion is that the probability for interband transitions across the band gap in materials like germanium or the III–V compounds is high, and we therefore expect to observe strong absorption. This is indeed the case, as we will see below. The discussion of germanium is complicated because it has an indirect band gap. We will therefore concentrate our attention on the III–V compound semiconductor gallium arsenide. GaAs has a direct band gap, and the level ordering follows the scheme shown in Fig. 3.3. The transitions across the gap are therefore both dipole-allowed and direct. This makes GaAs a standard example for considering direct interband transitions. It is also a very important material for optoelectronic applications.

3.3.2 The band structure of a direct gap III–V semiconductor

The high symmetry points of Brillouin zones are given symbolic names. The zone centre where $\mathbf{k} = (0, 0, 0)$ is called the Γ point. The zone edges along the (100) and (111) directions are called the X and L points respectively. In the Brillouin zone of the diamond or zinc-blende lattice, the wave vectors at the X and L points are $\mathbf{k} = \frac{2\pi}{a}(1, 0, 0)$ and $\frac{\pi}{a}(1, 1, 1)$ respectively, where a is the length of the cube edge of the face-centred cubic lattice from which the diamond or zinc-blende structure is derived. The $\Gamma \rightarrow X$ direction is labelled Δ , while the $\Gamma \rightarrow L$ direction is labelled Λ . See Appendix C for further details.

The band structure of GaAs in the energy range near the fundamental band gap is shown in Fig. 3.4. The energy E of the electrons in the different bands is plotted against the electron wave vector \mathbf{k} . GaAs has the zinc-blende structure, which is based on the face-centred cubic (f.c.c.) lattice. The band dispersion is shown for increasing \mathbf{k} along two different directions of the Brillouin zone. The right hand side of the figure corresponds to moving from the zone centre where $\mathbf{k} = (0, 0, 0)$ along the (100) direction to the zone edge at $\mathbf{k} = \frac{2\pi}{a}(1, 0, 0)$, a being the length of the cube edge in the f.c.c. lattice. The left hand side corresponds to moving from $\mathbf{k} = 0$ along the body diagonal direction until reaching the zone edge at $\mathbf{k} = \frac{\pi}{a}(1, 1, 1)$.

The figure is divided into a shaded region and an unshaded region. The shading represents the occupancy of the levels in the bands: bands that fall in the shaded region are below the Fermi level and are full of electrons. The three bands in the shaded region therefore correspond to valence band states. The single band above the shaded region is empty of electrons and is therefore the conduction band. The three bands in the valence band correspond to the

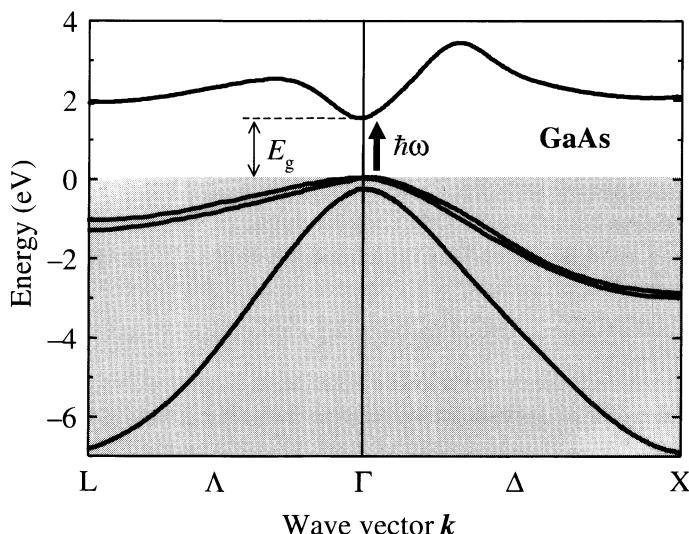


Fig. 3.4 Band structure of GaAs. The dispersion of the bands is shown for two directions of the Brillouin zone: $\Gamma \rightarrow X$ and $\Gamma \rightarrow L$. The Γ point corresponds to the zone centre with a wave vector of $(0, 0, 0)$, while the X and L points correspond respectively to the zone edges along the (100) and (111) directions. The valence bands are below the Fermi level and are full of electrons. This is indicated by the shading in the figure. After [1], copyright 1976 American Physical Society, reprinted with permission.

three p bonding orbitals shown in Fig. 3.3, while the single conduction band corresponds to the s antibonding state. This correspondence between the bands and the molecular orbitals is strictly valid only at the Γ point at the Brillouin zone centre. The atomic character (or more accurately, the symmetry) of the bands actually changes as k increases, and is only well defined at high symmetry points in the Brillouin zone such as Γ , X or L .

In this section we are interested in the transitions that take place across the band gap for small k values close to the Γ point. This means that the correspondence to Fig. 3.3 will be justified in our discussion here. We can therefore assume that the transitions are dipole-allowed, and concentrate on working out the density of states for the transition. To do this, it is helpful to make use of the simplified four-band model shown in Fig. 3.5. This model band diagram is typical of direct gap III–V semiconductors near $k = 0$. There is a single s -like conduction band and three p -like valence bands. All four bands have parabolic dispersions. The positive curvature of the conduction band on the E – k diagram indicates that it corresponds to an electron (e) band, while the negative curvature of the valence bands correspond to hole states. Two of the hole bands are degenerate at $k = 0$. These are known as the heavy (hh) and light hole (lh) bands, the heavy hole band being the one with the smaller curvature. The third band is split-off to lower energy by the spin–orbit coupling, and is known as the split-off (so) hole band. The energy difference between the maximum of the valence band and the minimum of the conduction band is the band gap E_g , while the spin–orbit splitting between the hole bands at $k = 0$ is usually given the symbol Δ .

The schematic diagram of Fig. 3.5 should be compared with the detailed band structure of GaAs shown previously in Fig. 3.4. The maxima of the valence band occur at the Γ -point of the Brillouin zone, while the conduction band has a ‘camel back’ structure, with minima at the Γ -point, the L -point and near the X -point. We can neglect the subsidiary minima at the L -point and near the X -point here because momentum conservation does not allow direct transitions to these states from the top of the valence band. The bands near the zone centre are all approximately parabolic, and so the simplified picture in Fig. 3.5 is valid near $k = 0$.

The three valence band states all have p -like atomic character, and so it is possible to have electric dipole transitions from each of the bands to the conduction band. Two such transitions are indicated on Fig. 3.5. As noted earlier, these absorption processes are represented by vertical arrows on the E – k diagram. This means that the k vector of the electron and hole created by the transition are the same (cf. eqn 3.12). The transition labelled 1 involves the excitation of an electron from the heavy hole band to the electron band. Transition 2 is the corresponding process originating in the light hole band. Direct transitions are also possible from the split-off band to the conduction band, but these are not shown in the figure for clarity.

3.3.3 The joint density of states

The frequency dependence of the absorption coefficient can now be calculated if we know the joint density of states factor given in eqn 3.14. This can be calculated analytically for the simplified band structure shown in Fig. 3.5.

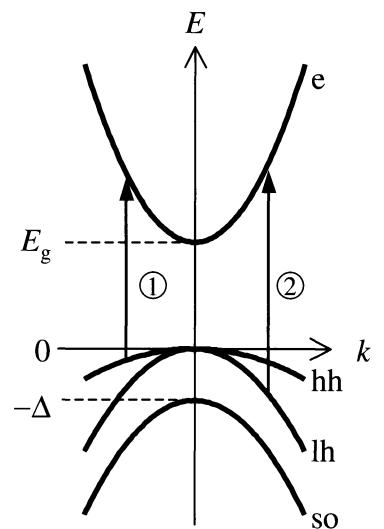


Fig. 3.5 Band structure of a direct gap III–V semiconductor such as GaAs near $k = 0$. $E = 0$ corresponds to the top of the valence band, while $E = E_g$ corresponds to the bottom of the conduction band. Four bands are shown: the heavy hole (hh) band, the light hole (lh) band, the split-off hole (so) band, and the electron (e) band. Two optical transitions are indicated. Transition 1 is a heavy hole transition, while transition 2 is a light hole transition. Transitions can also take place between the split-off hole band and the conduction band, but these are not shown for the sake of clarity. This four-band model was originally developed for InSb in reference [2].

The dispersion of the bands is determined by their respective effective masses, namely m_e^* for the electrons, m_{hh}^* for the heavy holes, m_{lh}^* for the light holes, and m_{so}^* for the split-off holes. This allows us to write the following E - k relationships for the conduction, heavy hole, light hole, and split-off hole bands respectively:

$$E_c(k) = E_g + \frac{\hbar^2 k^2}{2m_e^*} \quad (3.17)$$

$$E_{hh}(k) = -\frac{\hbar^2 k^2}{2m_{hh}^*} \quad (3.18)$$

$$E_{lh}(k) = -\frac{\hbar^2 k^2}{2m_{lh}^*} \quad (3.19)$$

$$E_{so}(k) = -\Delta - \frac{\hbar^2 k^2}{2m_{so}^*}. \quad (3.20)$$

It is evident from Fig. 3.5, that conservation of energy during a heavy hole or light hole transition requires that:

$$\hbar\omega = E_g + \frac{\hbar^2 k^2}{2m_e^*} + \frac{\hbar^2 k^2}{2m_h^*}, \quad (3.21)$$

where $m_h^* = m_{hh}^*$ or m_{lh}^* for the heavy or light hole transition respectively. We define the reduced electron–hole mass μ according to:

$$\frac{1}{\mu} = \frac{1}{m_e^*} + \frac{1}{m_h^*}. \quad (3.22)$$

This allows us to rewrite eqn 3.21 in the simpler form:

$$\hbar\omega = E_g + \frac{\hbar^2 k^2}{2\mu}. \quad (3.23)$$

We are interested in evaluating $g(E)$ with $E = \hbar\omega$. The joint electron–hole density of states can be worked out by substituting eqn 3.23 into eqns 3.14 and 3.15. This gives:

$$\text{For } \hbar\omega < E_g, \quad g(\hbar\omega) = 0.$$

$$\text{For } \hbar\omega \geq E_g, \quad g(\hbar\omega) = \frac{1}{2\pi^2} \left(\frac{2\mu}{\hbar^2} \right)^{\frac{3}{2}} (\hbar\omega - E_g)^{\frac{1}{2}}. \quad (3.24)$$

We therefore see that the density of states factor rises as $(\hbar\omega - E_g)^{\frac{1}{2}}$ for photon energies greater than the band gap.

3.3.4 The frequency dependence of the band edge absorption

Now that we have discussed the matrix element and the density of states, we can put it all together and deduce the frequency dependence of the absorption coefficient α . Fermi's golden rule given in eqn 3.2 tells us that the absorption rate for a dipole-allowed interband transition is proportional to the joint density

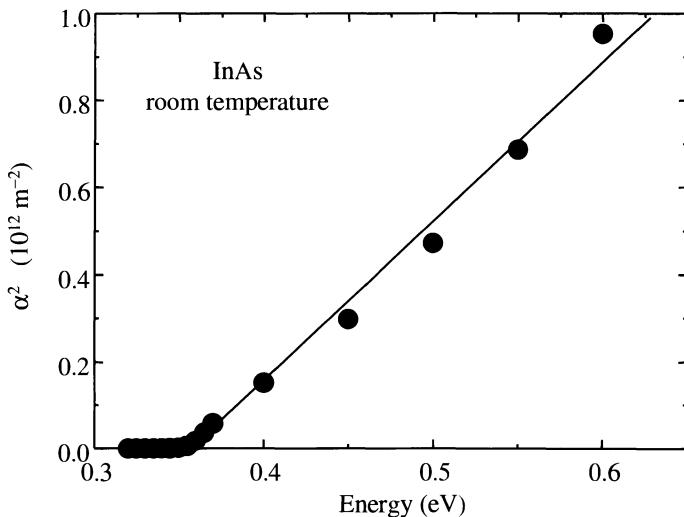


Fig. 3.6 Square of the optical absorption coefficient α versus photon energy for the direct gap III–V semiconductor InAs at room temperature. The band gap can be deduced to be 0.35 eV by extrapolating the absorption to zero. After [3].

of states given by eqn 3.24. We therefore expect the following behaviour for $\alpha(\hbar\omega)$:

$$\begin{aligned} \text{For } \hbar\omega < E_g, \quad \alpha(\hbar\omega) &= 0. \\ \text{For } \hbar\omega \geq E_g, \quad \alpha(\hbar\omega) &\propto (\hbar\omega - E_g)^{\frac{1}{2}}. \end{aligned} \quad (3.25)$$

There is no absorption if $\hbar\omega < E_g$, and the absorption increases as $(\hbar\omega - E_g)^{1/2}$ for photon energies greater than the band gap. We also expect that transitions with larger reduced masses will give rise to stronger absorption due to the $\mu^{3/2}$ factor in eqn 3.24.

The predictions of eqn 3.25 can be compared to experimental data. Figure 3.6 shows results for the absorption coefficient of the direct gap III–V semiconductor indium arsenide at room temperature. The graph plots α^2 against the photon energy in the spectral region close to the band gap. The straight line relationship between α^2 and $(\hbar\omega - E_g)$ indicates that the model we have developed is a good one. The band gap can be read from the data as the point at which the absorption goes to zero. This gives a value of 0.35 eV, which is in good agreement with values deduced from electrical measurements. Note that the values of the absorption coefficient are very large. This is a consequence of the very large density of states in the solid phase.

In many III–V semiconductors, including GaAs itself, it is found that the frequency dependence predicted by eqn 3.25 is only approximately obeyed. There are a number of reasons for this.

- We have neglected the Coulomb attraction between the electron and hole, which can enhance the absorption rate and cause exciton formation. These effects become stronger as the band gap gets larger and the temperature is lowered. This is why we have presented room temperature data for a semiconductor with a smallish band gap in Fig. 3.6. Excitonic effects are very significant in materials like GaAs even at room temperature. This point will be discussed further in Chapter 4, and is clearly apparent in the absorption data for GaAs shown in Fig. 4.3.

- The semiconductor may contain impurity or defect states with energies within the band gap, and these may allow absorption for photon energies less than the band gap. This point is discussed in Section 7.4.2.
- The parabolic band approximations embodied in the dispersion relations of eqns 3.17–3.20 is only valid near $k = 0$. As the photon energy increases above the band gap, the joint density of states will no longer obey the frequency dependence given in eqn 3.24. In these cases we must use the full band structure to evaluate the density of states. We will give a discussion of how this is done in Section 3.5 below.

Example 3.1

Indium phosphide is a direct gap III–V semiconductor with a band gap of 1.35 eV at room temperature. The absorption coefficient at 775 nm is $3.5 \times 10^6 \text{ m}^{-1}$. A platelet sample 1 μm thick is made with antireflection coated surfaces. Estimate the transmission of the sample at 620 nm.

Solution

The transmission can be calculated from eqn 1.6. The sample is antireflection coated, and so we may take $R = 0$. We therefore need to work out $\exp(-\alpha l)$, where α is the absorption coefficient at 620 nm, and $l = 1 \mu\text{m}$. 775 nm corresponds to a photon energy of 1.60 eV, which is greater than E_g . Similarly, 620 nm corresponds to a photon energy of 2.00 eV, which is also above E_g . We can therefore use eqn 3.25 and write:

$$\frac{\alpha(620 \text{ nm})}{\alpha(775 \text{ nm})} = \frac{(2.00 - E_g)^{\frac{1}{2}}}{(1.60 - E_g)^{\frac{1}{2}}} = 1.6$$

where we have used $E_g = 1.35 \text{ eV}$. This implies that $\alpha(620 \text{ nm}) = 5.6 \times 10^6 \text{ m}^{-1} \equiv 5.6 \mu\text{m}^{-1}$. We thus obtain the final result:

$$T(620 \text{ nm}) = \exp(-\alpha l) = \exp(-5.6) = 0.37 \text{ %}.$$

This value for T is only an estimate because we have ignored the excitonic effects and we have assumed that the parabolic band approximation is valid, even though we are quite a long way above E_g . The experimental value of $\alpha(620 \text{ nm})$ is actually about 15 % larger than the value calculated here.

3.3.5 The Franz–Keldysh effect

The modification of the band edge absorption by the application of an external electric field \mathcal{E} was studied independently by W. Franz and L.V. Keldysh in 1958. They showed that there are two main effects:

- The absorption coefficient for photon energies less than E_g is no longer zero, as stated in eqn 3.25, but now decreases exponentially with $(E_g - \hbar\omega)$. The frequency dependence of α is given by:

$$\alpha(\hbar\omega) \propto \exp\left(-\frac{4\sqrt{2m_e^*}}{3|e|\hbar\mathcal{E}}(E_g - \hbar\omega)^{3/2}\right). \quad (3.26)$$

This implies that the band edge shifts to lower energy as the field is increased. (See Exercise 3.11.)

- The absorption coefficient for $\hbar\omega > E_g$ is modulated by an oscillatory function. The oscillations in $\alpha(\hbar\omega)$ are called Franz–Keldysh oscillations.

These two effects are collectively known as the **Franz–Keldysh effect**. They are typically observed when the semiconductor is incorporated as a thin i-region at the junction of a p-n diode. This allows controllable fields to be applied by varying the bias on the device, as explained in Appendix D.

It can be seen from the Kramers–Kronig relationship given in eqn 2.36 that a change in the absorption coefficient will produce changes in the refractive index at frequencies below the band gap. Thus the application of the electric field modulates both the absorption and the refractive index of the material. This modulation of the optical constants by the electric field is an example of an **electro-optic effect**. The changes may be either linear or quadratic in the field. In Sections 11.3.1 and 11.4.1 of Chapter 11, we will explain how these effects can be described in terms of nonlinear optical susceptibility tensors.

The changes in the real and imaginary parts of the refractive index produced by the electric field imply that the reflectivity will also be changed through eqn 1.26. This is the basis of the technique of **electroreflectance**, in which the modulation of the reflectivity in response to an AC electric field is measured as a function of the photon energy. The electroreflectance technique is widely used to determine important band structure parameters.

3.3.6 Band edge absorption in a magnetic field

It is well known in classical physics that the application of a strong magnetic field with flux density B causes electrons to perform circular motion around the field at the **cyclotron frequency** ω_c given by (see Exercise 3.12):

$$\omega_c = \frac{eB}{m_0}. \quad (3.27)$$

In classical physics, the radius of the orbit and the energy can have any values, but in quantum physics, they are both quantized. The quantized energies are given by:

$$E_n = (n + \frac{1}{2})\hbar\omega_c, \quad (3.28)$$

where $n = 0, 1, 2, \dots$. These quantized energy levels are called **Landau levels**.

Consider a semiconductor in the presence of a strong magnetic field along the z direction. The motion of the electrons in the conduction band and holes in the valence band will be quantized in the x, y plane, but their motion will still be free in the z direction. Their energies within the bands will thus be given by:

$$E^n(k_z) = (n + \frac{1}{2})\frac{e\hbar B}{m^*} + \frac{\hbar^2 k_z^2}{2m^*}, \quad (3.29)$$

where m^* is the appropriate effective mass. The first term gives the energy of the quantized motion in the (x, y) plane, while the second describes the free motion along the z direction. In absolute terms relative to $E = 0$ at the top of

the valence band, the electron and hole energies are given by:

$$\begin{aligned} E_n^e(k_z) &= E_g + (n + \frac{1}{2}) \frac{e\hbar B}{m_e^*} + \frac{\hbar^2 k_z^2}{2m_e^*}, \\ E_n^h(k_z) &= -(n + \frac{1}{2}) \frac{e\hbar B}{m_h^*} - \frac{\hbar^2 k_z^2}{2m_h^*}. \end{aligned} \quad (3.30)$$

These are equivalent to eqns 3.17–3.19, which are valid at $B = 0$.

If the sample is illuminated when the field is applied, an interband transition can take place in which an electron is created in the conduction band and a hole is created in the valence band. It can be shown that the Landau level number n does not change during the interband transition. (See Exercise 3.12.) This selection rule implies that the electron and hole must have the same value of n . Furthermore, the k_z value of both particles must be the same because the photon has negligible momentum. Therefore, the transition energy will be given by:

$$\begin{aligned} \hbar\omega &= E_n^e(k_z) - E_n^h(k_z) \\ &= E_g + (n + \frac{1}{2}) \frac{e\hbar B}{\mu} + \frac{\hbar^2 k_z^2}{2\mu}, \end{aligned} \quad (3.31)$$

where μ is the reduced mass given in eqn 3.22. Equation 3.31 should be compared to eqn 3.23 which applies when $B = 0$. The term in k_z is unchanged, but the x and y components of \mathbf{k} are now quantized by the magnetic field.

The frequency dependence of the absorption coefficient which follows from eqn 3.31 is considered in detail in Exercise 3.13. In brief, we expect very high absorption at any photon energy that can satisfy eqn 3.31 with $k_z = 0$. This gives rise to a series of equally spaced peaks in the absorption spectrum with energies given by

$$\hbar\omega = E_g + (n + \frac{1}{2}) \frac{e\hbar B}{\mu}; \quad n = 0, 1, 2, \dots. \quad (3.32)$$

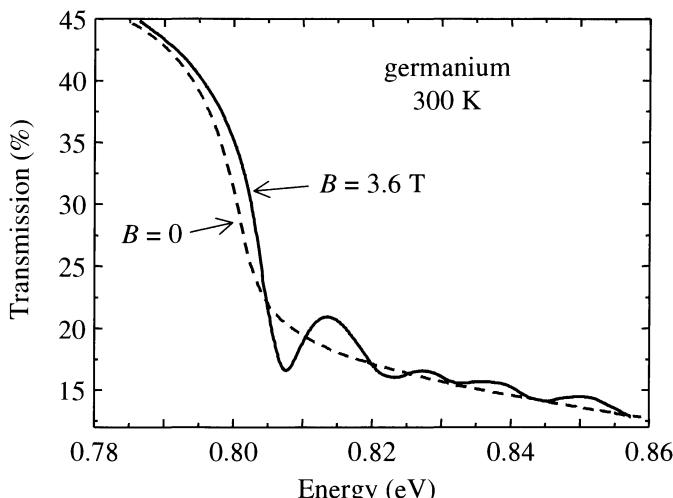


Fig. 3.7 Transmission spectrum of germanium for $B = 0$ and $B = 3.6$ T at 300 K. After [4], copyright 1957 American Physical Society, reprinted with permission.

One immediate consequence of this result is that we expect the absorption edge to shift to higher energy by $\hbar e B / 2\mu$ in the magnetic field.

Figure 3.7 shows the room temperature transmission spectrum of germanium at $B = 0$ and $B = 3.6$ T. We see that at $B = 3.6$ T the absorption edge is indeed shifted to higher energy and there is a regularly spaced series of dips in the transmission, as predicted by eqn 3.32. The spectral width of the dips is determined mainly by line broadening due to scattering. The electron effective mass can be determined from the energies of the minima in the transmission: see Exercise 3.13.

3.4 Band edge absorption in indirect gap semiconductors

In the previous two sections, we have been concentrating on direct interband transitions. As it happens, several of the most important semiconductors have indirect band gaps, most notably: silicon and germanium. Indirect gap semiconductors have their conduction band minimum away from the Brillouin zone centre, as shown schematically in Fig. 3.2(b). Transitions at the band edge must therefore involve a large change in the electron wave vector. Optical frequency photons only have a very small k vector, and it is not possible to make this transition by absorption of a photon alone: the transition must involve a phonon to conserve momentum.

Consider an indirect transition that excites an electron in the valence band in state (E_i, \mathbf{k}_i) to a state (E_f, \mathbf{k}_f) in the conduction band. The photon energy is $\hbar\omega$, while the phonon involved has energy $\hbar\Omega$ and wave vector \mathbf{q} . Conservation of energy demands that:

$$E_f = E_i + \hbar\omega \pm \hbar\Omega, \quad (3.33)$$

while conservation of momentum requires that:

$$\hbar\mathbf{k}_f = \hbar\mathbf{k}_i \pm \hbar\mathbf{q}. \quad (3.34)$$

The \pm factors allow for the possibility of phonon absorption or emission, with the $+$ sign corresponding to absorption, and the $-$ sign to emission. We have neglected the photon's momentum in eqn 3.34. This approximation was justified previously in connection with eqn 3.12.

Before considering the shape of the band edge absorption spectrum, we can first make a general point. Indirect transitions involve both photons and phonons. In quantum mechanical terms, this is a second-order process: a photon must be destroyed, and a phonon must be either created or destroyed. This contrasts with direct transitions which are first-order processes because no phonons are involved. The transition rate for indirect absorption is therefore much smaller than for direct absorption.

The smaller transition rate for indirect processes is clearly shown by the data given in Fig. 3.8, which compares the band edge absorption of silicon and GaAs. Silicon has an indirect band gap at 1.12 eV, while GaAs has a direct gap at 1.42 eV. We see that the absorption rises much faster with frequency in the direct gap material, and soon exceeds the indirect material even though its

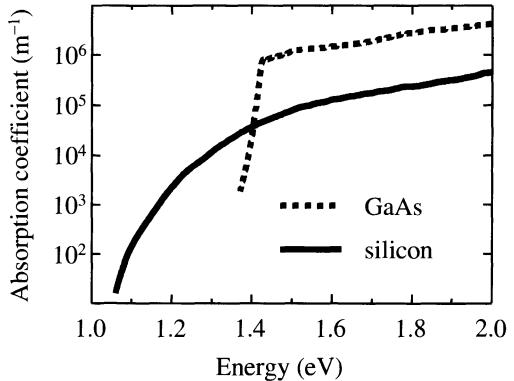


Fig. 3.8 Comparison of the absorption coefficient of GaAs and silicon near their band edges. GaAs has a direct band gap at 1.42 eV, while silicon has an indirect gap at 1.12 eV. Note that the vertical axis is logarithmic. After [3].

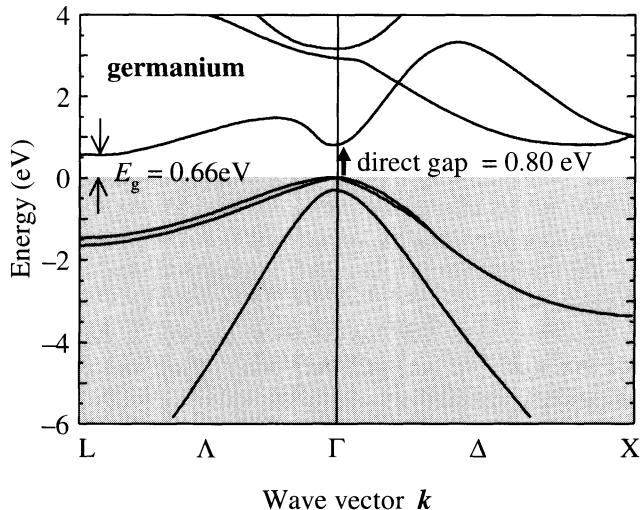


Fig. 3.9 Band structure of germanium. After [5], copyright 1988 Springer-Verlag, reprinted with permission.

band gap is larger. The absorption of GaAs is roughly an order of magnitude larger than that of silicon for energies greater than ~ 1.43 eV.

The derivation of the quantum mechanical transition rate for an indirect gap semiconductor is beyond the scope of this book. The results of such a calculation give the following result:

$$\alpha^{\text{indirect}}(\hbar\omega) \propto (\hbar\omega - E_g \mp \hbar\Omega)^2. \quad (3.35)$$

This shows that we expect the absorption to have a threshold close to E_g , but not exactly at E_g . The difference is $\mp \hbar\Omega$, depending on whether the phonon is absorbed or emitted. Note that the frequency dependence is different to that for direct gap semiconductors given in eqn 3.25. This provides a convenient way to determine whether the band gap is direct or not. Furthermore, the involvement of the phonons gives other tell-tale signs that the band gap is indirect, as we will discuss below.

Indirect absorption has been thoroughly studied in materials like germanium. The band structure of germanium is shown in Fig. 3.9. The overall shape of the band dispersion is fairly similar to that of GaAs given in Fig. 3.4. There is, however, one very important qualitative difference: the lowest conduction

It is not surprising that the band structures of GaAs and Ge are similar. Gallium and arsenic lie on either side of germanium in the periodic table, and so GaAs and Ge are approximately isoelectronic materials.

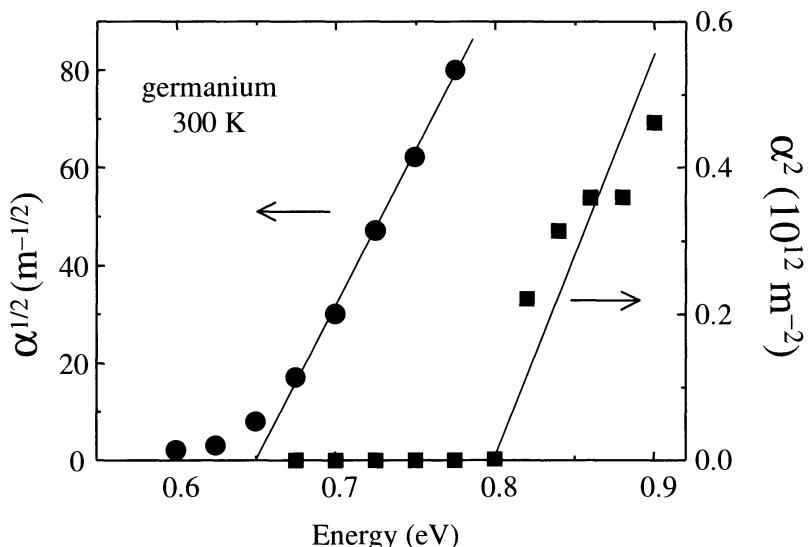


Fig. 3.10 Experimental data for the absorption coefficient of germanium at room temperature. After [3].

band minimum of germanium occurs at the L point, where $\mathbf{k} = \frac{\pi}{a}(1, 1, 1)$, and not at $k = 0$. This makes germanium an indirect gap semiconductor. The value of the indirect gap is 0.66 eV, which corresponds to the band gap determined by electrical measurements. This is 0.14 eV smaller than the direct gap at $k = 0$.

Figure 3.10 shows the results of absorption measurements on germanium at room temperature. $\sqrt{\alpha}$ is plotted against $\hbar\omega$ in the spectral region close to the band gap at 0.66 eV. The data fits well to a straight line, which confirms the prediction of eqn 3.35. The data extrapolates back to 0.65 eV, which indicates from eqn 3.35 that a phonon of energy ~ 0.01 eV has been absorbed. The wave vector \mathbf{q} of the phonon must be equal to that of an electron at the L-point of the Brillouin zone. The energies of the four different phonon modes with the required wave vector are listed in Table 3.1, from where we see that it must be TA phonons that are involved. The experimental data also shows a tail extending down to about 0.60 eV. This is caused by absorption of the higher frequency phonons and also multiphonon absorption.

The probability for phonon absorption is proportional to the number of phonons present, which is given by the Bose–Einstein formula:

$$f_{BE}(E) = \frac{1}{\exp(E/k_B T) - 1}. \quad (3.36)$$

The variation of the phonon populations with temperature implied by eqn 3.36 leads to a characteristic temperature dependence of the absorption edge. As we decrease T , the contributions due to phonon absorption gradually freeze out, starting with the highest energy phonons. At very low temperatures, we would not expect to observe any phonon absorption at all, because there would be no phonons excited with enough energy. On the other hand, phonon emission is possible at all temperatures. Thus at the lowest temperatures, the indirect absorption edge would be determined by phonon emission rather than phonon absorption. This behaviour contrasts with direct gap materials, in which the absorption edge merely shifts with the band gap as the temperature is varied.

Table 3.1 Phonon energies for germanium at the L point where $\mathbf{q} = \frac{\pi}{a}(1, 1, 1)$, a being the unit cell size. After [6].

Mode	$\hbar\Omega$ (eV)
Longitudinal acoustic (LA)	0.027
Transverse acoustic (TA)	0.008
Longitudinal optic (LO)	0.030
Transverse optic (TO)	0.035

The band diagram for germanium in Fig. 3.9 shows that direct transitions can occur if the photon energy exceeds 0.80 eV. In this case, we would expect that the absorption would begin to follow eqn 3.25 instead of eqn 3.35. This is indeed borne out by the data. The black squares in Fig. 3.10 show the absorption data on a new scale, with α^2 plotted against the photon energy. A clear threshold at 0.80 eV is observed, and the frequency dependence approximately obeys $\alpha^2 \propto (\hbar\omega - E_g^{\text{dir}})$, where $E_g^{\text{dir}} = 0.80$ eV is the band gap for direct transitions at the Γ point. Note that the direct absorption completely dominates over the indirect processes once we have crossed the second threshold at E_g^{dir} . The indirect absorption is much weaker, and is insignificant when plotted on the same scale as the direct absorption as in Fig. 3.10. This highlights the second-order nature of the indirect absorption.

3.5 Interband absorption above the band edge

Up to this point, we have been concentrating on the absorption near the band edge. As we will see in Chapter 5, the reason for doing this is that the optical properties at the band edge determine the emission spectra. This does not mean that the rest of the absorption spectrum is uninteresting: it is just more complicated to deal with because the parabolic band approximation does not apply. However, as we will see below, much useful information about the full band structure can be obtained from analysis of the overall spectrum.

It is not possible to give explicit formulae for the full frequency dependence of the absorption spectrum as we did for the band edge absorption in eqns 3.25 and 3.35. Instead, we have to work out dE/dk in eqn 3.14 from the full band structure. In this section we will illustrate how this is done for the case of silicon. The principles described here can be applied to other materials if the band structure is known.

Figure 3.11 shows the interband absorption spectrum of silicon up to 10 eV. Two features at about 3.5 eV and 4.3 eV are readily identified in the data. These two energies are labelled E_1 and E_2 and are related to aspects of the

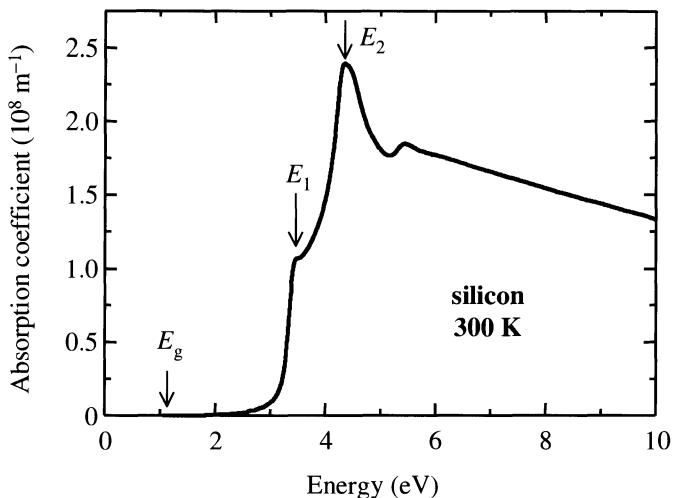


Fig. 3.11 Interband absorption spectrum of silicon at 300 K up to 10 eV. The transition at the band gap E_g is indirect and involves a phonon. The energies E_1 and E_2 correspond to critical points where the conduction and valence bands are parallel to each other. This can be seen more clearly in the band structure diagram given in Fig. 3.12. After [3].

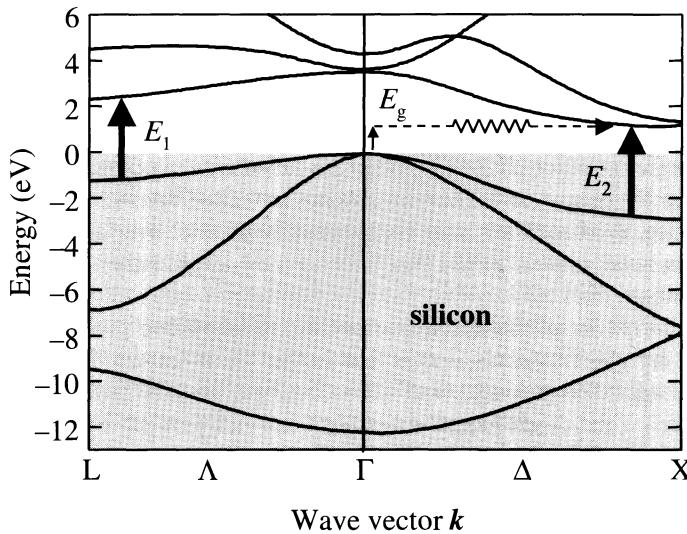


Fig. 3.12 Band structure of silicon. The band gap E_g is indirect and occurs at 1.1 eV. The conduction and valence bands are approximately parallel along the (111) and (100) directions near the zone edges at the L and X points. The separation of the bands in these regions are labelled E_1 (3.5 eV) and E_2 (4.3 eV) respectively. The absorption at these energies is very high due to the Van Hove singularities in the joint density of states. After [5], copyright 1988 Springer-Verlag, reprinted with permission.

band structure, as discussed further below. The absorption coefficient in the spectral region around E_1 and E_2 is extremely large, with values of α in excess of 10^8 m^{-1} . This should be compared to values of 10^2 – 10^6 m^{-1} in the spectral region immediately above the band gap E_g at 1.1 eV. (See Fig. 3.8.) Indeed, the band edge absorption is completely negligible on the scale of the data shown in Fig. 3.11. This is a consequence of two factors. Firstly, the band edge absorption is weak because it is indirect, and secondly, the density of states at the band edge is comparatively small. The measured absorption spectrum is actually dominated by direct absorption at photon energies where the density of states is very high.

Figure 3.12 shows the band structure of silicon along the (100) and (111) directions. The band gap E_g is indirect and has a value of 1.1 eV, with the conduction band minimum located near the X-point of the Brillouin zone. Direct transitions can take place between any state in the valence band and the conduction band states directly above it, if the transitions are dipole-allowed. The minimum direct separation between the conduction and valence bands occurs near the L point, where the transition energy is 3.5 eV. The energy of these transitions is labelled E_1 , and corresponds to the sharp increase in the absorption at 3.5 eV observed in the data shown in Fig. 3.11. The separation of the conduction and valence bands near the X point is also significant. This energy is labelled E_2 and corresponds to the absorption maximum at 4.3 eV.

The transitions near the L and X points are particularly important because of the ‘camel’s back’ shape of the conduction band. This means that the conduction band ends up having a negative curvature near these points of the Brillouin zone. The curvature is more or less the same as that in the valence band, so that the conduction and valence bands are approximately parallel to each other. This means that we can have direct transitions with the same photon energy for many different values of k . The joint density of states factor is therefore very high at E_1 and E_2 , and we expect the absorption to be correspondingly high. This is indeed observed in the experimental data: the absorption rises

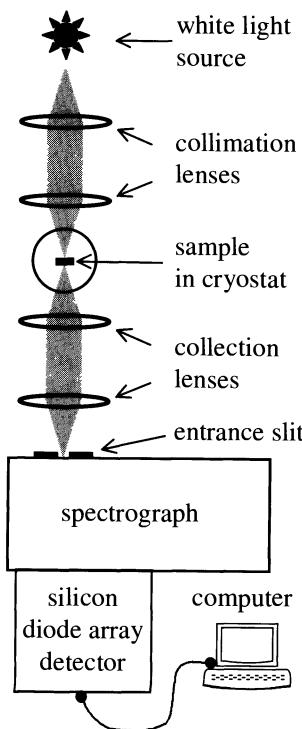


Fig. 3.13 Schematic diagram of an experimental arrangement to measure absorption spectra in the wavelength range 200–1000 nm using a silicon diode array detector. At longer wavelengths the spectrograph and silicon detector is replaced with a scanning monochromator and an infrared detector, or with a Fourier transform spectrometer.

sharply at E_1 and reaches a peak at E_2 . The absolute values of the absorption coefficient are extremely large: over 10^8 m^{-1} as we have already noted.

In a region of the Brillouin zone where the bands are parallel, the photon energy E for direct transitions does not depend on k . This implies that dE/dk is zero, and hence that the joint density of states $g(E)$ diverges (cf. eqn 3.14). The energies at which dE/dk vanishes are called **critical points**, and the corresponding divergences in the density of states are called **Van Hove singularities**. In practice, the bands are only approximately parallel over a portion of the Brillouin zone, and so $g(\hbar\omega)$ just become very large, rather than diverging completely.

The discussion of the absorption coefficient of silicon given here can be adapted to other materials if their band structure is known. The absorption strength will be proportional to the joint density of states, which will be particularly high if the conduction and valence bands are parallel to each other. An example of how this is done in the case of metals is discussed in Section 7.3.2 of Chapter 7.

3.6 Measurement of absorption spectra

The absorption coefficient of a material is usually determined by measuring the transmission of a thin platelet sample. A typical experimental arrangement to do this is shown in Fig. 3.13. The light from a low intensity white light source such as a tungsten bulb is passed through the sample, and the spectrum of the transmitted light is recorded with a spectrograph and a silicon diode array detector. The transmission coefficient is determined by calculating the ratio of the light on the detector with and without the sample present. The absorption coefficient is then calculated from the transmission using eqn 1.6, after measuring the reflectivities in a separate experiment. By placing the sample in a helium cryostat, the absorption coefficient can be measured as a function of temperature down to 2 K.

The arrangement shown in Fig. 3.13 can be used for measurements within the spectral response of silicon, namely ~ 200 –1000 nm. At longer wavelengths the spectrograph and silicon detector must be replaced by a scanning monochromator with a suitable infrared detector. The criteria used to select the optimal detector for a particular spectral region are discussed in Section 3.7.1. At wavelengths beyond $\sim 5 \mu\text{m}$, it is common to use Fourier transform spectrometers for absorption measurements.

In some materials, the absorption is so strong that it is impractical to use transmission measurements to determine α . We have already seen, for example, that the absorption coefficient of silicon exceeds 10^8 m^{-1} at some wavelengths, which would mean that the transmission of a very thin sample of thickness $0.1 \mu\text{m}$ would still be less than 0.01 %. In this case, the absorption is calculated from the imaginary part of the complex refractive index using eqn 1.16. κ itself is deduced from the measured reflectivity spectra $R(\hbar\omega)$ using eqn 1.26. This might seem impossible at first sight, because R depends on both n and κ . However, we know from Section 2.2.5 that n and κ are not completely independent variables and must be related to each other through the Kramers–Kronig relationships. Hence by self-consistent fitting of the re-

flectivity spectra using the Kramers–Kronig formulæ given in eqns 2.36 and 2.37, we can determine both n and κ from $R(\hbar\omega)$, and hence deduce α from κ .

3.7 Semiconductor photodetectors

The strong absorption found in semiconductors is the basis of semiconductor photodetectors. Light with photon energy greater than the band gap is absorbed in the semiconductor, and this creates free electrons in the conduction band and free holes in the valence band. The presence of the light can therefore be detected either by measuring a change in the resistance of the sample or by measuring an electrical current in an external circuit. In this section we consider the operating principles of two different types of detector, and then discuss the use of semiconductor detectors in solar cells.

3.7.1 Photodiodes

Figure 3.14 shows a schematic diagram of a photodiode detector. The detector consists of a p-n junction with a thin intrinsic (undoped) layer sandwiched in the depletion region, forming a p-i-n structure. The band alignments and electrostatics of this type of structure are discussed in Appendix D. The diode is operated in reverse bias. This ensures that there is only a very small current in the circuit when no light is present, and applies a very strong DC electric field \mathcal{E} across the i-region. Photons absorbed in the i-region generate electron–hole pairs, that are rapidly swept towards the contacts by the field, and hence into the external circuit. The current generated in this way is called the photocurrent.

Consider a photodiode of active length l illuminated by a light beam of optical power P and angular frequency ω . The flux of photons per unit time on the detector is $P/\hbar\omega$. From the definition of the absorption coefficient given in eqn 1.4, we can deduce that the fraction of light absorbed in a length l is equal to $(1 - e^{-\alpha l})$, where α is the absorption coefficient at frequency ω . Each absorbed photon generates one electron–hole pair, and we define the quantum efficiency η as the fraction of these carriers that flow into the external circuit. The magnitude of the photocurrent I_{pc} is thus given by:

$$I_{pc} = e\eta \frac{P}{\hbar\omega} (1 - e^{-\alpha l}). \quad (3.37)$$

We have assumed here that the top surface of the detector has been antireflection coated to prevent the wasteful reflection of incident photons. We have also assumed that the absorption in any layers above the active region is negligible.

The responsivity of the device is the ratio of the photocurrent I_{pc} to the optical power P , and is given by:

$$\text{Responsivity} = \frac{I_{pc}}{P} = \frac{\eta e}{\hbar\omega} (1 - e^{-\alpha l}) \quad \text{amps per watt.} \quad (3.38)$$

Equation 3.38 shows us that in order to obtain a large responsivity we need a high absorption and high quantum efficiency. Ideally, we would like to have both η and $(1 - e^{-\alpha l})$ to be equal to unity, in which case the responsivity is simply $e/\hbar\omega$. This sets an upper limit on the responsivity that can be achieved.

Many basic detectors just use p-n structures without the i-region. The light is absorbed in the depletion region at the junction, where there are no free carriers. The p-i-n structure is preferable because of the faster response times than can be achieved. It is, however, more complicated to make.

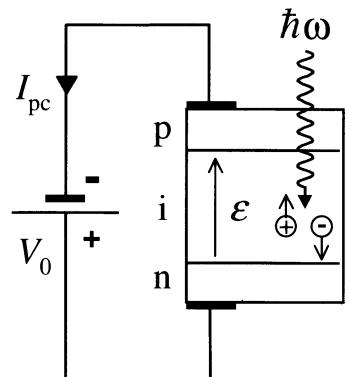


Fig. 3.14 Schematic diagram of a p-i-n photodiode. The diode is operated in reverse bias with a positive voltage V_0 applied to the n-region. This generates a strong DC electric field \mathcal{E} across the i-region. Absorption of photons in the i-region creates free electrons \ominus and holes \oplus that are attracted to the n-region and p-regions respectively by the field. The carriers that reach the doped regions flow into the external circuit, thereby generating the photocurrent I_{pc} .

For example, the maximum possible responsivity for a 2 eV photon ($\lambda = 620$ nm) is 0.5 AW^{-1} . Well designed photodiodes can come quite close to this ideal figure.

The design of practical photodiodes is based on several criteria.

- The choice of the semiconductor is made to optimize the responsivity while ensuring a fast response and low noise. The most fundamental criterion is that the band gap must be smaller than the photon energy. Having satisfied this criterion, we want E_g to be as large as possible to minimize the noisy dark current that arises from the thermal excitation of electrons and holes across the gap. At the same time, we want a material in which the electron and hole mobilities are high so that the photogenerated carriers can be swept quickly across the device and give a fast response time.
- Materials with direct band gaps are better than those with indirect gaps because the absorption is higher. With typical values of α over 10^6 m^{-1} for direct absorption, the thickness of the active layer needs only to be $\sim 1 \mu\text{m}$ to achieve very strong absorption. In an indirect gap semiconductor, greater thicknesses are required, which increases the constraints on the purity of the material. Furthermore, the direct gap materials can give faster response times because the thinner i-region reduces the transit time of the device.
- The top contact should be designed to transmit as much of the light into the i-region as possible. This means that the top contact should be made very thin. A better solution is to use different semiconductors for the p-n junction and i-region, such that the top contact has a larger band gap than the photons to be detected. This is possible with modern epitaxial semiconductor growth technology.

Table 3.2 Common semiconductor photodetectors. E_g : band gap, T : operating temperature, λ_{\max} : maximum wavelength that can be detected. The band gap of alloy semiconductors such as InGaAs and HgCdTe can be varied by altering the composition. The compositions listed here correspond to typical values used in detectors.

Semiconductor	E_g (eV)	T (K)	λ_{\max} (μm)
Si	1.1	300	1.1
In _{0.53} Ga _{0.47} As	0.75	300	1.65
Ge	0.66	300	1.9
Ge	0.73	77	1.7
InAs	0.42	77	3.0
InSb	0.23	77	5.2
Hg _{0.8} Cd _{0.2} Te	0.09	77	14

All these physical considerations have to be weighed against the manufacturing costs.

Table 3.2 gives a list of several common types of semiconductor photodetectors. Silicon is extensively used at visible and near-infrared wavelengths, despite the fact that the absorption is indirect. This choice is determined by the advanced technology of the silicon industry. Germanium detectors can be used out to $1.9 \mu\text{m}$, but for more demanding applications in the wavelength range $1\text{--}1.6 \mu\text{m}$, the III-V alloy semiconductor InGaAs is becoming increasingly important. This is because it has a direct gap and also has a higher electron mobility than Ge, which means that fast, efficient detectors can be made for the telecommunications wavelengths of $1.3 \mu\text{m}$ and $1.5 \mu\text{m}$.

At wavelengths beyond $1.9 \mu\text{m}$, narrow gap semiconductors such as InAs or InSb have to be used. These long wavelength detectors invariably require cryogenic cooling to suppress the thermal dark currents and achieve good signal to noise ratios. The II-VI alloy semiconductor HgCdTe is frequently used for wavelengths beyond $5 \mu\text{m}$. It has a band gap which can be varied according to the composition, and detectors with peak sensitivities in the range $5\text{--}14 \mu\text{m}$ are available. HgCdTe detectors are therefore able to cover several technologically important infrared wavelengths, especially $10.6 \mu\text{m}$, which corresponds to one of the infrared windows in the atmosphere and also to the emission lines of the CO₂ laser. In Section 6.7 of Chapter 6 we will describe

an alternative detector for $10.6\text{ }\mu\text{m}$ which has recently been developed using GaAs quantum wells. These quantum well detectors operate on a different principle to the interband detectors described here.

Example 3.2

Estimate the responsivity of a $10\text{ }\mu\text{m}$ thick antireflection coated silicon photodiode at 800 nm. Calculate the photocurrent generated when the diode is illuminated with a 1 mW beam from a semiconductor laser operating at this wavelength.

Solution

The responsivity is given by eqn 3.38. We can read a value of $\alpha \approx 1 \times 10^5\text{ m}^{-1} \equiv 0.1\text{ }\mu\text{m}^{-1}$ for silicon at 800 nm (1.55 eV) from Fig. 3.8. The device is antireflection coated, and we therefore assume that no optical power is lost at the front surface. A well designed photodiode will have negligible absorption in the top contact and quantum efficiency $\eta \approx 1$ at the operating wavelength. We therefore obtain:

$$\text{Responsivity} = \frac{e}{\hbar\omega}(1 - e^{-0.1 \times 10}) = 0.41 \text{ amps per watt.}$$

The photocurrent is given by the product of the responsivity and the optical power. The photocurrent will therefore be 0.41 mA.

3.7.2 Photoconductive devices

An alternative way to make a semiconductor photodetector is to use the photoconductive effect. This relies on the change of the conductivity of the material when illuminated by light. The conductivity is proportional to the density of free electrons and holes. The conductivity therefore increases due to the generation of free carriers after absorption of photons by interband transitions.

The devices consist of a sample with contacts at the ends so that a constant DC current can flow through the semiconductor between the contacts. The resistance between the contacts decreases upon illumination. This alters the voltage dropped across the device, and hence provides the detection mechanism. Photoconductive detectors are simpler to make than photodiodes, but tend to have slow response times.

3.7.3 Photovoltaic devices

Semiconductor photodiodes can also be operated in photovoltaic mode. In this mode of operation, the device does not have an external power supply. Instead it generates a photovoltage when irradiated by light. This in turn can be used to generate electrical power in an external load. This is the basis of operation of **solar cells**, which convert sunlight into electrical power.

The operating principle of a photovoltaic device relies on the relationship between the photocurrent and the applied bias in a photodiode. The photocurrent is sensitive to the bias because it affects the electric field \mathcal{E} across the depletion region. As explained in Appendix D, the field strength can be quite large even when the external bias is zero. This is because of the alignment of the Fermi levels in the p- and n-regions, which produces a voltage drop across the depletion region called the built-in voltage V_{bi} . The magnitude of V_{bi} is approximately equal to E_g/e . A forward bias approximately equal to V_{bi} must therefore be applied before \mathcal{E} drops to zero. The diode will produce a photocurrent on illumination provided that there is a field to sweep out the electrons and holes. Thus photocurrents can be produced at zero bias and even in forward bias, provided the forward bias voltage is less than V_{bi} .

Let us suppose that we replace the battery in Fig. 3.14 with an electrical load of resistance R . The voltage on the diode in the dark is zero. If the diode is illuminated, a photocurrent will be generated because the field due to the built-in voltage sweeps the carriers out of the i-region. This photocurrent flows through the load and the device therefore converts optical power to electrical power. The magnitude of the electrical power is $I_{pc}^2 R$. The direction of the photocurrent is such that the photovoltage $I_{pc} R$ across the load puts the diode in forward bias. Thus if the illumination level is increased from zero, the photocurrent will saturate as the photovoltage approaches V_{bi} . This limits the maximum amount of electrical power that can be generated. At present, the maximum power conversion efficiency achieved in a silicon solar cell is in the range 10–25 %.

Chapter summary

- Interband transitions occur when electrons jump to an excited state band by absorption of photons. The absorption process may be considered as the creation of an electron–hole pair.
- Interband absorption is only possible if the photon energy exceeds the band gap energy E_g . The absorption spectrum therefore shows a threshold at E_g .
- The absorption rate for direct transitions is proportional to the product of the joint density of states and the square of the electric dipole matrix element.
- The photon wave vector is negligible compared to that of the electron, and so the electron wave vector is unchanged in a direct transition. Direct transitions are represented by vertical arrows on E - k band diagrams.
- The frequency dependence of the absorption edge of a direct gap semiconductor at E_g is given by eqn 3.25. At higher frequencies, the absorption coefficient is determined by the detailed frequency dependence of the joint density of states. The absorption is particularly high at critical points.
- The application of an external electric field results in non-zero absorption below the band gap through the Franz–Keldysh effect. The application of a magnetic field causes the absorption edge to shift to higher energy.

- Interband transitions in indirect gap materials involve the absorption or emission of a phonon to conserve momentum in the process. Indirect absorption is much weaker than direct absorption since it is a second-order process.
- The frequency dependence of the absorption edge in an indirect gap material is given by eqn 3.35. This is different to that observed in direct gap semiconductors, and provides a way for determining the nature of the band gap experimentally.
- The absorption of light by interband transitions can be used to make photodetectors. The detection mechanism is based on the generation of a photocurrent or the increase of the conductivity after absorption of photons with energies greater than the band gap.

Further reading

The electronic states of solids are covered in the companion book of this series by Singleton (2001). They are also covered in all general solid state texts, for example Burns (1985), Ibach and Luth (1995) or Kittel (1996), and in more detail by Harrison (1999).

Detailed information on the interband absorption of semiconductors may be found in Klingshirn (1995), Pankove (1971), Seeger (1997), or Yu and Cardona (1996). Introductory treatments of the application of group theory to interband transitions can be found in Klingshirn (1995) or Yu and Cardona (1996). Yu and Cardona (1996) also give the derivation of the frequency dependence of the absorption coefficient in an indirect gap semiconductor (eqn 3.35).

The Franz–Keldysh effect is described in Klingshirn (1995), Seeger (1997), or Yu and Cardona (1996). Yu and Cardona (1996) explain the use of electroreflectance to determine band structure parameters, while Seeger (1997) gives a good discussion of the effect of magnetic fields on the band edge absorption.

The physics of semiconductor photodetectors is described in more detail in Bhattacharya (1997), Chuang (1995), Sze (1985), Wilson and Hawkes (1998), or Yariv (1997). Sze (1985) gives a good discussion of solar cells.

References

- [1] Chelikowsky, J.R. and Cohen, M.L. (1976). *Phys. Rev. B*, **14**, 556.
- [2] Kane, E.O. (1957). *J. Phys. Chem. Solids*, **1**, 249.
- [3] Palik, E.D. (1985). *Handbook of the optical constants of solids*. Academic Press, San Diego.
- [4] Zwerdling, S., Lax, B. and Roth, L.M. (1957). *Phys. Rev.*, **108**, 1402.
- [5] Cohen, M.L. and Chelikowsky, J. (1988). *Electronic structure and optical properties of semiconductors*. Springer-Verlag, Berlin.
- [6] Madelung, O. (1996). *Semiconductors, basic data* (2nd edn). Springer-Verlag, Berlin.

Exercises

- (3.1) Apply Born–von Karmen periodic boundary conditions (i.e. $e^{ikx} = e^{ik(x+L)}$ etc., where L is a macroscopic length) to show that the density of states in k -space is $1/(2\pi)^3$.

- (3.2) Show that the density of states for an electron with $E(k) = \hbar^2 k^2 / 2m^*$ is given by eqn 3.16.

- (3.3) The wave function of an atomic state with principal quantum number n , orbital quantum number l and magnetic quantum number m may be written in the form:

$$\psi_{nlm}(r, \theta, \phi) = R_{nl}(r)Y_{lm}(\theta, \phi),$$

where $R_{nl}(r)$ is the radial wave function, $Y_{lm}(\theta, \phi)$ is the spherical harmonic function, and (r, θ, ϕ) are the usual spherical polar coordinates. The spherical harmonic function itself may be written

$$Y_{lm}(\theta, \phi) = P_l^m(\cos \theta)e^{im\phi},$$

where $P_l^m(\cos \theta)$ is a polynomial function in $\cos \theta$. The parity of the spherical harmonic functions is equal to $(-1)^l$.

- (i) Explain what is meant by the ‘parity’ of an atomic wave function.
- (ii) By considering the parity of the wave functions, prove that the matrix element for an electric dipole transition defined by

$$M =$$

$$\left| \int_{r=0}^{\infty} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \psi_f^*(\mathbf{r})(-\epsilon \mathbf{r}) \psi_i(\mathbf{r}) r^2 \sin \theta dr d\theta d\phi \right|,$$

must be zero unless $|l|$ changes by an odd number during the transition.¹

- (iii) By writing the components of \mathbf{r} in spherical polar coordinates, prove that $\Delta m = 0$ if the light is polarized along the z direction, and $\Delta m = \pm 1$ for light polarized in the x or y direction.

- (3.4) Draw a schematic diagram of an experimental arrangement that could be used to obtain the absorption data shown in Fig. 3.6.

- (3.5) Explain how you would use optical absorption measurements to determine whether a semiconductor has a direct or indirect band gap.

- (3.6) Table 3.3 gives absorption data for gallium phosphide at 300 K. What can you deduce about the band structure of GaP from this data?

Table 3.3 Absorption coefficient α of GaP tabulated against photon energy E at 300 K. After [3].

E (eV)	α (m^{-1})	E (eV)	α (m^{-1})
2.2	3.12×10^1	2.7	7.39×10^5
2.3	7.79×10^3	2.8	3.35×10^6
2.4	2.72×10^4	2.9	5.38×10^6
2.5	6.43×10^4	3.0	6.81×10^6
2.6	1.44×10^5	3.1	8.64×10^6

- (3.7) Use the data given in Fig. 3.10 to estimate the absorption coefficient of germanium at 1200 nm.

- (3.8) The band parameters of the four-band model shown in Fig. 3.5 are given for GaAs in Table C.2.

(i) Calculate the k vector of the electron excited from the heavy hole band to the conduction band in GaAs when a photon of energy 1.6 eV is absorbed at 300 K. What is the corresponding value for the light hole transition?

(ii) Calculate the wave vector of the photon inside the crystal. Does this confirm the validity of the approximation given in eqn 3.12? The refractive index of GaAs at 1.6 eV is 3.7.

(iii) Calculate the ratio of the joint density of states for the heavy and light hole transitions.

(iv) What is the wavelength at which transitions from the split-off hole band become possible?

- (3.9)* In silicon the s -like antibonding orbital lies at a higher energy than the p -like antibonding orbitals, which contrasts with the ordering of the levels for Ge or GaAs shown in Fig. 3.3. This leads to major qualitative differences between the conduction band states of silicon and germanium at the Γ point, as can be seen by comparing Figs 3.9 and 3.12.

(i) Deduce the value of the direct band gap of silicon at the Γ point from the band structure diagram given in Fig. 3.12.

(ii) Explain qualitatively how the transitions at energies E_1 and E_2 can be dipole-allowed.

¹ By considering the properties of the function $P_l^m(\cos \theta)$, it is possible to prove that the selection rule on Δl is stricter than just being an odd number: Δl must in fact be equal to ± 1 .

* Exercises marked with an asterisk are more challenging.

- (3.10) Where would you expect to measure the optical absorption edge in germanium at 4 K? The indirect band gap is 0.74 eV at this temperature.
- (3.11) Estimate the electric field strength at which the band edge of GaAs is red-shifted by 0.01 eV. The electron effective mass is $0.067m_0$.
- (3.12)* Show that a classical particle of mass m and charge e performs circular orbits around a magnetic field with an angular frequency of eB/m , where B is the field strength. Show also that the selection rule for the Landau level number n during an interband transition is $\Delta n = 0$.
- (3.13)*
- (i) Show that the density of states of a particle which is free to move in one dimension only is proportional to $E^{-1/2}$, where E is the energy of the particle.
 - (ii) Draw a sketch of the frequency dependence of the optical absorption edge of a one-dimensional direct-gap semiconductor.
 - (iii) Explain why a bulk semiconductor in a strong magnetic field can be considered as a one-dimensional system. Hence explain the shape of the optical transmission spectrum of germanium at 300 K at 3.6 T given in Fig. 3.7.
 - (iv) Use the data in Fig. 3.7 to deduce values for the band gap and the electron effective mass of Ge on the assumption that $m_h^* \gg m_e^*$. Comment on the values you obtain.
- (3.14) The absorption coefficient of germanium is $4.6 \times 10^4 \text{ m}^{-1}$ at $1.55 \mu\text{m}$ and $7.5 \times 10^5 \text{ m}^{-1}$ at $1.30 \mu\text{m}$. Calculate the maximum responsivities of a germanium photodiode with a $10 \mu\text{m}$ thick absorbing layer at these two wavelengths.
- (3.15)
- (i) The capacitance of a reverse biased p-i-n photodiode can be calculated by treating the device as a parallel plate capacitor. Justify this approximation.
 - (ii) Calculate the capacitance of a silicon p-i-n photodiode with an area of 1 mm^2 and an i-region thickness of $10 \mu\text{m}$. The static dielectric constant of silicon is 11.9.
 - (iii) Estimate the time taken for the photogenerated electrons and holes to drift across the i-region when the reverse bias on the photodiode is 10 V. Assume that the built-in voltage is 1.1 V, and that the electron and hole mobilities of Si at room temperature are $0.15 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $0.045 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ respectively.
 - (iv) At what voltage would the electron transit time be equal to the RC time constant of the diode when connected to a 50Ω load?

4

Excitons

4.1 The concept of excitons	76
4.2 Free excitons	77
4.3 Free excitons in external fields	82
4.4 Free excitons at high densities	84
4.5 Frenkel excitons	86

In the previous chapter we discussed the absorption of photons by interband transitions. We saw that this process creates an electron in the conduction band and a hole in the valence band, but we neglected the effects of the mutual Coulomb attraction between them. As we will see in this chapter, the Coulomb interaction can give rise to the formation of new excitations of the crystal called excitons. These excitons have interesting optical properties and are important for optoelectronic applications.

We will encounter excitons in several different contexts throughout this book. In this chapter we will concentrate mainly on their effects on the absorption edge of bulk semiconductors. In Chapter 6 we will see how the excitonic effects can be enhanced by forming quantum well structures containing very thin layers of direct gap semiconductors such as GaAs. In Chapter 8 we will see how the excitonic effects have a strong influence on the optical properties of molecular materials. Finally, in Chapter 11 we will briefly study how the presence of excitons can give rise to useful nonlinear optical properties in bulk and quantum well semiconductors.

4.1 The concept of excitons

The absorption of a photon by an interband transition in a semiconductor or insulator creates an electron in the conduction band and a hole in the valence band. The oppositely charged particles are created at the same point in space and can attract each other through their mutual Coulomb interaction. This attractive interaction increases the probability of the formation of an electron–hole pair, and therefore increases the optical transition rate. Moreover, if the right conditions are satisfied, a bound electron–hole pair can be formed. This neutral bound pair is called an **exciton**. In the simplest picture, the exciton may be conceived as a small hydrogenic system similar to a positronium atom with the electron and hole in a stable orbit around each other.

Excitons are observed in many crystalline materials. There are two basic types:

- Wannier–Mott excitons, also called **free excitons**;
- Frenkel excitons, also called **tightly bound excitons**.

The Wannier–Mott excitons are mainly observed in semiconductors, while the Frenkel excitons are found in insulator crystals and molecular crystals.

The two generic types of exciton are illustrated schematically in Fig. 4.1. The diagrams show an electron and hole orbiting around each other within a crystal. The Wannier–Mott type excitons have a large radius that encompasses

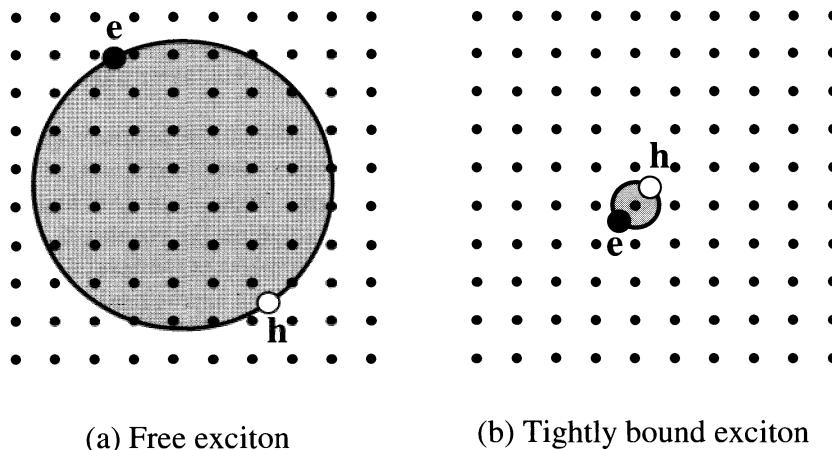


Fig. 4.1 Schematic diagram of: (a) a free exciton, and (b) a tightly bound exciton. The free excitons illustrated in (a) are also called Wannier–Mott excitons, while the tightly bound excitons illustrated in (b) are also called Frenkel excitons.

many atoms, and they are delocalized states that can move freely throughout the crystal; hence the alternative name of ‘free’ excitons. Frenkel excitons, by contrast, have a much smaller radius which is comparable to the size of the unit cell. This makes them localized states which are tightly bound to specific atoms or molecules; hence their alternative name of ‘tightly bound’ excitons. Tightly bound excitons are much less mobile than free excitons, and they have to move through the crystal by hopping from one atom site to another.

Stable excitons will only be formed if the attractive potential is sufficient to protect the exciton against collisions with phonons. Since the maximum energy of a thermally excited phonon at temperature T is $\sim k_B T$, where k_B is Boltzmann’s constant, this condition will be satisfied if the exciton binding energy is greater than $k_B T$. Wannier–Mott excitons have small binding energies due to their large radius, with typical values of around 0.01 eV. Since $k_B T \sim 0.025$ eV at room temperature, the excitons are only observed clearly at cryogenic temperatures in many materials. Frenkel excitons, on the other hand, have larger binding energies of the order 0.1–1 eV, which makes them stable at room temperature.

In the sections that follow, we will first describe the basic properties of free excitons, and then study how they are affected by external electric and magnetic fields. We will then discuss the interactions between excitons, which are the basis for the nonlinear optical properties of excitons discussed in Chapter 11. We close the chapter with a brief discussion of the optical properties of Frenkel excitons.

4.2 Free excitons

4.2.1 Binding energy and radius

In a free exciton, the average separation of the electrons and holes is much greater than the atomic spacing, as shown in Fig. 4.1(a). This is effectively the definition of a Wannier exciton, and it specifies more accurately what is meant by saying that the free exciton is a weakly bound electron–hole pair. Since the electron–hole separation is so large, it is a good approximation

to average over the detailed structure of the atoms in between the electron and hole and consider the particles to be moving in a uniform dielectric material. We can then model the free exciton as a hydrogenic system similar to positronium.

We know from atomic physics that the motion of hydrogenic atoms splits into the centre of mass motion and the relative motion. (See Exercise 4.1.) The centre of mass motion describes the kinetic energy of the atom as a whole, while the relative motion determines the internal structure. The energies of the bound states can be determined by finding the eigenvalues of the Schrödinger equation for the relative motion, or alternatively by using approximation techniques such as the variational method. (See Exercises 4.2–4.4). The main results are, however, well explained by using the Bohr model (cf. Exercise 4.5), and this is the procedure we will adopt here.

In applying the Bohr model to the exciton, we must take account of the fact that the electron and hole are moving through a medium with a high dielectric constant ϵ_r . We must also remember that the reduced mass μ will be given by eqn 3.22, instead of the value of $0.9995m_0$ that applies to the electron–proton system in a hydrogen atom. With these two qualifications, we can then just use the standard results of the Bohr model. The bound states are characterized by the principal quantum number n . The energy of the n th level relative to the ionization limit is given by

$$E(n) = -\frac{\mu}{m_0} \frac{1}{\epsilon_r^2} \frac{R_H}{n^2} = -\frac{R_X}{n^2}, \quad (4.1)$$

where R_H is the Rydberg constant of the hydrogen atom (13.6 eV). The quantity $R_X = (\mu/m_0\epsilon_r^2)R_H$ introduced here is the exciton Rydberg constant. The radius of the electron–hole orbit is given by

$$r_n = \frac{m_0}{\mu} \epsilon_r n^2 a_H = n^2 a_X \quad (4.2)$$

where a_H is the Bohr radius of the hydrogen atom (5.29×10^{-11} m) and $a_X = (m_0\epsilon_r/\mu)a_H$ is the exciton Bohr radius. Equations 4.1 and 4.2 show that the ground state with $n = 1$ has the largest binding energy and smallest radius. The excited states with $n > 1$ are less strongly bound and have a larger radius.

Table 4.1 lists the exciton Rydberg constant and Bohr radius for a number of direct gap III–V and II–VI semiconductors. A general pattern is easily noticed in the data, namely that R_X tends to increase and a_X to decrease as E_g increases. This is explained by the fact that ϵ_r tends to decrease and μ to increase as the band gap increases. From eqns 4.1 and 4.2, we see that this causes an increase in the exciton binding energy and a decrease in the radius. In insulators with band gaps greater than about 5 eV, a_X becomes comparable to the unit cell size, and the Wannier model is no longer valid. At the other extreme, R_X is so small in narrow gap semiconductors such as InSb that it is difficult to observe any free exciton effects at all. Hence, free exciton behaviour is best observed in semiconductors with medium-sized band gaps in the range ~ 1 –3 eV.

Table 4.1 Calculated Rydberg constant and Bohr radius of the free excitons in several direct gap III–V and II–VI compound semiconductors.

E_g : band gap,
 R_X : exciton Rydberg constant from eqn 4.1,
 a_X : exciton Bohr radius from eqn 4.2.
 The bracketed figures for InSb indicate that there has been no experimental confirmation of the values.

Crystal	E_g (eV)	R_X (meV)	a_X (nm)
GaN	3.5	23	3.1
ZnSe	2.8	20	4.5
CdS	2.6	28	2.7
ZnTe	2.4	13	5.5
CdSe	1.8	15	5.4
CdTe	1.6	12	6.7
GaAs	1.5	4.2	13
InP	1.4	4.8	12
GaSb	0.8	2.0	23
InSb	0.2	(0.4)	(100)

Example 4.1

- (i) Calculate the exciton Rydberg and Bohr radius for GaAs, which has $\epsilon_r = 12.8$, $m_e^* = 0.067m_0$ and $m_h^* = 0.2m_0$.
- (ii) GaAs has a cubic crystal structure with a unit cell size of 0.56 nm. Estimate the number of unit cells contained within the orbit of the $n = 1$ exciton. Hence justify the validity of assuming that the medium can be treated as a uniform dielectric in deriving eqns 4.1 and 4.2.
- (iii) Estimate the highest temperature at which it will be possible to observe stable excitons in GaAs.

Solution

- (i) We first need to calculate the reduced electron–hole mass μ , which is given by eqn 3.22. With $m_e^* = 0.067m_0$, and $m_h^* = 0.2m_0$, we find

$$\mu = \left(\frac{1}{0.067m_0} + \frac{1}{0.2m_0} \right)^{-1} = 0.05m_0.$$

We then insert this value of μ and $\epsilon_r = 12.8$ into eqns 4.1 and 4.2 to obtain:

$$R_X = \frac{0.05}{12.8^2} \times 13.6 \text{ eV} = 4.2 \text{ meV},$$

and

$$a_X = \frac{12.8}{0.05} \times 0.0529 \text{ nm} = 13 \text{ nm}.$$

- (ii) We see from eqn 4.2 that the radius of the $n = 1$ exciton is equal to a_X . The volume occupied by this exciton is $\frac{4}{3}\pi a_X^3$ which is equal to $9.2 \times 10^{-24} \text{ m}^3$. The volume of the cubic unit cell is equal to $(0.56 \text{ nm})^3 = 1.8 \times 10^{-28} \text{ m}^3$. Hence the exciton volume can contain 5×10^4 unit cells. Since this is a large number, the approximation of averaging the atomic structure to a uniform dielectric is justified.
- (iii) The $n = 1$ exciton has the largest binding energy with a value of 4.2 meV. This is equal to $k_B T$ at 49 K. Therefore, we would not expect the excitons to be stable above $\sim 50 \text{ K}$.

It is not immediately obvious what is the correct dielectric constant or hole effective mass to use for a III–V semiconductor such as GaAs. This is because ϵ_r varies with frequency (see Section 10.2), and the heavy and light hole bands are degenerate at $k = 0$ (see Fig. 3.5). As a rule of thumb, we use the value of ϵ_r for the photon energy that corresponds to R_X , and a weighted average of the heavy and light hole masses for m_h^* . In this example, R_X comes out to be 4.2 meV, which is in the far-infrared spectral region. We therefore use the static dielectric constant ϵ_{st} for ϵ_r .

4.2.2 Exciton absorption

Free excitons are typically observed in direct gap semiconductors such as GaAs. They are created during direct optical transitions between the valence and conduction bands. As discussed in Section 3.2 this creates an electron–hole pair in which the electron and hole have the same \mathbf{k} vector.

Excitons can only be formed if the electron and hole group velocities v_e and v_h are the same. This is a necessary condition for the electrons and holes to be

Free excitons can also be observed near the fundamental band gap of indirect semiconductors such as silicon and germanium. These indirect excitons are more difficult to conceptualize because the electron and hole have different \mathbf{k} vectors. The condition $v_e = v_h$ is satisfied because the electron at the conduction band minimum still has $v_e = 0$, even though it has a large \mathbf{k} vector. Experimental results give the binding energies of the free excitons in silicon and germanium as 14 meV and 4 meV respectively. These values are slightly higher than the general trends for direct gap semiconductors shown in Table 4.1. This is because of the larger electron mass at the zone edges compared to the Γ -point. It is difficult to observe indirect excitons in absorption because of the reduced probability for indirect transitions. They can, however, be clearly observed in emission experiments, as will be discussed briefly in Section 4.4.

able to move together as a bound pair. The group velocity of an electron in a band is given by (see eqn C.4):

$$v_g = \frac{1}{\hbar} \frac{\partial E}{\partial \mathbf{k}} . \quad (4.3)$$

This implies that the condition $v_e = v_h$ can only be satisfied if the gradients of the conduction and valence bands are the same at the point of the Brillouin zone where the transition occurs. All bands have zero gradient at the zone centre. Hence we can form excitons during a direct transition at $\mathbf{k} = 0$. In a direct gap semiconductor, these transitions correspond to photon energies of E_g (cf. eqn 3.23). Therefore we expect to observe strong excitonic effects in the spectral region close to the fundamental band gap.

The energy of the exciton created in a direct transition at $k = 0$ is equal to the energy required to create the electron–hole pair, namely E_g , less the binding energy due to the Coulomb interaction, which is given by eqn 4.1. Hence the energy of the exciton will be given by:

$$E_n = E_g - \frac{R_X}{n^2} . \quad (4.4)$$

Whenever the photon energy is equal to E_n , excitons can be formed. The probability for the formation of excitons is expected to be high, because it is energetically favourable for the exciton states to be formed compared to free electron–hole pairs. Therefore we expect to observe strong optical absorption lines at energies equal to E_n . These will appear in the optical spectra at energies just below the fundamental band gap. The band edge absorption spectrum expected when excitonic effects are included is illustrated schematically in Fig. 4.2.

Free excitons can only be observed in the absorption spectrum of very pure samples. This is because impurities release free electrons and holes that can screen the Coulomb interaction in the exciton and thereby strongly reduce the binding forces. For this reason, excitonic effects are not usually observed in doped semiconductors or metals, since they contain a very high density of free carriers. Charged impurities also generate electric fields, which tend to ionize the excitons, as discussed in Section 4.3.1.

4.2.3 Experimental data for free excitons in GaAs

Figure 4.3 gives experimental data for the excitonic absorption of undoped GaAs between 21 K and 294 K. As expected, the data show strong absorption lines at photon energies just below the fundamental band gap of GaAs. At 21 K a sharp line is observed just below the direct absorption edge. This corresponds to the $n = 1$ exciton. The line is too broad to permit observation of any of the excited states. As the temperature is increased, the band gap shifts to lower energy and the exciton line weakens. At room temperature where $k_B T \gg R_X$, the exciton line has completely gone.

The spectrum at 185 K shows a weak exciton line at the band edge even though $k_B T$ is almost four times greater than R_X . This indicates that the criterion for exciton stability used in Example 4.1(iii), namely $k_B T < R_X$, is too stringent. The main mechanism that causes dissociation of excitons is

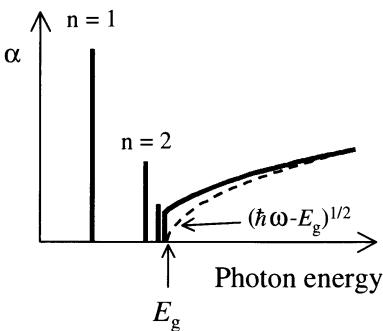


Fig. 4.2 Band edge absorption spectrum for a direct gap semiconductor with excitonic effects included. The dashed line shows the expected absorption when the excitonic effects are ignored.

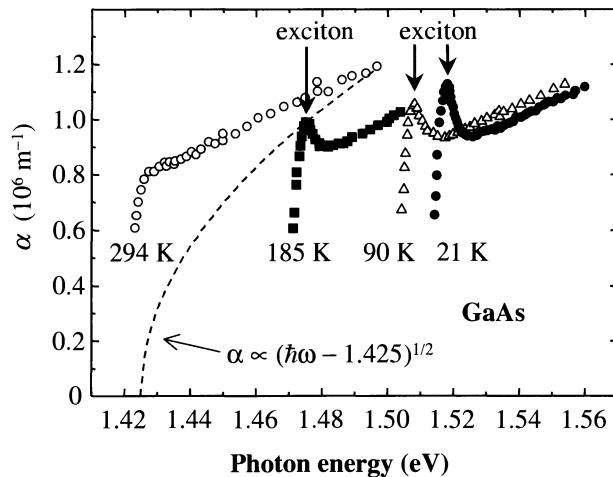


Fig. 4.3 Excitonic absorption of GaAs between 21 K and 294 K. The dashed line is an attempt to fit the absorption edge using eqn 3.25 with a value of E_g equal to 1.425 eV, which is appropriate for GaAs at 294 K. After [1], copyright 1962 American Physical Society, reprinted with permission.

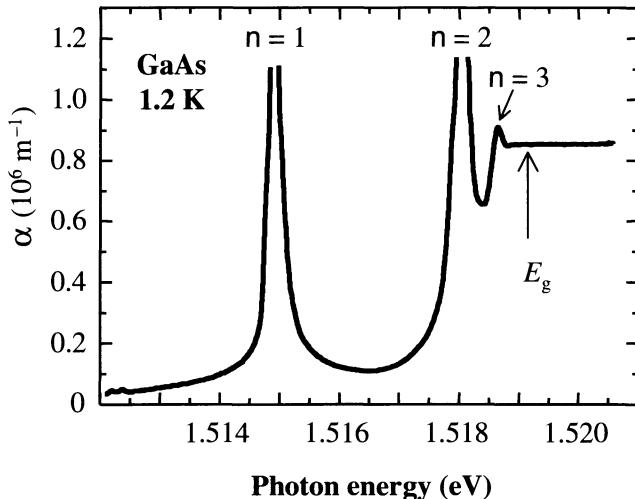


Fig. 4.4 Excitonic absorption of ultra pure GaAs at 1.2 K. After [2], copyright 1985 Excerpta Medica Inc., reprinted with permission.

collisions with longitudinal optic (LO) phonons. As the probability of such collisions increases, the lifetime of the excitons shortens. This leads to a corresponding broadening of the exciton line in the absorption spectrum. In GaAs, the relevant LO phonon has an energy of 35 meV, which has a thermal occupation of 11% at 185 K. (cf. eqn 3.36.) There are therefore still relatively few LO phonons in the crystal at this temperature, and so the collisional broadening does not yet exceed R_X , and the exciton line is just resolved.

The dashed line in Fig. 4.3 shows the frequency dependence of the absorption edge expected if excitonic effects are ignored. This line is obtained from eqn 3.25 with a value of 1.425 eV for E_g , which is appropriate for GaAs at 294 K. We see that the fit to the data is not good. This tells us that the Coulomb interaction between the electron and hole still enhances the absorption rate considerably, even though there are no clear exciton lines observed in the spectrum.

Figure 4.4 shows more recent data for the excitonic absorption of ultra pure GaAs at 1.2 K. The data clearly show the hydrogen-like energy spectrum of

In Chapter 6 we will discuss how the excitonic effects in materials such as GaAs can be enhanced in quantum well structures. This has made it possible to observe very strong free exciton absorption lines in GaAs quantum wells even at room temperature.

the exciton in the vicinity of the band gap. The exciton lines are more clearly resolved in this data set than in Fig. 4.3 because the temperature is lower and the sample purity is superior. As discussed above, the presence of impurities leads to screening of the Coulomb interaction by free carriers, while lower temperatures reduce the thermal broadening of the absorption lines.

Three exciton states can be clearly identified in the absorption spectrum shown in Fig. 4.4. The energies of the $n = 1$, $n = 2$ and $n = 3$ excitons are 1.5149 eV, 1.5180 eV, and 1.5187 eV respectively. These energies fit eqn 4.4 very well with $E_g = 1.5191$ eV and $R_X = 4.2$ meV. This value of E_g agrees well with other measurements, while the experimental figure of 4.2 meV for R_X is in excellent agreement with the value calculated in Example 4.1.

4.3 Free excitons in external fields

Free excitons are bound together by the electrostatic attraction between the negative electron and the positive hole. External electric and magnetic fields perturb the system through the forces exerted on the charged particles. The effects of these perturbations are discussed here, using the excitons in GaAs as an example.

4.3.1 Electric fields

When a DC electric field \mathcal{E} is applied to an exciton, the oppositely charged electrons and holes are pushed away from each other. It is shown in Exercise 4.10 that the order of magnitude of the electric field between the electron and hole in the ground state exciton is equal to $2R_X/eax$. If \mathcal{E} exceeds this value, the exciton will break apart. This effect is known as **field ionization**.

Electric fields are applied to excitons by incorporating the semiconductor as the i-region in a p-i-n diode structure, as discussed in Appendix D. The field strength across the i-region when a bias voltage V_0 is applied is given by eqn D.3 as:

$$\mathcal{E} = \frac{|V_{bi} - V_0|}{l_i}, \quad (4.5)$$

where V_{bi} is the built-in voltage of the diode and l_i is the intrinsic region thickness. The sign convention is such that positive V_0 corresponds to forward bias.

In a typical GaAs p-i-n diode, the i-region thickness is about $1\ \mu\text{m}$, and V_{bi} is about 1.5 V. Equation 4.5 then tells us that \mathcal{E} is $1.5 \times 10^6\ \text{V m}^{-1}$ at zero bias. At the same time we see from Table 4.1 that in GaAs $2R_X/eax$ is of order $6 \times 10^5\ \text{V m}^{-1}$, which is substantially less than the field strength at $V_0 = 0$. We would therefore expect the excitons to be ionized even before we apply bias to the diode.

Figure 4.5 shows experimental data for the field ionization of free excitons in a GaAs p-i-n diode with $l_i = 1.0\ \mu\text{m}$ at 5 K. In this experiment, the diode is illuminated with light, and the photocurrent generated at a given voltage and wavelength is recorded. The solid line is the photocurrent recorded in ‘flat band conditions’ ($V_0 = +1.44\ \text{V}$, $\mathcal{E} \approx 0$), while the dashed line is for $V_0 = +1.00\ \text{V}$, where $\mathcal{E} \approx 5 \times 10^5\ \text{V m}^{-1}$. In the flat band case we observe a well-resolved

The wavelength dependence of the photocurrent follows the absorption spectrum. We can see this from eqn 3.37, which shows that the photocurrent is proportional to $(1 - e^{-\alpha l})$. If αl is small, the photocurrent is directly proportional to α . If αl is not small, the photocurrent will still show peaks at wavelengths where α is a maximum.

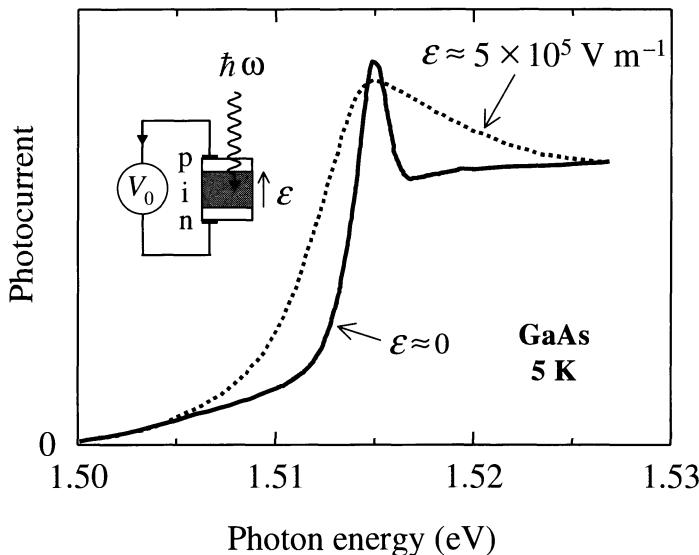


Fig. 4.5 Field ionization of the free excitons in GaAs at 5 K. The data was taken on a GaAs p-i-n diode with an i-region thickness of $1.0\ \mu\text{m}$. The solid line corresponds to ‘flat band’ conditions (forward bias = $+1.44\text{ V}$, $\epsilon \approx 0$), while the dashed line is for a forward bias of $+1.00\text{ V}$, where $\epsilon \approx 5 \times 10^5 \text{ V m}^{-1}$. No exciton lines are resolved at zero bias. (After G. von Plessen and A.M. Tomlinson, Personal communication)

exciton line at 1.515 eV. However, once we reduce the bias by only a very small amount, we rapidly approach the ionization field, and the exciton broadens significantly. At zero bias (not shown), we are well above the ionization field, and no exciton lines are resolved in the spectrum.

From the discussion above, it is clear that excitonic effects do not play a large part in the physics of bulk semiconductor diodes. The excitons will only be observed over a small range of forward bias voltages just less than V_{bi} . Therefore, the physics of bulk semiconductors in electric fields is dominated more by the effect of the field on the band states, namely the Franz–Keldysh effect discussed in Section 3.3.5. As we will see in Chapter 6, this is not the case for the enhanced free excitons in GaAs quantum wells. These show very interesting electric field effects even at room temperature.

4.3.2 Magnetic fields

The application of a magnetic field perturbs the free excitons by applying magnetic forces to the electron and hole. The strength of the perturbation is set by the exciton cyclotron energy $\hbar\omega_c$, which is given by

$$\hbar\omega_c = \hbar \frac{eB}{\mu}, \quad (4.6)$$

where B is the magnetic flux density. This is similar to the formula for individual electrons given in eqn 3.27, except that the reduced electron–hole effective mass μ appears instead of the individual electron mass.

The behaviour can be divided into the weak and strong field limits, with the transition point set by the ratio of the exciton Rydberg energy to the cyclotron energy. If $R_X \gg \hbar\omega_c$, we are in the weak field regime, whereas $R_X \ll \hbar\omega_c$ corresponds to the strong field regime. In GaAs, the transition between the two limits occurs around 2 T for the $n = 1$ exciton: see Exercise 4.12.

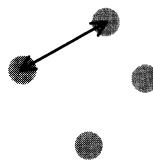
In the weak field limit we treat the magnetic field as a perturbation on the excitons. The ground state of a hydrogen atom has no net magnetic moment

because it is spherically symmetric. Thus the interaction between the $n = 1$ exciton and the B -field will be described by diamagnetic effects. The diamagnetic energy shift is given by (see Exercise 4.13):

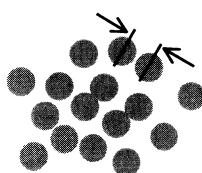
$$\delta E = +\frac{e^2}{12\mu} r_n^2 B^2. \quad (4.7)$$

The shift is positive because Lenz's law tells us that the field induces a magnetic moment that opposes the applied field. This induced dipole then interacts with the field to give an energy shift proportional to $+B^2$.

In the strong field limit, the interaction of the electrons and holes with the field is stronger than their mutual Coulomb interaction. We therefore consider the Landau energy of the individual electrons and holes first, as in Section 3.3.6. We then add on the Coulomb interaction as a small perturbation. The details of this analysis are beyond the scope of this book. The end result is that the excitonic effects cause a small shift in the energies of the optical transitions between the Landau levels.



(a) Low density
Separation » diameter



(b) High density
Separation ≈ diameter

Fig. 4.6 Distribution of the free excitons within a crystal. (a) Low densities: the excitons are randomly distributed throughout the excitation volume and the interexciton separation is large. (b) High densities: the wave functions overlap when the exciton-exciton separation becomes comparable to the exciton diameter.

4.4 Free excitons at high densities

Wannier excitons behave as if they are hydrogen-like atoms moving freely through the crystal. The atoms in a gas of hydrogen are agitated by thermal motion and interact with each other whenever they get close together. The simplest type of interaction is the tendency to form the H_2 molecule, but other phenomena such as Bose-Einstein condensation are also possible. Excitons show a similar variety of phenomena such as the tendency to form molecules or condense to a liquid phase. The type of behaviour observed in any one material depends very much on the conditions that apply and the details of the interactions between the excitons.

We first consider an experiment in which we take a powerful laser and tune it to one of the exciton absorption lines. The laser creates excitons in the sample, with a density that is proportional to the laser power. At low powers, the density of the excitons is small, and the separation between the excitons is large, as sketched in Fig. 4.6(a). The exciton-exciton interactions are negligible in these conditions. As the power is increased, the density of excitons increases. Eventually, the density will be high enough that the exciton wave functions begin to overlap, as sketched in Fig. 4.6(b). At this point, we expect that the exciton-exciton interactions will become very significant.

We can see from Fig. 4.6(b) that exciton wave function overlap occurs when the exciton-exciton distance is equal to the exciton diameter. The density at which this occurs is called the **Mott density** N_{Mott} . It is given approximately by the inverse volume of the exciton:

$$N_{\text{Mott}} \approx \frac{1}{\frac{4}{3}\pi r_n^3}. \quad (4.8)$$

From Table 4.1 and eqn 4.2, we find that the Mott density for the $n = 1$ excitons in GaAs is $1.1 \times 10^{23} \text{ m}^{-3}$. This density is easily achievable with a focussed laser beam.

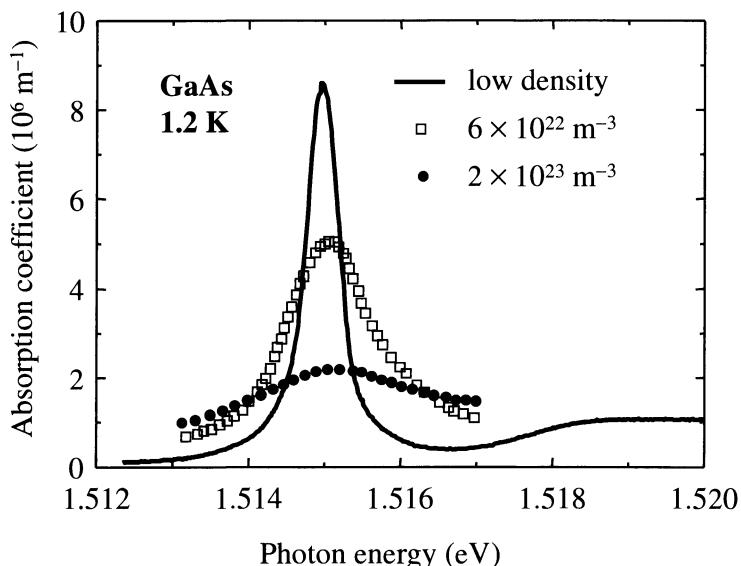


Fig. 4.7 Absorption coefficient of GaAs in the spectral region close to the band edge at 1.2 K at three different excitation powers. The carrier densities generated for the two higher power levels are indicated. After [2], copyright 1985 Excerpta Medica Inc., reprinted with permission.

When the exciton density approaches N_{Mott} , a number of effects can occur. In GaAs the collisions between the excitons cause the exciton gas to dissociate into an electron–hole plasma. This causes exciton broadening with a reduction in the absorption strength. Figure 4.7 shows the absorption coefficient at the $n = 1$ exciton in GaAs at 1.2 K at three different excitation powers. The weakening and broadening of the exciton line as the carrier density is increased is clearly observed in the data. The density at which these effects occur agrees well with the value of $1.1 \times 10^{23} \text{ m}^{-3}$ given by eqn 4.8. The saturation of the exciton absorption with increasing power is an example of a **nonlinear** optical effect: the absorption coefficient depends on the intensity of the light. We will return to discuss applications of these nonlinear effects in Section 11.4.3 of Chapter 11.

Another effect that can be observed at high exciton densities in other materials is the formation of exciton molecules called **biexcitons**. This is the equivalent process to the formation of an H₂ molecule from two isolated hydrogen atoms. Biexcitons have been observed in a number of compound semiconductors, including CdS, ZnSe, ZnO and especially copper chloride. CuCl has a band gap at 3.40 eV, and the ground state exciton is observed at 3.20 eV, implying that $R_X = 0.2$ eV. At high densities, a new feature is observed in the absorption spectrum at 3.18 eV. This is attributed to biexciton formation. The energy difference between the two features tells us that the binding energy of the biexciton is 0.02 eV. Attempts to observe biexcitons in materials like GaAs have been hindered by the nonlinear saturation effects described above.

In silicon and germanium at high densities, yet another effect occurs. At low densities the excitons may be considered to be in a gaseous phase. As the density increases, the excitons condense to form a liquid. The liquid phase manifests itself in the formation of **electron–hole droplets**, which are observed in the recombination radiation of the excitons at high densities. The droplet

appears as a broad feature at lower energy than the free excitons.

The final high density effect that we consider here is **Bose–Einstein condensation**. At high temperatures, the particles in a non-interacting boson gas are distributed between the possible energy levels of the system according to Bose–Einstein statistics. As the temperature is lowered, the distribution undergoes a radical change, and a macroscopic number of particles accumulate in the ground state. The critical temperature T_c at which this occurs is given by:

$$N = 2.612 \left(\frac{mk_B T_c}{2\pi\hbar^2} \right)^{\frac{3}{2}}, \quad (4.9)$$

where N is the number of particles per unit volume and m is the particle mass. At T_c the thermal de Broglie wavelength is comparable to the interparticle separation, and quantum effects are to be expected. (See Exercise 4.16.)

Bose–Einstein condensation has been observed in many boson systems. One of the best studied examples is liquid helium. In this case, N is fixed, and eqn 4.9 predicts a phase transition as the liquid is cooled through T_c at 2.2 K. However, the physics of Bose–Einstein condensation is complicated in liquid helium by the strong interactions between the atoms. To achieve pure Bose–Einstein condensation behaviour, we require that the interactions between the bosons are negligible. This suggests that we need highly dilute gaseous systems such that the interparticle separation is very large. However, from eqn 4.9 we see that the transition temperature for such a dilute system would be very low. It has been an outstanding recent achievement of atomic physics to succeed in observing Bose–Einstein condensation in extremely dilute gases of atoms at temperatures below 1 μ K.

Excitons consist of two spin $\frac{1}{2}$ particles, and so their total spin is either 0 or 1. This means that they are bosons. There have been many attempts to study condensation phenomena, but practically all of the claimed observations have been disputed. It is actually very difficult to prove definitively that condensation has occurred. Two of the most promising candidate systems that have been studied to date are the spin-0 excitons in copper oxide (Cu_2O) and the biexcitons in CuCl.

4.5 Frenkel excitons

The free exciton model that leads to eqns 4.1 and 4.2 breaks down when the predicted radius becomes comparable to the interatomic spacing. This occurs in large band gap materials with small dielectric constants and large effective masses. In these materials we observe Frenkel excitons rather than Wannier excitons.

Frenkel excitons are localized on the atom site at which they are created, as shown in Fig. 4.1(b). The excitons may therefore be considered as excited states of the individual atoms or molecules on which they are localized. They have very small radii and correspondingly large binding energies, with typical values ranging from about 0.1 eV to several eV. This means that Frenkel excitons are usually stable at room temperature. The excitons can propagate through the crystal by hopping from atom site to site in the same way that spin excitations propagate through crystals as magnon waves.

The theoretical treatment of Frenkel excitons requires techniques more akin to atomic or molecular physics than solid state physics. There is no simple model similar to the one that led to eqns 4.1 and 4.2 for free excitons. The calculation of the exciton energies usually follows a tight binding approach, in order to emphasize the correspondence to the atomic or molecular states from which the excitons are derived. The calculation is further complicated by the fact that the coupling between the excitons and the crystal lattice is usually very strong. This leads to ‘self-trapping’ effects, in which the exciton produces a local distortion of the lattice, which then causes further localization of the exciton wave functions.

Frenkel excitons have been observed in many inorganic and organic materials. The properties of some of the more widely studied crystals are described briefly below.

4.5.1 Rare gas crystals

The rare gases from group VIII of the periodic table, namely neon, argon, krypton and xenon, crystallize at cryogenic temperatures. The band gap ranges from 21.6 eV in neon to 9.3 eV in xenon. Neon in fact has the largest band gap of any crystal known in nature. The excitonic absorption of these materials has been thoroughly studied, and the results are summarized in Table 4.2. The exciton transitions all occur in the vacuum ultraviolet spectral range, and the binding energies are very large.

It has been found experimentally that there is a close correspondence between the $n = 1$ exciton energies in the crystals and the optical transitions of the isolated atoms. For example, the energy of the $n = 1$ exciton in xenon crystals coincides almost exactly with the lowest energy absorption line of xenon atoms in the gaseous phase, namely the $5p^6 \rightarrow 5p^56s$ transition. This underlines the point made earlier that the localized nature of the Frenkel excitons makes them equivalent to excited states of the individual atoms or molecules. This correspondence gets weaker for the excitons with larger values of n . As the radius increases with n , the excitons become more and more delocalized, and it eventually becomes valid to use the Wannier model.

4.5.2 Alkali halides

Frenkel excitons are readily observable in the optical spectra of alkali halide crystals. These have large direct band gaps in the ultraviolet spectral region ranging from 5.9 eV in NaI to 13.7 eV in LiF. LiF has the widest band gap of any practical optical material: only argon and neon crystals have larger band gaps, but neither of these are solids at room temperature.

Table 4.3 lists the band gap of selected alkali halide crystals, together with the energy and binding energy of the $n = 1$ exciton. The data show that E_g tends to increase with decreasing anion and cation size. The exciton binding energy follows a similar general trend. Detailed spectroscopy has established that the excitons are localized at the negative (halogen) ions.

Figure 4.8 shows the absorption spectrum of two representative alkali halide crystals at room temperature, namely NaCl and LiF. Both spectra show a strong excitonic absorption line below the band gap. The binding energies are 0.8 eV

The self-trapping of electrons or holes is caused by the electron–phonon coupling. These polaronic effects will be discussed in Section 10.4.

Table 4.2 Properties of Frenkel excitons in rare gas crystals

T_m : melting temperature in K,
 E_g : band gap,
 E_1 : energy of the $n = 1$ exciton,
 E_b : binding energy of the $n = 1$ exciton.
All energies are given in eV. After [3].

Crystal	T_m	E_g	E_1	E_b
Ne	25	21.6	17.5	4.1
Ar	84	14.2	12.1	2.1
Kr	116	11.7	10.2	1.5
Xe	161	9.3	8.3	1.0

Table 4.3 Properties of Frenkel excitons in selected alkali halide crystals

E_g : band gap,
 E_1 : energy of the $n = 1$ exciton line,
 E_b binding energy of the $n = 1$ exciton.
All energies are given in eV. After [3].

Crystal	E_g	E_1	E_b
KI	6.3	5.9	0.4
KBr	7.4	6.7	0.7
KCl	8.7	7.8	0.9
KF	10.8	9.9	0.9
NaI	5.9	5.6	0.3
NaBr	7.1	6.7	0.4
NaCl	8.8	7.9	0.9
NaF	11.5	10.7	0.8
CsF	9.8	9.3	0.5
RbF	10.3	9.5	0.8
LiF	13.7	12.8	1.9

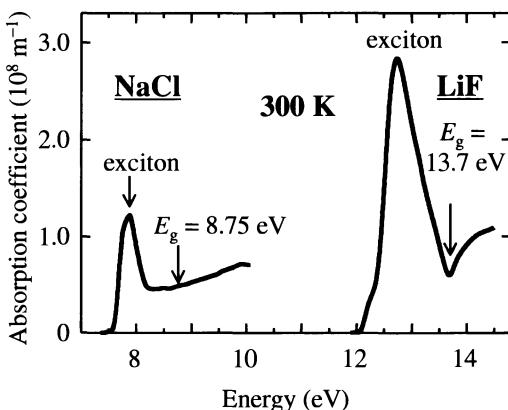


Fig. 4.8 Absorption spectra of NaCl and LiF at room temperature. After [4].

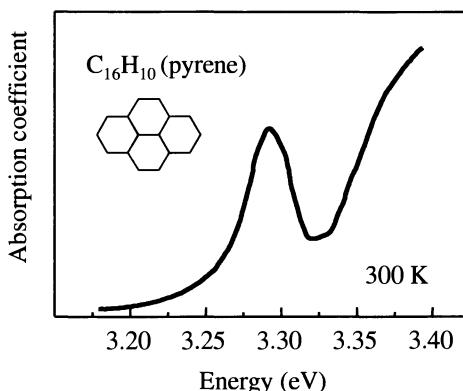


Fig. 4.9 Absorption spectrum of pyrene ($\text{C}_{16}\text{H}_{10}$) single crystals at room temperature. After [5], reprinted with permission.

and 1.9 eV respectively. These values are well above $k_{\text{B}}T$ at room temperature, which explains why the excitons are observed so strongly. The fine structure of the excitons due to the excited states can be observed by cooling the crystals. Note that the absorption coefficient at the exciton lines is extremely large, with values over 10^8 m^{-1} in both materials.

4.5.3 Molecular crystals

Frenkel excitons can be observed in many molecular crystals and organic thin film structures. In most cases, there is a very strong correspondence between the optical transitions of the isolated molecules and the excitons observed in the solid state. This is a consequence of the fact that the molecular crystals are held together by relatively weak van der Waals forces, so that the molecular levels are only weakly perturbed when condensing to the solid state.

Figure 4.9 shows the fundamental absorption edge of pyrene crystals at room temperature. The pyrene molecule has a composition of $\text{C}_{16}\text{H}_{10}$ and is an example of an aromatic hydrocarbon, that is, a carbon-hydrogen compound based on benzene rings. The 4-ring structure of pyrene is given in the inset. The absorption spectrum shows a clear excitonic peak at 3.29 eV. Other aromatic hydrocarbons such as anthracene ($\text{C}_{14}\text{H}_{10}$) also show very strong excitonic effects, but the optical spectra are more complicated because of the strong cou-

pling to the vibrational modes of the molecule. These effects will be discussed in more detail in Section 8.4 in Chapter 8. The pyrene spectrum is relatively simple because the 4-ring structure makes the molecule very rigid and reduces the effects of the vibrational coupling.

Frenkel excitons are also very important in conjugated polymers, such as polydiacetylene (PDA). Single crystals of PDA can be grown, but the optical properties are often studied by using amorphous films coated onto glass substrates. The strong excitonic effects in conjugated polymers have acquired considerable technological significance in recent years, following the development of organic light emitting diodes for use in display technology. The optical properties of organic semiconductors such as PDA will be discussed in more detail in Sections 8.5 and 8.6 of Chapter 8.

Pyrene, anthracene and the other aromatic hydrocarbons are examples of *conjugated* molecules. The π electrons of the benzene rings form large delocalized molecular orbitals with optical transitions in the blue/UV spectral range. Polydiacetylene is another example of a conjugated molecule. In Section 1.4.4 we mentioned that the most interesting molecular materials, from the point of view of their optical properties, are those with conjugated bonds. This point will be developed further in Chapter 8.

Chapter summary

- Excitons are electron–hole pairs bound together in stable orbits by the mutual Coulomb attraction between them.
- There are two types of excitons. Wannier (free) excitons have a large radius and move freely throughout the crystal. Frenkel (tightly bound) excitons are localized on individual atoms sites.
- The properties of free excitons can be calculated by treating them as hydrogen-like atoms. The binding energies and radii are given by eqns 4.1 and 4.2 respectively.
- Free excitons are observed in semiconductors at photon energies just below E_g . They have fairly small binding energies, and are observed most clearly at low temperatures. They are easily ionized by electric fields.
- Free excitons can interact with each other, and they show a rich variety of phenomena at high densities due to the exciton–exciton interactions.
- Frenkel excitons have very small radii and large binding energies. They are easily observed at room temperature in insulator crystals and molecular materials. There is a strong correspondence between the excitons observed in the solid state and the excited states of the individual atoms or molecules of which the solid is composed.

Further reading

Supplementary reading on excitons may be found in most of the standard solid state texts such as Burns (1985) or Kittel (1996). More detailed information on free excitons in semiconductors may be found in Klingshirn (1995), Pankove (1971), Seeger (1997), or Yu and Cardona (1996).

Dexter and Knox (1965) is a classic text on excitons, while Rashba and Sturge (1982) is a more recent authoritative reference work. Reynolds and Collins (1981) give a good overview of excitonic physics, while Song and Williams (1993) give a thorough discussion of the properties of Frenkel excitons.

An overview of high density exciton effects may be found in Klingshirn (1995). The general phenomenon of Bose–Einstein condensation is discussed in most texts on statistical mechanics, for example, Mandl (1988). Griffin *et al.* (1995) give a review of measurements of condensation in a wide variety of systems. Reviews of Bose condensation effects in atomic systems may be found in Wieman (1997) or Ketterle (1999).

References

- [1] Sturge, M.D. (1962). *Phys. Rev.*, **127**, 768.
- [2] Fehrenbach, G.W., Schäfer, W. and Ulbrich, R.G. (1985). *J. Luminescence*, **30**, 154.
- [3] Song, K.S. and Williams, R.T. (1993). *Self-trapped excitons*. Springer-Verlag, Berlin.
- [4] Palik, E.D. (1985). *Handbook of the optical constants of solids*. Academic Press, San Diego.
- [5] Matsui, A. and Nishimura, H. (1980). *J. Phys. Soc. Jap.*, **49**, 657.

Exercises

- (4.1) Write down the Schrödinger equation for the hydrogen atom. By defining the centre of mass and relative coordinates for the electron and proton, show that the Hamiltonian of the system can be split into two parts, one describing the free motion of the whole atom and the other describing the internal energy of the atom due to the Coulomb energy and orbital motion.
- (4.2) The Hamiltonian for the relative motion of an electron–hole pair in a semiconductor is given by:

$$\hat{H} = -\frac{\hbar^2}{2\mu} \nabla^2 - \frac{e^2}{4\pi\epsilon_0\epsilon_r r}.$$

- (i) Explain the origin of the two terms in the Hamiltonian.
(ii) Show that the wave function $\Psi(r, \theta, \phi) = C \exp(-r/a_0)$ is a solution of the Schrödinger equation

$$\hat{H}\Psi = E\Psi,$$

and find the values of E and a_0 . Find also the value of the normalization constant C .

- (4.3) Find the radius at which the radial probability density of the hydrogenic wave function given in the previous

question reaches its maximum value. Compare this to the expectation value $\langle r \rangle$ defined by

$$\langle r \rangle = \int_{r=0}^{\infty} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \Psi^* r \Psi r^2 \sin\theta dr d\theta d\phi.$$

- (4.4)* In the variational method, we make an enlightened guess of the wave function, and then vary its parameters to minimize the expectation value of the energy.¹ The variational principle says that the wave function that gives the minimum energy is the best approximation to the true wave function, and that the corresponding expectation value of the energy is the best approximation of the true energy.

- (i) Explain why the following function is a sensible guess for the wave function of the ground state of the exciton system:

$$\Psi(r, \theta, \phi) = \left(\frac{1}{\xi}\right)^{\frac{3}{2}} \frac{1}{\sqrt{\pi}} \exp\left(-\frac{r}{\xi}\right).$$

¹ This exercise illustrates the use of the variational method to obtain approximate solutions for the wave function and energy of the ground state. These can of course be found by brute force solution of the Schrödinger equation, but the variational method is more intuitive, and can be easily adapted to more complex problems where no analytic solutions are possible.

- (ii) Calculate the expectation value for the energy of an exciton with wave function Ψ :

$$\langle E \rangle = \int \int \int \Psi^* \hat{H} \Psi r^2 \sin \theta dr d\theta d\phi,$$

where \hat{H} is the Hamiltonian given in Exercise 4.2.

- (iii) Find the value of ξ that minimizes $\langle E \rangle$, and compute $\langle E \rangle$ for this value of ξ .
- (iv) Compare the minimal values of E and ξ obtained in part (iii) to those obtained in Exercise 4.2, and comment on your answer.
- (4.5) (i) State the assumptions of the Bohr model of the hydrogen atom.
- (ii) Use the Bohr model to show that the energy and radius of a hydrogenic atom with reduced mass μ in a medium with a relative dielectric constant ϵ_r are given by eqns 4.1 and 4.2 respectively.
- (iii) How does $E(n)$ compare to the exact solution of the Schrödinger equation (cf. Exercise 4.2)?
- (iv) How does r_n relate to the conclusions of Exercise 4.3?
- (4.6) Calculate the binding energy and radius of the $n = 1$ and $n = 2$ free excitons in zinc sulphide (ZnS) which has $m_e^* = 0.28m_0$, $m_h^* = 0.5m_0$ and $\epsilon_r = 7.9$. Would you expect these excitons to be stable at room temperature?
- (4.7) Calculate the difference in the wavelengths of the $n = 1$ and $n = 2$ excitons in InP, which has $E_g = 1.424$ eV, $m_e^* = 0.077m_0$, $m_h^* = 0.2m_0$ and $\epsilon_r = 12.4$.
- (4.8) At 4 K the $n = 1$ exciton in GaAs has a peak absorption coefficient of 3×10^6 m⁻¹ at 1.5149 eV, with a full width at half maximum equal to 0.6 meV. By applying the bound oscillator model discussed in Chapter 2 to the exciton, determine the magnitude and energy of the local maximum in the refractive index just below the exciton absorption line. The non-resonant refractive index of GaAs at energies below the band gap is 3.5.
- (4.9) Excitons can absorb photons by making transitions to excited states in exactly the same way that hydrogen atoms do. Calculate the wavelength of the photon required to promote an exciton in GaAs ($\mu = 0.05m_0$, $\epsilon_r = 12.8$) from the $n = 1$ to the $n = 2$ state.
- (4.10) Use the Bohr model to show that the magnitude of the electric field between the electron and hole in the ground state of a free exciton is equal to $2R_X/e\alpha_X$.
- (4.11) Direct excitons may be formed in germanium at low temperatures using photon energies close to the direct band gap at 0.898 eV. Calculate the binding energy and

radius of the ground state exciton, taking $m_e^* = 0.038m_0$, $m_h^* = 0.1m_0$, and $\epsilon_r = 16$. Calculate the voltage at which the field across the excitons will be equal to the ionization field in a germanium p-i-n diode, which has $V_{bi} = 0.74$ eV and an i-region thickness of 2 μm .

- (4.12) Show that the magnetic field strength at which the exciton cyclotron energy is equal to the exciton Rydberg energy is given by:

$$B = \frac{\mu^2}{\epsilon_r^2 m_0 \hbar} \left(\frac{R_H}{e} \right).$$

Evaluate this field strength for GaAs with $\mu = 0.05m_0$ and $\epsilon_r = 12.8$.

(4.13)* Verify by using eqn A.14 that a vector potential of the form $\mathbf{A} = \frac{1}{2}(-yB_0, xB_0, 0)$ produces a constant magnetic flux density of magnitude B_0 in the z direction. By following an analysis similar to the one that leads to eqn B.17 in Appendix B, show that the diamagnetic energy shift of an electron in an atom with a wave function ψ is given by:

$$\Delta E = \frac{e^2 B_0^2}{8m_0} \langle \psi | x^2 + y^2 | \psi \rangle.$$

Hence derive eqn 4.7.

- (4.14) Calculate the diamagnetic energy shift of the $n = 1$ exciton of GaAs in a magnetic field of 1.0 T. What is the shift in the wavelength of the exciton caused by applying the field? Take $\mu = 0.05m_0$, and the energy of the exciton at $B = 0$ to be 1.515 eV.
- (4.15) Estimate the Mott densities for the $n = 1$ and $n = 2$ excitons in gallium nitride (GaN), which has $m_e^* = 0.2m_0$, $m_h^* = 1.2m_0$ and $\epsilon_r = 10$.
- (4.16) Show that the de Broglie wavelength λ_{deB} of a particle of mass m with thermal kinetic energy $\frac{3}{2}k_B T$ is given by:

$$\lambda_{deB} = \frac{h}{(3mk_B T)^{\frac{1}{2}}}.$$

Calculate the ratio of the interparticle separation to λ_{deB} at the Bose-Einstein condensation temperature.

- (4.17) Calculate the Bose-Einstein condensation temperature for the spin-0 excitons in cuprous oxide when the density is 10^{24} m⁻³. The electron and hole effective masses are $1.0m_0$ and $0.7m_0$ respectively.
- (4.18) The values of μ and ϵ_r for sodium iodide (NaI) are $0.18m_0$ and 2.9 respectively. The unit cell size is 0.65 nm. Would you expect the Wannier model to be valid for the $n = 1$ exciton? What about the $n = 2$ exciton?

* Exercises marked with an asterisk are more challenging.

5

Luminescence

5.1	Light emission in solids	92
5.2	Interband luminescence	95
5.3	Photoluminescence	98
5.4	Electroluminescence	103

In Chapter 3 we considered how light can be absorbed in solids by exciting interband transitions. Then in Chapter 4 we considered how the absorption spectrum is modified by the interactions that lead to the formation of excitons. We now consider the reverse process in which electrons in an excited state drop to lower levels by emitting photons. This is the solid state equivalent to light emission in atoms by spontaneous emission, which is reviewed in Appendix B.

The physical mechanisms responsible for light emission in solids vary considerably from material to material. In this chapter we will start by giving a few general principles that apply to all materials, and then focus on the emission of light by interband transitions in bulk semiconductors. This will provide the framework for discussing the light emission processes in semiconductor quantum wells in Chapter 6, and will also serve as a general introduction to the light emission processes in other types of materials.

5.1 Light emission in solids

Atoms emit light by spontaneous emission when electrons in excited states drop down to a lower level by radiative transitions. In solids the radiative emission process is called **luminescence**. Luminescence can occur by a number of mechanisms, but in this book we will mainly be considering just two:

- **Photoluminescence**: the re-emission of light after absorbing a photon of higher energy.
- **Electroluminescence**: the emission of light caused by running an electrical current through the material.

The physical processes involved in both photoluminescence and electroluminescence are more complicated than those in absorption. This is because the generation of light by luminescence is intimately tied up with the energy relaxation mechanisms in the solid. Furthermore, the shape of the emission spectrum is affected by the thermal distributions of the electrons and holes within their bands. Therefore, we have to consider the emission rates and the thermal spread of the carriers before we can gain a good understanding of the emission efficiency and the luminescence spectrum.

Figure 5.1 gives an overview of the main processes that occur when light is emitted from a solid. The photon is emitted when an electron in an excited state drops down into an empty state in the ground state band. For this to be possible, we must first inject electrons, which then relax to the state from where the emission occurs. This could be the bottom of the conduction band, but it might also be a discrete level. The photon cannot be emitted unless the lower

The diagram in Fig. 5.1 applies to emission between bands, but the basic idea that the carriers relax to the lowest excited state level before emitting the photon is usually applicable even if the levels are discrete.

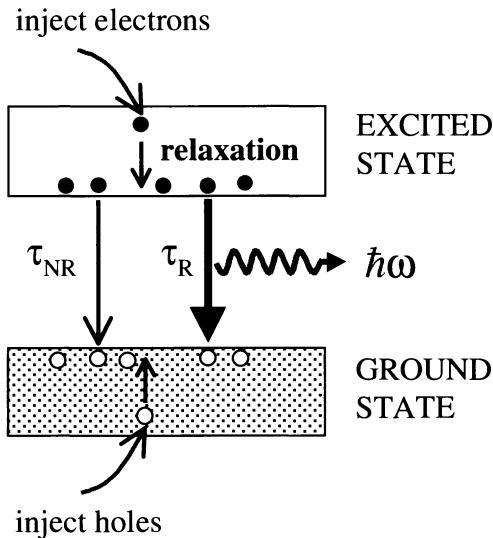


Fig. 5.1 General scheme of luminescence in a solid. Electrons are injected into the excited state band and relax to the lowest available level before dropping down to empty levels in the ground state band by emitting a photon. These empty levels are generated by the injection of holes. The radiative recombination rate is determined by the radiative lifetime τ_R . Radiative emission has to compete with non-radiative recombination, which has a time constant τ_{NR} . The luminescent efficiency is determined by the ratio of τ_R to τ_{NR} , and is given by eqn 5.5.

level for the transition is empty, because the Pauli principle does not permit us to put two electrons into the same level. The empty lower level is produced by injecting holes into the ground state band in an entirely analogous way to the injection of the electrons into the excited state.

The spontaneous emission rate for radiative transitions between two levels is determined by the Einstein A coefficient. (See Appendix B.) If the upper level has a population N at time t , the radiative emission rate is given by:

$$\left(\frac{dN}{dt} \right)_{\text{radiative}} = -AN . \quad (5.1)$$

This shows that the number of photons emitted in a given time is proportional to both the A coefficient of the transition and also to the population of the upper level. The rate equation can be solved to give:

$$N(t) = N(0) \exp(-At) = N(0) \exp(-t/\tau_R) , \quad (5.2)$$

where $\tau_R = A^{-1}$ is the **radiative lifetime** of the transition.

Equation B.11 in Appendix B tells us that the Einstein A coefficient is directly proportional to the B coefficient, which determines the absorption probability. This means that transitions which have large absorption coefficients also have high emission probabilities and short radiative lifetimes. However, the fact that the absorption and emission probabilities are closely related to each other does not imply that the absorption and emission spectra are the same. This is because of the population factor that enters eqn 5.1. A transition might have a high emission probability, but no light will be emitted unless the upper level is populated.

We can summarize these points by writing the luminescent intensity at frequency ν as:

$$I(h\nu) \propto |M|^2 g(h\nu) \times \text{level occupancy factors}, \quad (5.3)$$

where the occupancy factors give the probabilities that the relevant upper level is occupied and the lower level is empty. The other two terms are the matrix

element and the density of states for the transition, which determine the quantum mechanical transition probability by Fermi's golden rule. (See Section B.2 in Appendix B.)

The occupancy factors that enter eqn 5.3 will be discussed in detail in Section 5.3. The main point is that the electrons relax very rapidly to the lowest levels within the excited state band, and then form a thermal distribution that can be calculated by statistical mechanics. In normal circumstances the electrons will relax to within $\sim k_B T$ of the bottom of the excited state band. The holes follow a similar series of relaxation processes. The light is emitted between the electron and hole states that are thermally occupied, and will therefore only be emitted within a narrow energy range from the lowest levels in the excited state band. This contrasts with the absorption spectrum, where photons can be absorbed to any state within the excited state band, no matter how far it is above the bottom of the band.

Radiative emission is not the only mechanism by which the electrons in an excited state can drop down to the ground state. The alternative pathway between the excited state and ground state bands in Fig. 5.1 indicates the possibility of **non-radiative** relaxation. The electron might, for example, lose its excitation energy as heat by emitting phonons, or it may transfer the energy to impurities or defects called 'traps'. If these non-radiative relaxation processes occur on a faster time scale than the radiative transitions, very little light will be emitted.

The luminescent efficiency η_R can be calculated by writing down the rate equation for the population of the excited state when non-radiative processes are possible:

$$\left(\frac{dN}{dt} \right)_{\text{total}} = -\frac{N}{\tau_R} - \frac{N}{\tau_{NR}} = -N \left(\frac{1}{\tau_R} + \frac{1}{\tau_{NR}} \right). \quad (5.4)$$

The two terms on the right hand side of eqn 5.4 represent the radiative and non-radiative rates respectively. τ_{NR} is the non-radiative lifetime. η_R is given by the ratio of the radiative emission rate to the total de-excitation rate. This is obtained by dividing eqn 5.1 by eqn 5.4 to obtain

$$\eta_R = \frac{AN}{N(1/\tau_R + 1/\tau_{NR})} = \frac{1}{1 + \tau_R/\tau_{NR}}, \quad (5.5)$$

where we have used the fact that $A = \tau_R^{-1}$. If $\tau_R \ll \tau_{NR}$ then η_R approaches unity and the maximum possible amount of light is emitted. On the other hand, if $\tau_R \gg \tau_{NR}$ then η_R is very small and the light emission is very inefficient. Thus efficient luminescence requires that the radiative lifetime should be much shorter than the non-radiative lifetime.

The principles discussed here are very general and apply to a wide range of light emission phenomena in solids. In the rest of this chapter we will concentrate on the luminescence generated by interband transitions in a bulk semiconductor. In subsequent chapters we will consider the light emission processes in semiconductor quantum wells (Chapter 6), molecular materials (Chapter 8) and luminescent impurities (Chapter 9).

5.2 Interband luminescence

Interband luminescence occurs in a semiconductor when an electron that has been excited into the conduction band drops back to the valence band by the emission of a photon. This simultaneously reduces the number of electrons in the conduction band and holes in the valence band by one. Interband luminescence thus corresponds to the annihilation of an electron–hole pair, and is known as radiative electron–hole recombination. This should be contrasted with interband absorption, which is equivalent to the creation of an electron–hole pair.

We noted in Chapter 3 that there are very important differences between the optical properties of direct and indirect band gap materials. This is particularly true when we come to consider the interband emission processes. We must therefore consider them separately, beginning with direct gap materials.

5.2.1 Direct gap materials

Figure 5.2 shows the band diagram for an interband luminescence process in a direct gap semiconductor. The photons are emitted when electrons at the bottom of the conduction band recombine with holes at the top of the valence band. As discussed in Chapter 3, the optical transitions between the valence and conduction bands of typical direct gap semiconductors are dipole-allowed and have large matrix elements. This implies through eqn B.28 that the radiative lifetime will be short, with typical values in the range 10^{-8} – 10^{-9} s. (See Exercise 5.3.) The luminescent efficiency is therefore expected to be high.

The processes by which the electrons and holes are injected into the bands will be discussed in Sections 5.3 and 5.4. We will also see in Section 5.3.1 that the injected electrons and holes relax very rapidly to the lowest energy

Holes move *upwards* on energy band diagrams when they relax. This is because band diagrams show electron energies, not hole energies. The hole energy is zero at the top of the valence band and increases as we move further *down* into the valence band.

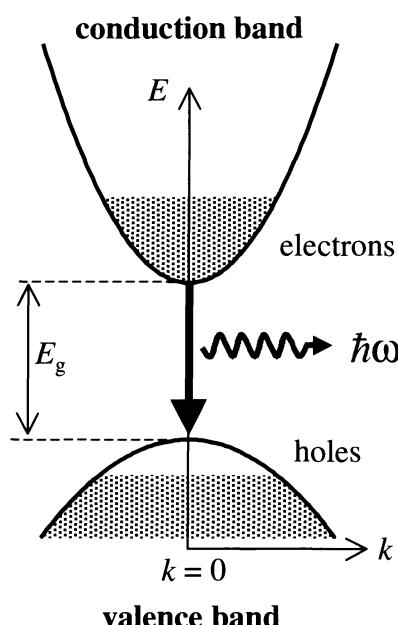


Fig. 5.2 Schematic diagram of the interband luminescence process in a direct gap semiconductor. The shading indicates that the states are occupied by electrons. The filled states at the bottom of the conduction band and the empty states at the top of the valence band are created by injecting electrons and holes into the semiconductor.

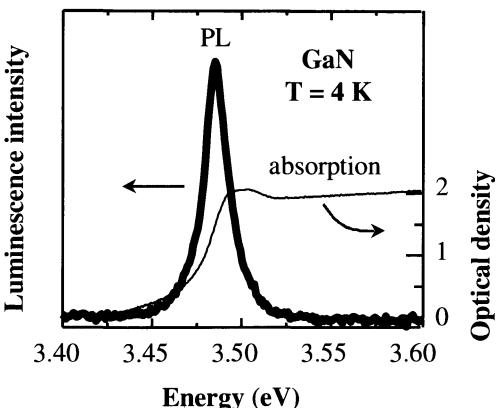


Fig. 5.3 Luminescence spectrum (thick line) and absorption (thin line) of a GaN epilayer of thickness $0.5\ \mu\text{m}$ at $4\ \text{K}$. The photoluminescence (PL) was excited by absorption of $4.9\ \text{eV}$ photons from a frequency doubled copper vapour laser. (After K.S. Kyhm and R.A. Taylor, Personal communication)

states within their respective bands by emitting phonons. This means that the electrons accumulate at the bottom of the conduction band before they recombine, as indicated in Fig. 5.2. Similarly, the holes accumulate at the top of the valence band.

Since the momentum of the photon is negligible compared to the momentum of the electron, the electron and hole that recombine must have the same \mathbf{k} vector (cf. eqn 3.12). Therefore, the transition is represented by a downward vertical arrow on the band diagram, as indicated in Fig. 5.2. The emission takes place near $k = 0$, and corresponds to a photon of energy E_g . No matter how we excite the electrons and holes in the first place, we always obtain luminescence at energies close to the band gap.

Figure 5.3 shows the luminescence and absorption spectra of the direct gap semiconductor gallium nitride at $4\ \text{K}$. The band gap is $3.5\ \text{eV}$ at this temperature. The luminescence spectrum consists of a narrow emission line close to the band gap energy, while the absorption shows the usual threshold at E_g with continuous absorption for $h\nu > E_g$.

The data shown in Fig. 5.3 illustrate the point that the emission and absorption spectra are not the same, even though they are determined by the same matrix element. The band gap corresponds to the threshold for optical absorption, but to the energy of the optical emission.

5.2.2 Indirect gap materials

Figure 5.4 illustrates the processes that occur during interband emission in an indirect gap material. This is the reverse of the indirect absorption process shown in Fig. 3.2(b). In an indirect gap material, the conduction band minimum and valence band maximum are at different points in the Brillouin zone. Conservation of momentum requires that a phonon must either be emitted or absorbed when the photon is emitted.

The requirement of emitting both a phonon and a photon during the transition makes it a second-order process, with a relatively small transition probability. The radiative lifetime is therefore much longer than for direct transitions. We can see from eqn 5.5 that this makes the luminescent efficiency small, because of the competition with non-radiative recombination. For this reason, indirect gap materials are generally bad light emitters. They are only used when

The difference between the absorption and emission properties of semiconductors imply that the criteria for choosing the best material to act as an emitter or detector for a particular wavelength are different. When we are designing an emitter, we must choose a material that has a band gap corresponding to the desired wavelength. Detectors, on the other hand, will work at any wavelength provided that the photon energy exceeds E_g .

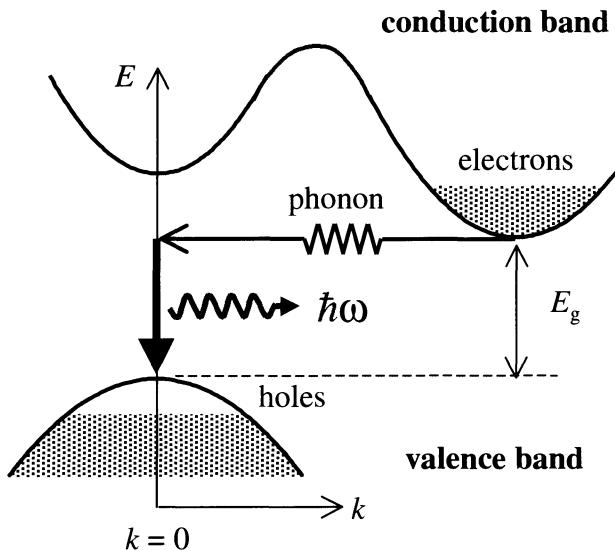


Fig. 5.4 Schematic diagram of the interband luminescence process in an indirect gap material. The transition must involve the absorption or emission of a phonon to conserve momentum.

there is no alternative direct gap material available. Two of the most important semiconductors, namely silicon and germanium, have indirect band gaps and are therefore not used as light emitters.

Example 5.1

The band gap of the III-V semiconductor alloy $\text{Al}_x\text{Ga}_{1-x}\text{As}$ at $k = 0$ varies with composition according to $E_g(x) = (1.420 + 1.087x + 0.438x^2)$ eV. The material is direct for $x \leq 0.43$, and indirect for larger values of x . Light emitters for specific wavelengths can be made by appropriate choice of the composition.

- Calculate the composition of the alloy in a device emitting at 800 nm.
- Calculate the range of wavelengths than can usefully be obtained from an AlGaAs emitter.

Solution

- The photons at 800 nm have an energy of 1.55 eV. The device will emit at the band gap wavelength, so we must choose x such that $E_g(x) = 1.55$ eV. On substituting into the relationship for $E_g(x)$, we find $x = 0.11$.
- The long wavelength limit is set by the smallest band gap that can be obtained in the alloy, namely 1.420 eV for $x = 0$. The short wavelength limit is set by the largest direct band gap that can be obtained, namely 1.97 eV for $x = 0.43$. The useful emission range is therefore 1.42–1.97 eV, or 630–870 nm. Alloy compositions with $x > 0.43$ are not useful because indirect gap materials have very low luminescent efficiencies.

5.3 Photoluminescence

In this section we consider the re-emission of light by interband luminescence after a direct gap semiconductor has been excited by a photon with energy greater than E_g . As noted at the start of Section 5.1, this process is called photoluminescence.

5.3.1 Excitation and relaxation

The band diagram corresponding to the photoluminescence process in a direct gap material is given in Fig. 5.5(a). This is a more detailed version of the diagram already given in Fig. 5.2. Photons are absorbed from an excitation source such as a laser or lamp, and this injects electrons into the conduction band and holes into the valence band. This will be possible if the frequency ν_L of the source is chosen so that $h\nu_L$ is greater than E_g .

It is apparent from Fig. 5.5(a) that the electrons are initially created in states high up in the conduction band. The electrons do not remain in these initial states for very long, because they can lose their energy very rapidly by emitting phonons. This process is indicated by the cascade of transitions within the conduction band shown in Fig. 5.5(a). Each step corresponds to the emission of a phonon with the correct energy and momentum to satisfy the conservation laws. The electron–phonon coupling in most solids is very strong and these scattering events take place on time scales as short as ~ 100 fs (i.e. $\sim 10^{-13}$ s). This is much faster than the radiative lifetimes which are in the nanosecond range, and the electrons are therefore able to relax to the bottom of the conduction band long before they have had time to emit photons. The same conditions apply to the relaxation of the holes in the valence band.

After the electrons and holes have relaxed as far as they can by phonon emission, they must wait at the bottom of the bands until they can emit a photon or recombine non-radiatively. This leaves time to form thermal

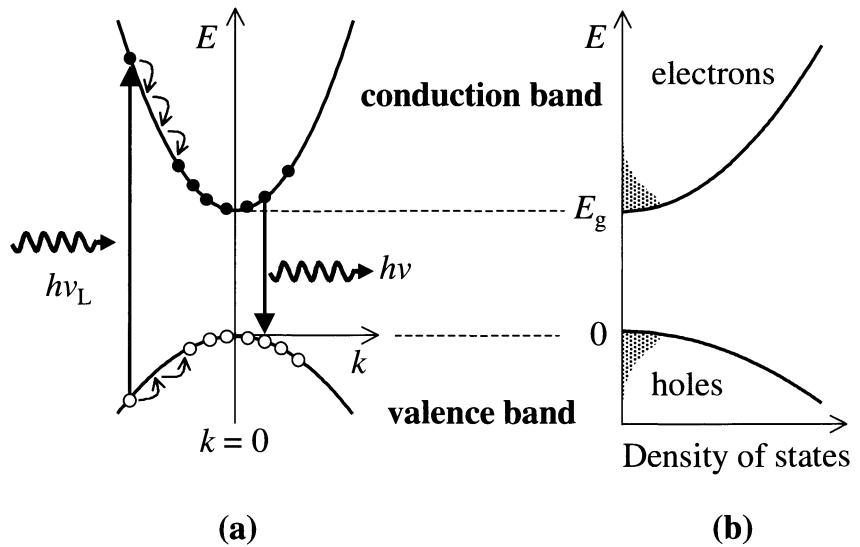


Fig. 5.5 (a) Schematic diagram of the processes occurring during photoluminescence in a direct gap semiconductor after excitation at frequency ν_L . The electrons and holes rapidly relax to the bottom of their bands by phonon emission before recombining by emitting a photon. (b) Density of states and level occupancies for the electrons and holes after optical excitation. The distribution functions shown by the shading apply to the classical limit where Boltzmann statistics are valid.

distributions, as sketched in Fig. 5.5(b). The shading indicates the occupancy of the available states. These occupancy factors can be calculated by applying statistical physics to the electron and hole distributions.

The distributions of the optically excited electrons and holes in their bands can be calculated by Fermi–Dirac statistics. The total number density N_e of electrons is determined by the power of the illumination source (see Exercises 5.6 and 5.7), and must satisfy the following equation:

$$N_e = \int_{E_g}^{\infty} g_c(E) f_e(E) dE , \quad (5.6)$$

where $g_c(E)$ is the density of states in the conduction band and $f_e(E)$ is the Fermi–Dirac distribution for the electrons. $g_c(E)$ is given by eqn 3.16 with m^* replaced by m_e^* :

$$g_c(E) = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{\frac{3}{2}} (E - E_g)^{\frac{1}{2}} . \quad (5.7)$$

Likewise, $f_e(E)$ is given by the Fermi–Dirac formula at temperature T :

$$f_e(E) = \left[\exp \left(\frac{E - E_F^c}{k_B T} \right) + 1 \right]^{-1} . \quad (5.8)$$

Note that we have added a superscript c to the Fermi level E_F to indicate that it only applies to the electrons in the conduction band. This is needed because we are in a situation of **quasi-equilibrium** in which there is no unique Fermi energy, and the electrons and holes have different Fermi levels.

The Fermi integrals can be put in a more transparent form by changing the variables such that we start the electron energy at the bottom of the conduction band. We then combine eqns 5.6–5.8 to obtain

$$N_e = \int_0^{\infty} \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}} \left[\exp \left(\frac{E - E_F^c}{k_B T} \right) + 1 \right]^{-1} dE , \quad (5.9)$$

where E_F^c is now measured relative to the bottom of the conduction band. In the same way, we can write

$$N_h = \int_0^{\infty} \frac{1}{2\pi^2} \left(\frac{2m_h^*}{\hbar^2} \right)^{\frac{3}{2}} E^{\frac{1}{2}} \left[\exp \left(\frac{E - E_F^v}{k_B T} \right) + 1 \right]^{-1} dE , \quad (5.10)$$

for the holes, where $E = 0$ corresponds to the top of the valence band and the energy is measured downwards. The Fermi energy for the holes E_F^v is also measured downwards from the top of the valence band. Note that N_e must equal N_h here because the photoexcitation process creates equal numbers of electrons and holes.

Equations 5.9 and 5.10 can be used to determine the electron and hole Fermi energies for a given carrier density. Once these are known, the occupancy factors required to calculate the emission spectrum using eqn 5.3 can be computed. Unfortunately, the general solution of eqns 5.9 and 5.10 requires numerical methods. However, the equations simplify in two important limits. These are discussed separately below.

When we apply statistical mechanics to the carriers generated by optical excitation, it is important to realize that we are dealing with a non-equilibrium situation: there are more electrons and holes present than there would normally be just from the thermal excitation of electrons across the band gap. The system is therefore in a state of ‘quasi-equilibrium’. This means that the electrons and holes form thermal distributions but with separate Fermi energies. This should be contrasted with full thermal equilibrium in which the electrons and holes share the same Fermi energy. Full thermal equilibrium can only be restored by turning off the excitation source, or by waiting for the excess electrons and holes created by a pulsed light source to recombine.

The Fermi–Dirac function of the holes has the same form as eqn 5.8. The distribution function $f_h(E)$ gives the probability that the state is occupied by a hole. This is equal to the probability that the state is *unoccupied* by an electron.

5.3.2 Low carrier densities

At low carrier densities, the electron and hole distributions will be described by classical statistics. The distributions shown in Fig. 5.5(b) are drawn for this limit. In this situation the occupancy of the levels is small and we can ignore the $+1$ factor in eqn 5.8. The occupancies are then just given by Boltzmann statistics:

$$f(E) \propto \exp\left(-\frac{E}{k_B T}\right). \quad (5.11)$$

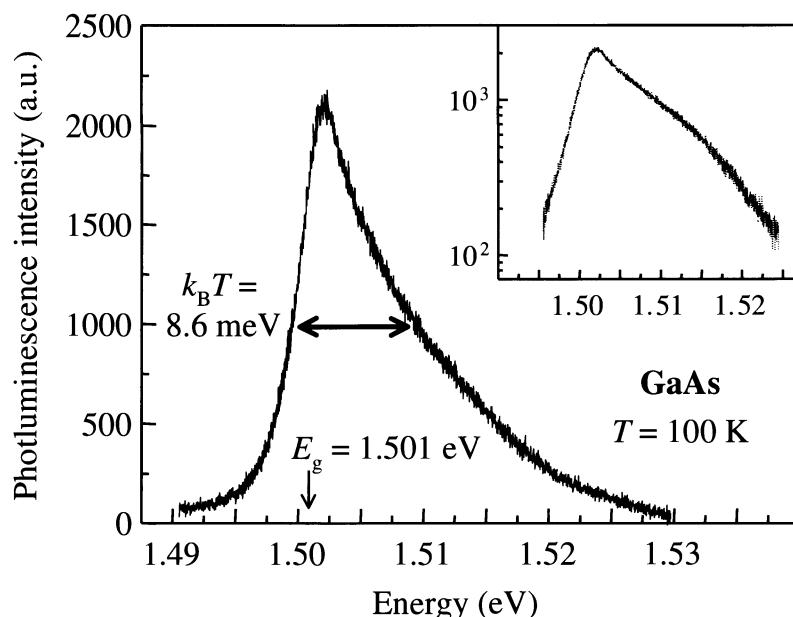
Equation 5.11 will be valid for the electrons if E_F^c is large and negative. Exercise 5.9 explores this limit. It is reasonably obvious that it will be valid at low carrier densities and high temperatures.

The frequency dependence of the emission spectrum in the classical limit can be calculated if we assume that the matrix element in eqn 5.3 is independent of frequency. We can then evaluate all the factors in eqn 5.3 and obtain:

$$I(h\nu) \propto (h\nu - E_g)^{1/2} \exp\left(-\frac{h\nu - E_g}{k_B T}\right). \quad (5.12)$$

The $(h\nu - E_g)^{1/2}$ factor arises from the joint density of states for the interband transition (cf. eqn 3.24). The final factor arises from the Boltzmann statistics of the electrons and holes: see Exercise 5.8. The luminescence spectrum described by eqn 5.12 rises sharply at E_g and then falls off exponentially with a decay constant of $k_B T$ due to the Boltzmann factor. We thus expect a sharply peaked spectrum of width $\sim k_B T$ starting at E_g .

Figure 5.6 shows the photoluminescence spectrum of GaAs at a temperature of 100 K. The spectrum was obtained using 1.96 eV photons from a helium-neon laser as the excitation source. The spectrum shows a sharp rise at E_g due to the $(h\nu - E_g)^{1/2}$ factor in eqn 5.12, and then falls off exponentially due to the



At very low temperatures, the emission spectrum from a direct gap semiconductor begins to depart from the form predicted by eqn 5.12, even for very low carrier densities. This is caused by the formation of excitons, and the possibility of radiative recombination involving impurities.

Fig. 5.6 Photoluminescence spectrum of GaAs at 100 K. The excitation source was a helium neon laser operating at 632.8 nm. The inset gives a semilogarithmic plot of the same data. (After A.D. Ashmore and M. Hopkinson, Personal communication)

Boltzmann factor. The full width at half maximum of the emission line is very close to $k_B T$, as expected. The fact that the high energy decay is exponential is clearly shown by the semilogarithmic plot of the same data given in the inset. The slope of the decay is consistent with the carrier temperature of 100 K.

5.3.3 Degeneracy

At high carrier densities, the classical limit will no longer be valid. The Fermi energies will be positive, and it is essential to use Fermi–Dirac statistics to describe the electron and hole distributions. This situation is called **degeneracy**.

In the extreme limit of $T = 0$, all the states up to the Fermi energy are filled and all states above it are empty. The Fermi energies can be calculated explicitly (see Exercise 5.10) and are given by:

$$E_F^{c,v} = \frac{\hbar^2}{2m_{e,h}^*} (3\pi^2 N_{e,h})^{\frac{2}{3}}. \quad (5.13)$$

The distribution of the carriers in this limit is shown in Fig. 5.7. Electron–hole recombination can occur between any states in which there is an electron in the upper level and a hole in the lower level. Recombination is thus possible for a range of photon energies between E_g and $(E_g + E_F^c + E_F^v)$. We therefore expect to observe a broad emission spectrum starting at E_g up to a sharp cut-off at $(E_g + E_F^c + E_F^v)$.

As finite temperatures the carriers will still be degenerate provided that $E_F^{c,v} \gg k_B T$, where $E_F^{c,v}$ is calculated using eqn 5.13. As T increases, the Fermi–Dirac functions smear out around the Fermi energies, and we expect to observe that the cut-off at $(E_g + E_F^c + E_F^v)$ will be broadened over an energy range $\sim k_B T$.

Figure 5.8 shows the emission spectrum of the III–V alloy semiconductor $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ in the degenerate limit. $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ has a direct gap of 0.81 eV at the lattice temperature T_L of 10 K. The spectra were obtained using the techniques of time-resolved photoluminescence spectroscopy described in Section 5.3.4 below. The figure shows the emission spectrum recorded at two different times after the sample has been excited with an ultrashort (< 8 ps) pulse from a dye laser operating at 610 nm. Each pulse has an energy of 6 nJ and is able to excite an initial carrier density of $2 \times 10^{24} \text{ m}^{-3}$.

The spectrum taken 24 ps after the pulse arrives rises sharply at E_g , and then shows a flat plateau up to ~ 0.90 eV. The spectrum then gradually falls off to zero at higher energies. The flat plateau is a signature of the degenerate carriers, while the high energy tail is an indication that the effective carrier temperature is higher than T_L due to the ‘hot carrier’ effect. In this case, the effective carrier temperature is 180 K. At 250 ps the carrier density is lower because a significant number of the electrons and holes have recombined, and the carriers have also cooled to a temperature of 55 K. At still longer times, the spectrum continues to narrow as the carrier density decreases and the carriers cool further towards the lattice temperature of 10 K. Eventually, the carrier density falls to the point where classical statistics are appropriate, and the emission only occurs at energies close to E_g . The analysis of this data is explored in more detail in Exercise 5.14.

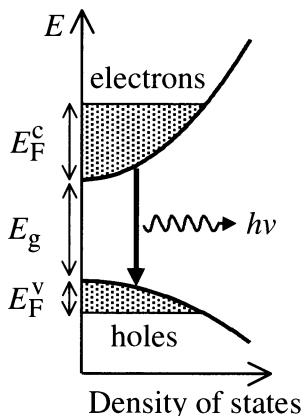


Fig. 5.7 Occupancy of the conduction and valence band states in the degenerate limit at $T = 0$. The electrons and holes have separate Fermi energies E_F^c and E_F^v respectively which are determined by the number of carriers injected into the bands. The conduction and valence bands are filled up to their respective Fermi levels, as shown by the shading.

Effective temperatures higher than T_L are possible because the carriers are not in full thermal equilibrium with the lattice. The carriers are ‘hot’ in the same sense that boiling water that has just been poured into a cold cup is hot: the temperatures are different initially, but gradually converge as heat flows from the water to the cup. In the case we are considering here, the electrons and holes are created high up the bands. This gives them a large amount of kinetic energy, which implies that their initial effective temperature is very high. This follows because the temperature is just a measure of the distribution of the carriers among the energy levels of the system. The temperature decreases rapidly as energy flows from the carriers to the lattice by phonon emission. The cooling towards T_L is therefore determined by the electron–phonon interactions in the material. Experimental values of the effective carrier temperature can be obtained by detailed modelling of the luminescence spectra.

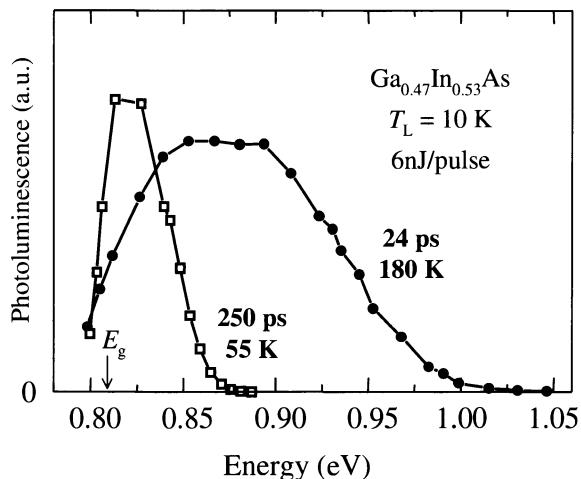


Fig. 5.8 Time-resolved photoluminescence spectra of the direct gap III-V alloy semiconductor $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ at a lattice temperature T_L of 10 K. The sample was excited with laser pulses at 610 nm with an energy of 6 nJ and a duration of 8 ps. This generated an initial carrier density of $2 \times 10^{24} \text{ m}^{-3}$. Spectra are shown for time delays of 24 ps (filled circles) and 250 ps (open squares). The effective carrier temperature at the two time delays is indicated. After [1], copyright 1984 American Institute of Physics, reprinted with permission.

5.3.4 Photoluminescence spectroscopy

Photoluminescence spectroscopy is mainly used as a diagnostic and development tool in semiconductor research. The usual goal is to develop electroluminescent devices such as light-emitting diodes and lasers. This is usually only achieved after the emission mechanisms have been studied in detail using photoluminescence spectroscopy.

Photoluminescence spectra can be recorded with an experimental arrangement such as the one shown in Fig. 5.9. The sample is mounted in a variable temperature cryostat and is illuminated with a laser or bright lamp with photon energy greater than E_g . If a liquid helium cryostat is used, sample temperatures from 2 K upwards are easily obtained. The luminescence is emitted at lower frequencies and in all directions. A portion is collected with a lens and focussed onto the entrance slit of a spectrometer. The spectrum is recorded by scanning the spectrometer and measuring the intensity at each wavelength with a sensitive detector such as a photomultiplier tube. Alternatively, the whole spectrum is recorded at once using an array of detectors such as a charge coupled device (CCD).

A number of useful variations of the basic photoluminescence technique have been developed over the years. In **photoluminescence excitation spectroscopy** (PLE), the sample is excited with a tunable laser, and the intensity of the luminescence at the peak of the emission is measured as the laser wavelength is tuned. Since the shape of the emission spectrum is independent of the way the carriers are excited, the signal strength is simply proportional to the carrier density, which in turn is determined by the absorption coefficient. (See Exercise 5.6.) Hence the signal is proportional to the absorption coefficient at the laser wavelength. This might seem to be a very complicated way to measure the absorption, but it is actually very useful. Many semiconductor samples are grown as thin layers on top of a thick substrate which is opaque at the wavelengths of interest. This makes it impossible to perform direct transmission measurements, and the use of the PLE technique allows the absorption spectrum to be measured in conditions where it would not be possible otherwise.

In time-resolved photoluminescence spectroscopy the sample is excited with a very short light pulse and the emission spectrum is recorded as a function of time after the pulse arrives. The spectra are obtained using the arrangement shown in Fig. 5.9 but with an ultrafast pulse laser as the excitation source. Lasers emitting pulses shorter than 1 ps are now readily available, and the time resolution is usually limited by the response time of the detector. Time resolutions down to ~ 100 ps can be obtained using photomultiplier tubes, while resolutions down to 1 ps or better are possible using ‘streak camera’ or ‘up-conversion’ techniques. The time-dependence of the emission spectrum gives direct information about the carrier relaxation and recombination mechanisms, and allows the radiative lifetimes to be measured. Figure 5.8 gives an example of the data that can be obtained using this technique.

5.4 Electroluminescence

Electroluminescence is the process by which luminescence is generated while an electrical current flows through an optoelectronic device. There are two main types of device:

- light emitting diodes (LEDs)
- laser diodes.

We will look at both of these devices in this section, after first discussing the general physical principles which determine their operation.

5.4.1 General principles of electroluminescent devices

Figure 5.10 shows the layer structure and circuit diagram for a typical electroluminescent device. The device consists of several epitaxial layers grown on top of a thick crystal substrate. The epitaxial layers consist of a p-n diode with a thin active region at the junction. The diode is operated in forward bias with a current flowing from the p-layer through to the n-layer underneath. The luminescence is generated in the active region by the recombination of electrons that flow in from the n-type layer with holes that flow in from the p-type side.

The microscopic mechanisms that determine the emission spectrum are exactly the same as the ones discussed in the context of photoluminescence in Sections 5.3.1–5.3.3. The only difference is that the carriers are injected electrically rather than optically. At room temperature we therefore expect a single emission line of width $\sim k_B T$ at the band gap energy E_g . Hence E_g determines the emission wavelength.

We pointed out in Section 5.2 that the radiative efficiency of indirect gap materials is low. Commercial electroluminescent devices are therefore made from direct gap compounds. Any direct gap semiconductor can, in principle, be used for the active region, but in practice only a few materials are commonly employed. The main factors that determine the choice of the material are:

- (1) the size of the band gap;
- (2) constraints relating to lattice matching;
- (3) the ease of p-type doping.

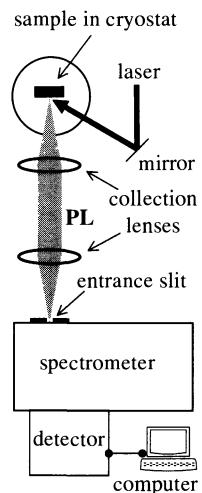


Fig. 5.9 Experimental arrangement used for the observation of photoluminescence (PL) spectra. The sample is excited with a laser or lamp with photon energy greater than the band gap. The spectrum is obtained by recording the emission as a function of wavelength using a computer-controlled spectrometer and detector. In photoluminescence excitation spectroscopy (PLE), the detection wavelength is fixed and the excitation wavelength is scanned. In time-resolved photoluminescence spectroscopy, a pulsed laser is used, and the emission at each wavelength is recorded on a fast detector as a function of time after the pulse has arrived.

We concentrate here on electroluminescence from inorganic semiconductors. In Section 8.6, we will discuss an alternative approach to making visible LEDs using the burgeoning technology of light emitting polymers.

In the past, some indirect gap materials have been used for lack of practical direct gap alternatives. For example, gallium phosphide based devices have been used for yellow and green LEDs, while silicon carbide has been used for the blue spectral region. The active regions of these devices are often doped to promote recombination via impurities and hence increase the luminescent quantum efficiency. The advent of efficient direct gap nitride LEDs in 1995 has made these indirect gap devices obsolete.

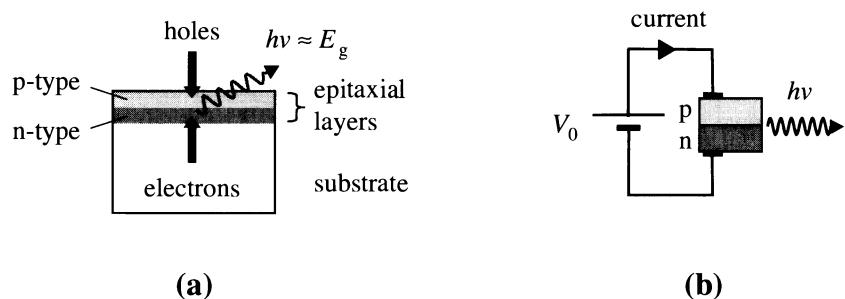


Fig. 5.10 (a) Layer structure and (b) circuit diagram for a typical electroluminescent device. The thin active region at the junction of the p- and n-layers is not shown, and the dimensions are not drawn to scale. The thickness of the epitaxial layers will be only $\sim 1\text{ }\mu\text{m}$, whereas the substrate might be $\sim 500\text{ }\mu\text{m}$ thick. The lateral dimensions of the device might be several millimetres.

The first point is obvious: the band gap determines the emission wavelength. The second and third points are practical ones relating to the way the devices are made. These are discussed further below.

The term **lattice matching** relates to the relative size of the lattice parameters of the epitaxial layers and the substrate. The thin epitaxial layers are grown on top of a substrate crystal, as shown in Fig. 5.10(a). This is done for practical reasons. It is hard to grow large crystals with sufficient purity to emit light efficiently. We therefore grow thin ultra-pure layers on top of a substrate of poorer optical quality by various techniques of crystal **epitaxy**. The crystal growth conditions constrain the epitaxial layers to form with the same unit cell size as the substrate crystal. This means that the epitaxial layers will be highly strained unless they have the same lattice constant as the substrate, that is, that we have ‘lattice matching’ between the epitaxial layers and the substrate. If this condition is not satisfied, crystal dislocations are likely to form in the epitaxial layers, which would severely degrade the optical quality.

Figure 5.11 plots the band gap of a number of III-V materials used in electroluminescent devices against their lattice constant. The lattice constants of the commonly used substrate crystal are indicated at the top of the figure. The materials separate into two distinct groups. On the right we have the arsenic and phosphorous compounds which crystallize with the cubic zinc blende structure, while on the left we have the nitride compounds which have the hexagonal wurtzite structure. We will discuss the cubic materials first, and then consider the nitrides afterwards.

For many years, the optoelectronics industry has been mainly based on GaAs. GaAs emits in the infrared at 870 nm, and by mixing it with AlAs to form the alloy $\text{Al}_x\text{Ga}_{1-x}\text{As}$, light emitters for the range 630–870 nm can be produced. (See Example 5.1.) Lattice-matched AlGaAs can easily be grown on GaAs substrates because of the convenient coincidence that the lattice constants of GaAs and AlAs are almost identical. AlGaAs emitters are widely used in local area fibre optic networks operating around 850 nm, and also for red LEDs.

AlGaAs is an example of a ‘ternary’ alloy which contains three elements. ‘Quaternary’ alloys such as $(\text{Al}_y\text{Ga}_{1-y})_x\text{In}_{1-x}\text{P}$ can also be formed. All of these arsenic and phosphorous alloys suffer from the problem that they become indirect as the band gap gets larger. This limits their usefulness to the red and near-infrared spectral range.

Applications in the fibre optics industry require light emitting devices that operate around $1.3\text{ }\mu\text{m}$ and $1.55\text{ }\mu\text{m}$. These are the wavelengths where silica

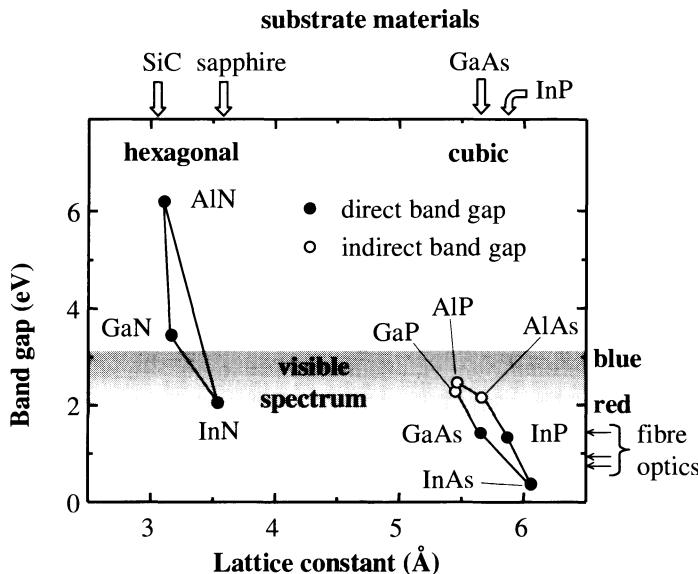


Fig. 5.11 Band gap of selected III-V semiconductors as a function of the their lattice constant. The materials included in the diagram are the ones commonly used for making LEDs and laser diodes. The lattice constants of readily available substrate crystals are indicated along the top axis. The nitride materials on the left grow with the hexagonal wurtzite structure, whereas the phosphides and arsenides on the right have the cubic zinc blende structure. After [2].

fibres have the lowest dispersion and loss respectively. Emitters for these wavelengths tend to be made from the quaternary alloy $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$. Lattice matching to InP substrates can be achieved if $x \approx 0.47y$. This allows a whole range of direct gap compounds to be made with emission wavelengths varying from $0.92 \mu\text{m}$ to $1.65 \mu\text{m}$. See Table 5.1.

Until fairly recently, it has been very difficult to make efficient electroluminescent devices for the green and blue spectral regions using III-V compounds. This is because of the problem that has already been mentioned, namely that the arsenic and phosphorous compounds become indirect as the band gap gets larger. However, in 1995 Shuji Nakamura at Nichia Chemical Industries in Japan made an important breakthrough and reported the successful development of LEDs based on gallium nitride compounds. GaN has a direct band gap of 3.5 eV at 4 K (see Fig. 5.3) and 3.4 eV at room temperature. By alloying it with InN, which has a direct gap of 1.9 eV at 300 K , the emission wavelength can be varied between 360 nm (ultraviolet) and 650 nm (red). This enables the entire visible spectrum to be covered using nitrides for the blue and green colours, and AlGaInP alloys for the reds.

It is interesting to consider why it took so long to develop the nitride devices. It was well known that the nitrides would in principle make good blue/green emitters, but no commercial devices were available. The reason for this relates to the third point on our list of factors affecting the choice of electroluminescent materials, namely the difficulty of p-type doping. This is a problem that has also dogged other wide band gap materials. For example, the direct gap II-VI compounds like ZnSe and CdSe should also, in principle, make good LEDs for the blue/green/yellow spectral regions, but they have never found widespread commercial application due to the doping problem.

P-type doping is difficult in wide band gap semiconductors because they have very deep acceptor levels. The energies of the acceptors are given by eqn 7.29 with m_e^* replaced by m_h^* . The high value of m_h^* and the relatively small

Table 5.1 Band gap energy E_g and emission wavelength λ_g for several compositions of the direct band gap quaternary III-V alloy $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$. The compositions indicated all satisfy the lattice-matching condition for InP substrates, namely $x \approx 0.47y$. After [2].

x	y	E_g (eV)	λ_g (μm)
0	0	1.35	0.92
0.27	0.58	0.95	1.30
0.40	0.85	0.80	1.55
0.47	1	0.75	1.65

We will see in Section 9.5 that the nitride emitters can also be combined with phosphors to make efficient white light sources.

There is another point that is surprising about Nakamura's development of nitride LEDs. It is apparent from Fig. 5.11 that the best substrate material from the point of view of lattice-matching considerations is silicon carbide. However, SiC is very expensive, and the commercial devices tend to be grown on the cheaper sapphire substrates. This combination is far from satisfying the lattice-matching condition, but the devices still work very efficiently. One factor that has made this possible is the growth of a thick 'buffer' layer between the substrate and the active region, to prevent the crystal dislocations affecting the light-emitting regions too adversely.

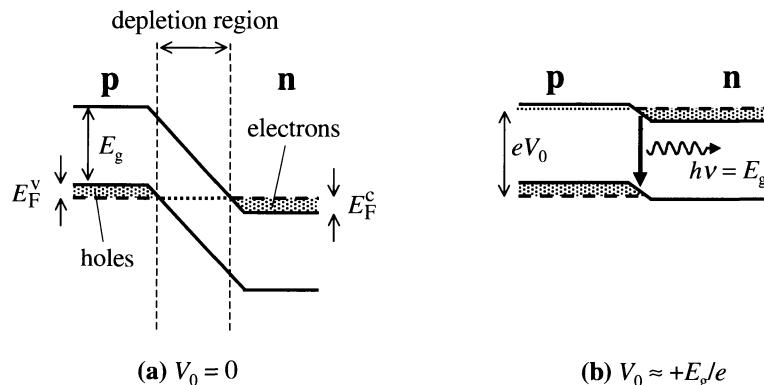


Fig. 5.12 Band diagram of a light emitting diode at (a) zero bias, and (b) forward bias $V_0 \approx E_g/e$. The diode consists of a p-n diode with heavily doped p and n regions. The dashed lines indicate the positions of the Fermi levels in the p and n regions. Light is emitted in (b) when the electrons in the n-region recombine with holes in the p-region at the junction.

value of ϵ_r increases the acceptor energies, and hence reduces the number of holes which are thermally excited into the valence band at room temperature. This last point follows from the Boltzmann factor (eqn 5.11) with E equal to the acceptor binding energy, which is significantly larger than $k_B T$. The low hole density gives the layers a high resistivity, which causes ohmic heating when the current flows and hence device failure. Nakamura's breakthrough came after discovering new techniques to activate the holes in p-type GaN by annealing the layers in nitrogen at 700 °C.

In the next chapter we will describe how the use of quantum well layers has led to further developments in the field of electroluminescent materials. In fact, many commercial devices now routinely use quantum wells in the active region. This is especially true in laser diodes, but it is also increasingly so for LEDs as well.

5.4.2 Light emitting diodes

The operating principle of an LED can be understood with reference to the band diagram shown in Fig. 5.12. The p and n regions are both very heavily doped to produce degenerate distributions of holes in the p-region and electrons in the n-region. In thermal equilibrium at zero bias, the Fermi energy must be the same at every point in the device. The bands therefore align with the Fermi energies of the p- and n-regions at the same energy, as shown in Fig. 5.12(a). At the junction, a depletion region is formed, with neither electrons nor holes present. No light can be emitted, because there is no point within the device where there is a significant population of both electrons and holes.

The situation is different when a forward bias of $V_0 \sim E_g/e$ is applied to drive a current through the device. This shifts the Fermi levels relative to each other as shown in Fig. 5.12(b). The depletion region shrinks, allowing the electrons in the n-region to diffuse into the p-region, and vice versa. This creates a region at the junction where both electrons and holes are present. The electrons recombine with the holes, emitting photons at energy E_g by interband luminescence. The electrons and holes that recombine are replenished by the current flowing through the device from the external circuit, which was given previously in Fig. 5.10(b).

Note that this is a different type of degeneracy to that considered in Section 5.3.3. In this case we have full thermal equilibrium at $V_0 = 0$, and there is a unique Fermi energy at each point in the device. Degeneracy here means that the carrier density produced by the doping is so large that the Fermi energies are positive with respect to the band edges.

Figure 5.13 shows the spectrum of a forward-biased GaAs p-i-n diode with a current of 1 mA flowing through the device. The light is generated in the thin i-region at the junction between the p- and n-regions. As mentioned previously, GaAs has a band gap of 1.42 eV at room temperature, which gives emission in the near-infrared around 870 nm. The full width at half maximum of the emission line is 58 meV, which is about twice $k_B T$ at 293 K.

5.4.3 Diode lasers

Semiconductor lasers are more difficult to make than LEDs, but they give superior performance in terms of their output efficiency, spectral linewidth, beam quality and response speed. They are therefore used for the more demanding applications, leaving the simpler ones for the cheaper LED devices. They are mainly made from GaAs-based materials, and operate in the red and near-infrared spectral regions. However, blue laser diodes have recently become available following the development of efficient nitride-based emitters.

The acronym ‘laser’ stands for ‘Light Amplification by Stimulated Emission of Radiation’. As the name suggests, laser operation is based on the quantum mechanical process of **stimulated emission**. This should be distinguished from the process of **spontaneous emission** that is responsible for luminescence. (See Section B.1 in Appendix B.) Stimulated emission causes an *increase* in the photon number as the light interacts with the atoms of the medium, which in turn leads to optical amplification. This contrasts with the process of absorption which *reduces* the number of photons, and hence causes attenuation.

Consider the interaction between a light wave of frequency ν and a medium containing atoms with an electronic transition at energy $h\nu$, as illustrated in Fig. B.2. The absorption processes causes beam attenuation, while stimulated emission causes amplification. The transition rates for the two processes are given by eqns B.5 and B.6 respectively. In the normal conditions of thermal equilibrium, the population of the lower level N_1 will be greater than the population of the upper level N_2 by the Boltzmann factor given in eqn B.8. This means that the absorption rate exceeds the stimulated emission rate, and there is net beam attenuation. However, if we were somehow to arrange for N_2 to be larger than N_1 , then the reverse would be true. The stimulated emission rate would exceed the absorption rate, and there would be net beam amplification. The condition with $N_2 > N_1$ is called **population inversion**. It is a necessary condition for laser oscillation to occur.

In Section 5.3.1 we explained how the carrier distributions after injection of electrons and holes are not those of full thermal equilibrium. The top of the valence band is empty of electrons, while the bottom of the conduction band is filled with them. We therefore have population inversion at the band gap frequency E_g/e . This gives rise to net optical gain, which can be used to obtain laser operation if an optical cavity is provided.

Figure 5.14 shows a schematic diagram of a laser cavity formed from a gain medium with mirrors at either end. This is the typical arrangement for a semiconductor laser diode, which usually consists of just the semiconductor chip itself. The reflectivities of the semiconductor-air surfaces at the edge of the crystal are typically around 30 %. (See Exercise 5.16.) This may be sufficient in itself to obtain lasing, although in what follows we assume that

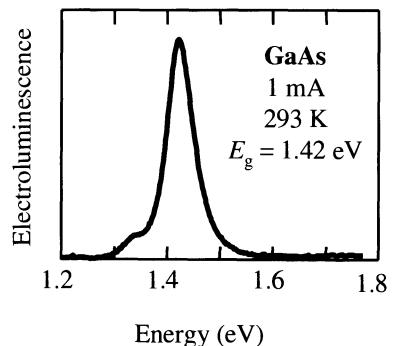


Fig. 5.13 Electroluminescence spectrum of a GaAs LED at room temperature. (After A.D. Ashmore, Personal communication)

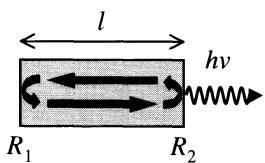


Fig. 5.14 Schematic diagram of a laser cavity formed by reflections from the end surfaces of the gain medium of length l . The reflectivities of the surfaces are taken to be R_1 and R_2 respectively, with $R_1 \gg R_2$.

the reflectivities R_1 and R_2 at the two ends are different, and that $R_1 \gg R_2$.

On passing a current through the p-n junction of a laser diode, light at frequency $\nu \approx E_g/h$ is generated by electroluminescence. This light is reflected back and forth within the cavity, and experiences gain due to the population inversion between the conduction and valence bands. At some particular value of the injection current I_{in} called the threshold current I_{th} , the laser will begin to oscillate. For current values above I_{th} , the light output power of the laser increases linearly with I_{in} . This is illustrated in Fig. 5.15(a). The output power is coupled out of the cavity by transmission through the mirror with the lower reflectivity, which is called the output coupler of the laser.

Once the laser is oscillating, the emission spectrum will be determined by the resonant longitudinal modes of the optical cavity. The resonant modes must satisfy the condition that they form standing waves between the mirrors, and hence that there are an integer number of half-wavelengths within the cavity. This condition can be written:

$$\text{integer} \times \frac{\lambda'}{2} = l. \quad (5.14)$$

where λ' is the wavelength inside the crystal, which is equal to λ/n , λ being the air wavelength and n the refractive index. This means that the frequencies of the longitudinal modes must satisfy:

$$\nu = \text{integer} \times \frac{c}{2nl}. \quad (5.15)$$

The laser will oscillate at one or several of these resonant frequencies. The best lasers are single longitudinal mode devices, and have emission line widths in the MHz range. This is many orders of magnitude smaller than that of the equivalent LED.

The condition for stable oscillation of the laser is that the light intensity in the cavity should not change with time. This implies that the gain in the laser medium must exactly balance any losses suffered by the light during a round-trip of the cavity. This condition allows us to work out the value of the gain in the medium when the laser is oscillating.

We assume that there is population inversion inside the medium, and hence that there is optical amplification at the transition frequency ν . We define the incremental gain coefficient γ_ν as:

$$dI = +\gamma_\nu dx \times I(x). \quad (5.16)$$

This is exactly the same definition as for the absorption coefficient in eqn 1.3, except that the intensity is now growing with distance rather than diminishing. Integration of eqn 5.16 yields:

$$I(x) = I_0 e^{\gamma_\nu x}. \quad (5.17)$$

We follow the light at frequency ν around a round-trip of the cavity shown in Fig. 5.14. In stable laser oscillation, the increase of the intensity due to the gain must exactly balance the losses due to the imperfect reflectivity of the end mirrors and any other losses that may be present in the medium. This condition may be written:

$$R_1 R_2 e^{2\gamma_\nu l} e^{-2\alpha_b l} = 1. \quad (5.18)$$

The factor of 2 in the two exponentials allows for the fact that the light passes through the gain medium twice during a round trip. The attenuation coefficient α_b in eqn 5.18 accounts for scattering losses and absorption due to processes other than interband transition, for example impurity absorption. The oscillation condition in eqn 5.18 can be re-written:

$$\gamma_{th} = \alpha_b - \frac{1}{2l} \ln(R_1 R_2). \quad (5.19)$$

This defines the threshold gain γ_{th} required to make the laser oscillate. Direct gap semiconductors such as GaAs have very large gain coefficients due to their high density of states and short radiative lifetimes. This makes it possible to overcome the output coupling losses with cavity lengths of order 1 mm or less.

We assume that the gain coefficient increases linearly with the injection current I_{in} , as indicated in Fig. 5.15(b). When $I_{in} = I_{th}$, the gain reaches the value γ_{th} defined by eqn 5.19, at which point the laser begins to oscillate. Once the laser is oscillating, the gain must be clamped at the value of γ_{th} , because otherwise the gain would exceed the losses, and the stability condition set out in eqn 5.18 would not hold. Thus for $I_{in} > I_{th}$, the extra electrons and holes injected into the junction do not produce any more gain, but recombine directly by stimulated emission, and cause the output power to increase, as indicated in Fig. 5.15(a).

The output power P_{out} above threshold can be written:

$$P_{out} = \eta \frac{h\nu}{e} (I_{in} - I_{th}). \quad (5.20)$$

where η is the quantum efficiency. η defines the fraction of injected electron-hole pairs that generate laser photons. The quantum efficiency determines the **slope efficiency** in watts per amp through

$$\text{slope efficiency} = \frac{P_{out}}{(I_{in} - I_{th})} = \frac{\eta h\nu}{e}. \quad (5.21)$$

In an ideal laser diode we would have $\eta = 1$ and the slope efficiency would be equal to the theoretical maximum of $h\nu/e$. Many of the best diode lasers come quite close to this ideal limit.

The main reason why η would be less than unity in a real laser diode relates to issues of **optical confinement** and **electrical confinement**. The device will not work efficiently unless we can arrange that the injection current is confined to the same part of the device where the light is confined. This is not necessarily an easy task due to the inherently planar nature of semiconductor lasers. The devices have very small dimensions (e.g. 1 μm) in the vertical (z) direction, and much larger directions (e.g. several hundred microns) in the horizontal x, y plane. The light is generated in the thin active region, and is emitted from the edge of the chip. In such a planar structure, the light tends to spread out in the y, z plane, while the current tends to spread out in the x, y direction. This leads to the possibility that the current and light might not overlap properly in the x, y plane, in which case we would have poor quantum efficiency.

There are many different ways to achieve optical and electrical confinement, and we can understand the basic principles by looking at a specific example. Figure 5.16 gives a schematic diagram of an oxide-confined GaAs-AlGaAs

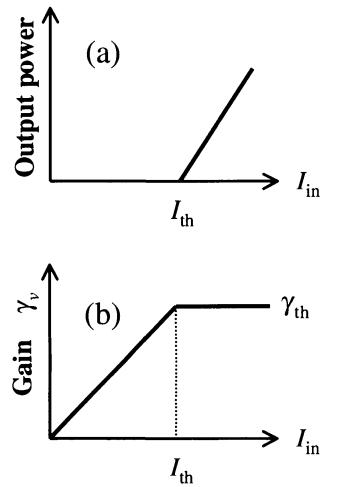


Fig. 5.15 (a) Power output and (b) gain coefficient γ_v as a function of injection current I_{in} in a semiconductor laser diode. I_{th} is the threshold injection current, and γ_{th} is the threshold gain required for stable laser oscillation.

New types of lasers called vertical cavity surface emitting lasers have different geometries to the planar lasers discussed here. The light is emitted from the top of the chip, rather than from its sides. See the references given in the further reading list for more details.

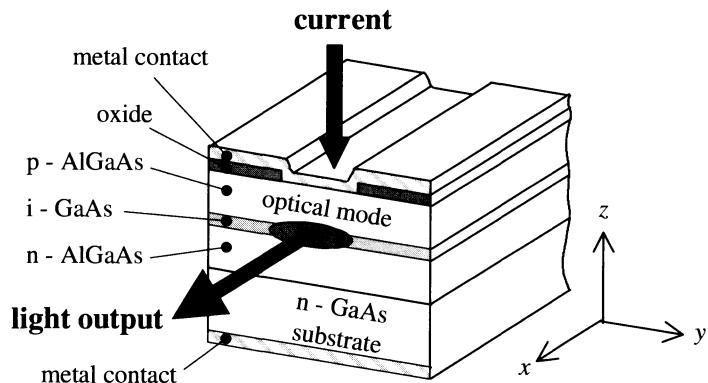


Fig. 5.16 Schematic diagram of an oxide-confined GaAs-AlGaAs heterojunction stripe laser. The current flows in the $-z$ direction, while the light propagates in the $\pm x$ direction. The stripe is defined by the gap in the insulating oxide layers deposited on the top of the device during the fabrication process. The active region is the intrinsic GaAs layer at the junction between the n- and p-type AlGaAs cladding layers.

heterostructure stripe laser. The ‘stripe’ is defined by the gap in the oxide layers deposited on the top of the device during the fabrication process. The current flows in the $-z$ direction, between the top and bottom metal contacts. The top contact only connects to the p-region in between the insulating oxide layers, and so the current is confined to the long thin rectangular strip of the x, y plane defined by the fabrication process.

The light, on the other hand, propagates in the $\pm x$ direction. The shape of the laser mode in the y, z plane is determined by **waveguide** effects. This refers to the confinement of a light beam in the direction perpendicular to its propagation instead of the usual divergence due to diffraction. The confinement in the z direction is achieved through the tendency of the light to propagate in the region with the largest refractive index, which can be understood in terms of repeated total internal reflections at the interfaces between the high and low refractive index materials. This vertical confinement is easily achieved in **heterojunction** devices such as the one shown in Fig. 5.16. In the example given, the active region is made of GaAs, which has a higher refractive index than the AlGaAs ‘cladding’ layers on either side.

The optical confinement in the y direction is more difficult. It is either achieved by index guiding or gain guiding. Index-guiding is the same effect as that used to produce the vertical confinement. The lateral patterning of the top of the chip can produce small variations in the effective refractive index in the y direction through strain or other effects. Gain guiding, on the other hand, follows as a consequence of current confinement. The semiconductor layers have very strong absorption at the laser wavelength except in the regions where there is gain due to population inversion. Hence the optical mode will be extremely lossy except in the gain regions defined by the current confinement. This is the case with the example shown in Fig. 5.16.

The reader is referred to the references given in the further reading list for more detailed information about the many different types of semiconductor laser that have been made. In the next chapter, we will explain how the use of quantum wells in the active region has led to superior performance and greater flexibility in the emission wavelength.

A heterojunction is one with different materials at the junction, and contrasts with a homojunction, where all the materials are the same. The names are derived from the Greek words *heteros*, meaning ‘other’, and *homos* meaning ‘same’. The heterojunction laser was independently invented by Zhores I. Alferov and Herbert Kroemer in 1963, for which they were awarded the Nobel Prize for physics in 2000.

Chapter summary

- Luminescence is the generic name for light emission by spontaneous emission in solids. Photoluminescence is the re-emission of light following absorption of higher energy photons. Electroluminescence is the luminescence generated by electrical excitation.
- The emission rate is proportional to the matrix element for the transition, the density of states, and the occupancy factors of the upper and lower levels.
- Transitions with high absorption coefficients have short radiative lifetimes. Efficient luminescence is only obtained when the radiative lifetime is shorter than the non-radiative lifetime.
- Interband luminescence occurs when an electron in the conduction band drops to the valence band with the emission of a photon. The process is equivalent to the recombination of an electron–hole pair. The transition is represented by a downward vertical arrow on the band diagram.
- The interband luminescence spectrum is usually independent of the way the material is excited. The emission wavelength corresponds to the fundamental band gap of the material.
- Direct gap materials have short radiative lifetimes ($\sim 1\text{ ns}$) and are strong emitters. Indirect gap materials have much longer radiative lifetimes and are generally very inefficient emitters.
- The carriers generated by photoexcitation rapidly relax to the bottom of their bands before recombining, and come to a state of quasi-equilibrium with separate Fermi energies for the electrons and holes. The luminescence spectrum can be calculated from the thermal distributions of the carriers.
- Light emitting diodes consist of p-n diodes with the light emitting material in the active region at the junction between the p and n layers. Light is emitted when the diode is forward biased. LEDs are usually made from direct gap semiconductors.
- The injection of electrons and holes into the conduction and valence bands can produce population inversion at the band gap frequency. This can support laser operation if the gain due to stimulated emission balances the round trip losses in the optical cavity.
- Semiconductor lasers are usually planar structures with the light emitted from the edge of the chip. The cavity is formed between the end mirrors at the air–semiconductor interfaces.

Further reading

A good introductory overview of luminescent processes in solids may be found in Elliott and Gibson (1974). Interband luminescence in semiconductors is discussed in Pankove (1971) and Yu and Cardona (1996). More detailed discussions may be found in Landsberg (1991) or Voos *et al.* (1980).

An authoritative discussion of time-resolved luminescence spectroscopy may be found in Shah (1999).

The physics of electroluminescence is discussed in most optoelectronics texts, for example Bhattacharya (1997), Chuang (1995), Sze (1981), Sze (1985), or Wilson and Hawkes (1998). The development of nitride light emitters is discussed in Nakamura *et al.* (2000), while detailed information about semiconductor laser diodes may be found in Silfvast (1996), Svelto (1998) or Yariv (1997).

References

- [1] Kash, K. and Shah, J. (1984). *Appl. Phys. Lett.*, **45**, 401.
- [2] Madelung, O. (1996). *Semiconductors, basic data* (2nd edn). Springer-Verlag, Berlin.

Exercises

- (5.1) Explain why it is difficult to make light emitting devices out of indirect gap materials.
- (5.2) When a direct gap semiconductor is excited by absorption of photons with energy greater than the band gap, it is generally found that the luminescence spectrum is independent of the excitation frequency. Explain this phenomenon.
- (5.3)* The wave functions for atomic hydrogen may be written in the form:

$$\Psi_{nlm}(r, \theta, \phi) = R_{nl}(r) Y_{l,m}(\theta, \phi).$$

The radial wave functions for the $1s$ and $2p$ states are given by:

$$R_{10}(r) = \frac{2}{a_H^{3/2}} e^{-r/a_H},$$

and

$$R_{21}(r) = \frac{r}{\sqrt{24} a_H^{5/2}} e^{-r/2a_H},$$

where a_H is the Bohr radius of hydrogen. The spherical harmonic functions of the same states are given by

$$Y_{0,0}(\theta, \phi) = \frac{1}{\sqrt{4\pi}},$$

$$Y_{1,0}(\theta, \phi) = \sqrt{\frac{3}{4\pi}} \cos \theta,$$

and

$$Y_{1,\pm 1}(\theta, \phi) = \mp \sqrt{\frac{3}{8\pi}} e^{\pm i\phi} \sin \theta.$$

Use eqn B.29 in Appendix B to calculate the Einstein A coefficient for the $2p \rightarrow 1s$ transition. Hence calculate the radiative lifetime of the $2p$ state.

- (5.4) The radiative lifetime τ_R of the laser transition in titanium doped sapphire is $3.9 \mu\text{s}$. The lifetime τ of the excited state is measured to be $3.1 \mu\text{s}$ at 300 K and $2.2 \mu\text{s}$ at 350 K . Explain why τ is different from τ_R , and suggest a reason why τ decreases with increasing temperature. Calculate the radiative efficiencies at the two temperatures.
- (5.5) A semiconductor crystal is found to emit efficiently at 540 nm when excited with the 488 nm line from an argon ion laser. Use the data in Table C.3 to make a guess at what the crystal is.
- (5.6) A continuous wave laser beam is incident on a material which has an absorption coefficient of α at the laser frequency ν .
 - (i) Show that electron–hole pairs are generated at a rate equal to $I\alpha/\hbar\nu$ per unit volume per unit time, where I is the intensity in the material.
 - (ii) By considering the balance between carrier generation and recombination in steady state conditions, show that the carrier density N within the illumi-

* Exercises marked with an asterisk are more challenging.

- nated volume is equal to $I\alpha\tau/\hbar\nu$, where τ is the recombination lifetime of the electrons and holes.
- (iii) Calculate N when a laser beam of power 1 mW is focussed to a circular spot of radius $50 \mu\text{m}$ on an antireflection coated sample with an excited state lifetime of 1 ns. Take the absorption coefficient to be $2 \times 10^6 \text{ m}^{-1}$ at the laser wavelength of 514 nm.
- (5.7) A very short laser pulse at 780 nm is incident on a thick crystal which has an absorption coefficient of $1.5 \times 10^6 \text{ m}^{-1}$ at this wavelength. The pulse has an energy of 10 nJ and is focussed to a circular spot of radius $100 \mu\text{m}$.
- Calculate the initial carrier density at the front of the sample.
 - If the radiative and non-radiative lifetimes of the sample are 1 ns and 8 ns respectively, calculate the time taken for the carrier density to drop to 50 % of the initial value.
 - Calculate the total number of luminescent photons generated by each laser pulse.
- (5.8) Explain why the emission probability for an interband transition is proportional to the product of the electron and hole occupancy factors f_e and f_h respectively. In the classical limit where Boltzmann statistics apply, show that the product $f_e f_h$ is proportional to $\exp(-(h\nu - E_g)/k_B T)$.
- (5.9) Show that in the classical limit the number of electrons in the conduction band is given by
- $$N_e = \exp\left(\frac{E_F^C}{k_B T}\right) \times \frac{1}{2\pi^2} \left(\frac{2m_e^* k_B T}{\hbar^2}\right)^{\frac{3}{2}} \int_0^\infty x^{\frac{1}{2}} e^{-x} dx,$$
- where $\int_0^\infty x^{\frac{1}{2}} e^{-x} dx = \sqrt{\pi}/2$. Evaluate E_F^C using this formula for electrons in GaAs at 300 K ($m_e^* = 0.067m_0$) at (a) $N_e = 1 \times 10^{20} \text{ m}^{-3}$ and (b) $N_e = 1 \times 10^{24} \text{ m}^{-3}$. Hence discuss whether the approximations used to derive this equation are justified in the two cases.
- (5.10) Show that at $T = 0$ the Fermi integrals given in eqns 5.9 and 5.10 simplify to:
- $$N_{e,h} = \int_0^{E_F^{C,V}} \frac{1}{2\pi^2} \left(\frac{2m_{e,h}^*}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}} dE.$$
- Evaluate the integral to derive eqn 5.13.
- (5.11) A laser excites a semiconductor which has $m_e^* = 0.1m_0$ and $m_h^* = 0.5m_0$. Calculate the electron and hole Fermi energies for carrier densities of (a) $1 \times 10^{21} \text{ m}^{-3}$, and (b) $1 \times 10^{24} \text{ m}^{-3}$, on the assumption that the distributions are degenerate. Write down a condition on the temperature for the degeneracy conditions to apply in each case and comment on the answers you obtain.
- (5.12) Show that in the degenerate limit during photoluminescence, the k vectors corresponding to the conduction and valence band Fermi energies are the same, even though the Fermi energies are different.
- (5.13)* The photoluminescence spectrum of CdTe, which has a direct band gap at 1.61 eV and a refractive index of 2.7, is measured using the apparatus shown in Fig. 5.9. An argon ion laser of power 1 mW and photon energy 2.41 eV is focussed to a small spot on the sample. The luminescence is collimated with a lens of diameter 25 mm and focal length 100 mm.
- Calculate the solid angle subtended by the lens at the sample.
 - Estimate the fraction of the photoluminescence collected by the lens. Assume that the luminescence is emitted uniformly in all directions inside the crystal, and is then both reflected and refracted at the front surface.
 - Calculate the total luminescent power emitted by the atoms in terms of the radiative quantum efficiency η_R of the sample.
 - Hence estimate the luminescent power collected by the collimation lens in terms of η_R .
- (5.14)* Figure 5.8 shows the emission spectrum from the direct gap semiconductor $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ at two time delays after $2 \times 10^{24} \text{ m}^{-3}$ carriers have been excited using an ultra-short laser pulse.
- Calculate the electron Fermi energy for the initial carrier density if $T = 0$. ($m_e^* = 0.041m_0$.)
 - Calculate the hole Fermi energy in the same conditions, on the assumption that the densities of states from the light and heavy hole bands can just be added together. ($m_{hh}^* = 0.47m_0$ and $m_{lh}^* = 0.05m_0$.)
 - The effective carrier temperature for the 24 ps spectrum is 180 K. Are the carriers degenerate?
 - Explain the shape of the 24 ps spectrum, given that the band gap of $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ is 0.81 eV.
 - Use the data at 250 ps to obtain a rough estimate of the carrier density at this time delay. Hence estimate the average lifetime of the carriers.
- (5.15) GaN and InN are direct gap semiconductors with band gaps of 3.4 eV and 1.9 eV respectively. A light emitting diode made from the alloy $\text{Ga}_x\text{In}_{1-x}\text{N}$ is found to emit at 500 nm. Estimate the composition of the alloy, on the assumption that $E_g(x)$ varies linearly with x .
- (5.16) GaAs has a refractive index of 3.5 at its band gap.
- Calculate the reflectivity of the air–GaAs interface.
 - Calculate the frequency separation of the longitudinal modes of a GaAs laser diode of length 1 mm.

- (iii) The laser diode of part (ii) is coated so that one end of the chip has a reflectivity of 95 %. The other end is uncoated. Calculate the threshold gain coefficient for the laser if the scattering and other impurity losses are negligibly small.
- (5.17) A laser diode emits at 830 nm when operating at an injection current of 100 mA.
- (i) Calculate the maximum possible power that can be emitted by the device.
- (ii) Calculate the power conversion efficiency, if the actual power output is 50 mW and the operating voltage is 1.9 V.
- (iii) The threshold current of the laser is 35 mA. What is the slope efficiency and the quantum efficiency?

Semiconductor quantum wells

6

In this chapter we will give an overview of the optical properties of quantum confined semiconductor structures. These are artificial structures in which the electrons and holes are confined in one or more directions. We will concentrate mainly on quantum well structures in which there is confinement in just one dimension. The concept of a quantum well was invented by Esaki and Tsu in 1970, and it has proven to be a very important step forward in semiconductor physics. As we will see, quantum wells have very interesting optical properties which readily lend themselves to applications in optoelectronics.

The optical properties of quantum wells are derived from the physics of interband absorption, excitons and interband luminescence discussed in Chapters 3–5. The discussion we give here presupposes that these subjects have been assimilated, and the topics covered provide a good opportunity to practice and develop the principles that we have been learning in these previous chapters.

6.1 Quantum confined structures

The optical properties of solids do not usually depend on their size. Ruby crystals, for example, have the same red colour irrespective of how big they are. This statement is only true as long as the dimensions of the crystal are large. If we make very small crystals, then the optical properties do in fact depend on the size. A striking example of this is semiconductor doped glasses. These contain very small semiconductor microcrystals within a colourless glass host, and the colour of the filter can be altered just by changing the size of the crystals.

The size dependence of the optical properties in very small crystals is a consequence of the **quantum confinement** effect. The Heisenberg uncertainty principle tells us that if we confine a particle to a region of the x axis of length Δx , then we introduce an uncertainty in its momentum given by:

$$\Delta p_x \sim \frac{\hbar}{\Delta x}. \quad (6.1)$$

If the particle is otherwise free, and has a mass m , the confinement in the x direction gives it an additional kinetic energy of magnitude

$$E_{\text{confinement}} = \frac{(\Delta p_x)^2}{2m} \sim \frac{\hbar^2}{2m(\Delta x)^2}. \quad (6.2)$$

This confinement energy will be significant if it is comparable to or greater than the kinetic energy of the particle due to its thermal motion in the x direction.

6.1	Quantum confined structures	115
6.2	Growth and structure of semiconductor quantum wells	117
6.3	Electronic levels	119
6.4	Optical absorption and excitons	124
6.5	The quantum confined Stark effect	131
6.6	Optical emission	133
6.7	Intersubband transitions	135
6.8	Bloch oscillators	135
6.9	Quantum dots	136

Table 6.1 Number of degrees of freedom tabulated against the dimensionality of the quantum confinement.

Structure	Quantum confinement	Number of free dimensions
Bulk	none	3
Quantum well/superlattice	1-D	2
Quantum wire	2-D	1
Quantum dot/box	3-D	0

This condition may be written:

$$E_{\text{confinement}} \sim \frac{\hbar^2}{2m(\Delta x)^2} > \frac{1}{2}k_B T, \quad (6.3)$$

The principle of equipartition of energy tells us that we have a thermal energy of $k_B T/2$ for each degree of freedom of the motion.

We could in principle observe quantum size effects in a semiconductor layer of thickness $1 \mu\text{m}$ if we were to work at extremely low temperatures: see Exercise 6.1.

Certain types of quantum well structures are called superlattices (see Section 6.2), while quantum dots are sometimes called quantum boxes.

and tells us that quantum size effects will be important if

$$\Delta x \sim \sqrt{\frac{\hbar^2}{mk_B T}}. \quad (6.4)$$

This is equivalent to saying that Δx must be of the same order of magnitude as the de Broglie wavelength $\lambda_{\text{deB}} \equiv p_x/\hbar$ for the thermal motion.

The criterion given in eqn 6.4 gives us an idea of how small the structure must be if we are to observe quantum confinement effects. For an electron in a typical semiconductor with $m_e^* = 0.1m_0$ at room temperature, we find that we must have $\Delta x \sim 5 \text{ nm}$. Thus a ‘thin’ semiconductor layer of thickness $1 \mu\text{m}$ is not thin by the standards of the electrons. It is in fact a bulk crystal which would not exhibit any quantum size effects. To observe quantum size effects we need thinner layers.

Table 6.1 summarizes the three basic types of quantum confined structures that can be produced. The structures are classified as to whether the electrons are confined in one, two or three dimensions. These structures are respectively called:

- **quantum wells** (1-D confinement);
- **quantum wires** (2-D confinement);
- **quantum dots** (3-D confinement).

Table 6.1 also lists the number of degrees of freedom associated with the type of quantum confinement. The electrons and holes in bulk semiconductors are free to move within their respective bands in all three directions, and hence they have three degrees of freedom. The electrons and holes in a quantum well, by contrast, are confined in one direction, and therefore only have two degrees of freedom. This means that they effectively behave as two-dimensional (2-D) materials. Similarly, quantum wire structures have 1-D physics, while quantum dots have ‘0-D’ physics. This last point means that the motion of the electrons and holes is quantized in all three dimensions, so that they are completely localized to the quantum dot.

The very small crystal dimensions required to observe quantum confinement effects have to be produced by special techniques.

- Quantum well structures are made by techniques of advanced epitaxial crystal growth. This will be explained in Section 6.2.
 - Quantum wire structures are made by lithographic patterning of quantum well structures, or by epitaxial growth on patterned substrates.
 - Quantum dots structures can be made by lithographic patterning of quantum wells or by spontaneous growth techniques.

In the sections that follow, we will mainly concentrate almost entirely on quantum well structures. This is because they have the most mature technologies, and are already widely used in many commercial applications. We will not mention quantum wires at all, because these are only being developed at the research level at present, due to the difficulties associated with the fabrication techniques. In Section 6.9 we will briefly consider the optical properties of quantum dots such as those found in semiconductor doped glasses.

A typical example of the spontaneous growth of quantum dots is the formation of microcrystals in semiconductor doped glasses mentioned above.

6.2 Growth and structure of semiconductor quantum wells

Semiconductor quantum wells are examples of **heterostructure** crystals. Heterostructures are artificial crystals that contain layers of different materials grown on top of a thicker substrate crystal. The structures are made by the specialized epitaxial crystal growth techniques introduced previously in Section 5.4.1. The two most important ones are **molecular beam epitaxy** (MBE) and **metal-organic chemical vapour deposition** (MOCVD), which is also called **metal-organic vapour phase epitaxy** (MOVPE) by some authors. The layer thicknesses of the crystal grown by these techniques can be controlled with atomic precision. This makes it easy to achieve the thin layer thicknesses required to observe quantum confinement of the electrons in a semiconductor at room temperature.

Figure 6.1(a) shows a schematic diagram of the simplest type of quantum well that can be grown. In this particular case, we show a GaAs/AlGaAs structure, grown on a GaAs substrate. The structure consists of a GaAs layer of

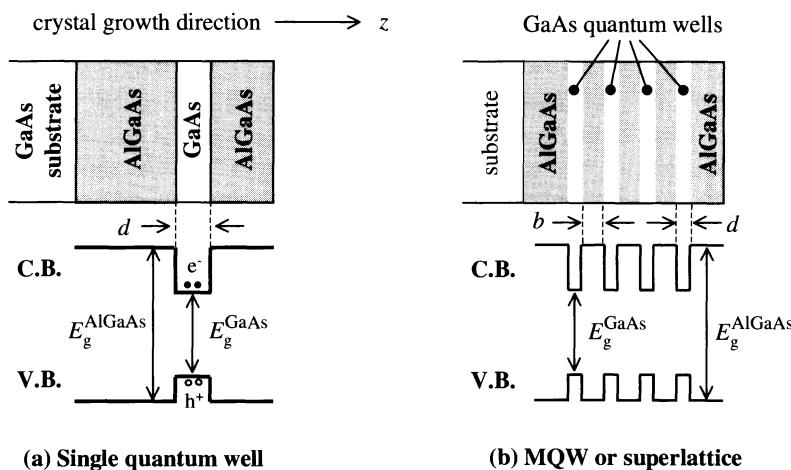


Fig. 6.1 (a) Schematic diagram of a single GaAs/AlGaAs quantum well. The quantum well is formed in the thin GaAs layer sandwiched between AlGaAs layers which have a larger band gap. The lower half of the figure shows the spatial variation of the conduction band (C.B.) and the valence band (V.B.). (b) Schematic diagram of a GaAs/AlGaAs multiple quantum well (MQW) or superlattice structure. The distinction between an MQW and a superlattice depends on the thickness b of the barrier separating the quantum wells.

thickness d sandwiched between much thicker layers of the alloy semiconductor AlGaAs. d is chosen so that the motion of the electrons in the GaAs layer is quantized according to the criterion given in eqn 6.4. We set up axes so that the z axis corresponds to the crystal growth direction, while the x and y axes lie in the plane of the layers. We thus have quantized motion if the z direction, and free motion in the x , y plane.

The bottom half of Fig. 6.1(a) shows the spatial variation of the conduction and valence bands that corresponds to the change of the composition along the z direction. The band gap of AlGaAs is larger than that of GaAs, and the bands line up so that the lowest conduction and valence band states of the GaAs lie within the gap of the AlGaAs. This means that electrons in the GaAs layer are trapped by potential barriers at each side due to the discontinuity in the conduction band. Similarly, holes are trapped by the discontinuity in the valence band. These barriers quantize the states in the z direction, but the motion in the x , y plane is still free. We thus effectively have a two-dimensional system in which the electrons and holes are quantized in one direction and free in the other two.

Epitaxial techniques are very versatile, and they allow the growth of a great variety of quantum well structures. Figure 6.1(b) shows one such variant derived from the single well structure shown in Fig. 6.1(a). The crystal consists of a series of repeated GaAs quantum wells of width d separated from each other by AlGaAs layers of thickness b . This type of structure is either called a **multiple quantum well (MQW)** or a **superlattice**, depending on the parameters of the system. The distinction depends mainly on the value of b .

MQWs have large b values, so that the individual quantum wells are isolated from each other, and the properties of the system are essentially the same as those of single quantum wells. They are often used in optical applications to give a usable optical density. It would be very difficult to measure the optical absorption of a single 10 nm thick quantum well, simply because there is so little material to absorb the light. By growing many identical quantum wells, the absorption will increase to a measurable value.

Superlattices, by contrast, have much thinner barriers. The quantum wells are then coupled together by tunnelling through the barrier, and new extended states are formed in the z direction. Superlattices have additional properties over and above those of the individual quantum wells.

Quantum well structures of the type shown in Fig. 6.1 can only be made if the physical properties of the constituent compounds are favourable to the formation of the artificial crystals. We have already noted in Section 5.4.1 that the unit cell size of GaAs and AlAs (and hence also the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ alloy) are almost identical: see Fig. 5.11. This means that both the GaAs and AlGaAs layers in the quantum well structure are lattice matched to the GaAs substrate, enabling dislocation-free crystals to be grown.

In recent years it has been realized that it is also possible to make quantum wells from materials with different unit cell sizes. This allows much more flexibility in the combinations of materials that can be used. The mismatch in the lattice constants introduces strain into the structure, but high quality crystals can still be grown provided the total layer thickness is kept below a critical value. We will briefly mention the application of these non-lattice-matched quantum wells in light emitting diodes and laser diodes in Section 6.6.

Esaki and Tsu's original proposal specifically considered one of the special properties of superlattices, namely the possibility of observing Bloch oscillations. This quantum effect was predicted in the 1930s, but has never been observed in ordinary solids. In Section 6.8 we will explain how the advent of the superlattice led to the successful observation of this effect.

6.3 Electronic levels

The wave functions and energies of the quantized states in the conduction and valence bands of a quantum well can be calculated by using Schrödinger's equation and the effective mass approximation. Fortunately, we do not have to solve the Schrödinger equation in three dimensions because the problem separates naturally between the free motion in the x , y plane and the quantized motion in the z direction. In this section, we will first explain how this separation of variables works, and then go on to discuss the quantized states in the z direction in two different approximations. We will treat the electron and holes separately here, postponing till Section 6.4 the discussion of the effects of electron–hole Coulomb interaction that leads to the formation of excitons.

6.3.1 Separation of the variables

The electrons and holes in a quantum well layer are free to move in the x , y plane but are confined in the z direction. This allows us to write the wave functions in the form:

$$\Psi(x, y, z) = \psi(x, y) \varphi(z), \quad (6.5)$$

and then solve separately for $\psi(x, y)$ and $\varphi(z)$. The states of the system are described by two parameters: a wave vector \mathbf{k} to specify the free motion in the x , y plane, and a quantum number n to indicate the energy level for the z direction. The total energy is then obtained by adding together the separate energies for the z and x , y motion, according to:

$$E^{\text{total}}(n, \mathbf{k}) = E_n + E(\mathbf{k}), \quad (6.6)$$

where E_n is the quantized energy of the n th level.

We can deal with the x , y plane motion very quickly. Since the motion is free, the electron and hole wave functions are described by plane waves of the form:

$$\psi_k(x, y) = \frac{1}{\sqrt{A}} e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (6.7)$$

where \mathbf{k} is the wave vector of the particle, and A is the normalization area. Note that \mathbf{k} and \mathbf{r} only span the two-dimensional x , y plane here. The energy corresponding to this motion is just the kinetic energy determined by the effective mass:

$$E(\mathbf{k}) = \frac{\hbar^2 \mathbf{k}^2}{2m^*}. \quad (6.8)$$

The total energy for an electron or hole in the n th quantum level is therefore given by:

$$E^{\text{total}}(n, \mathbf{k}) = E_n + \frac{\hbar^2 \mathbf{k}^2}{2m^*}. \quad (6.9)$$

6.3.2 Infinite potential wells

The calculation of the wave functions and energies for the quantized states in the z direction is determined by the spatial dependence of the conduction and

valence bands. We begin by considering the simplest case in which we assume that the confining barriers are infinitely high. This allows us to model the states by those of a 1-D potential well with infinite barriers, as shown in Fig. 6.2.

We consider a quantum well of thickness d and define position and energy coordinates such that the potential is zero for $-d/2 < z < +d/2$ (i.e. inside the well) and ∞ elsewhere, as indicated in Fig. 6.2. The choice of $z = 0$ at the centre of the well is convenient, since it corresponds to the symmetry axis of the potential. The Schrödinger equation within the well is:

$$-\frac{\hbar^2}{2m^*} \frac{d^2\varphi(z)}{dz^2} = E\varphi(z). \quad (6.10)$$

Since the barriers are infinitely high, there is no probability that the particle can tunnel out of the well. The solutions of eqn 6.10 are therefore subject to the boundary condition that $\varphi = 0$ at the interfaces.

It can be checked by substitution that the normalized wave functions that satisfy eqn 6.10 and the boundary conditions are of the form:

$$\varphi_n(z) = \sqrt{\frac{2}{d}} \sin\left(k_n z + \frac{n\pi}{2}\right) \quad (6.11)$$

where n is an integer that gives the quantum number of the state, and

$$k_n = \frac{n\pi}{d}. \quad (6.12)$$

This form of wave function describes a standing wave inside the well with nodes at the interfaces. The energy that corresponds to the n th level is given by

$$E_n = \frac{\hbar^2 k_n^2}{2m^*} = \frac{\hbar^2}{2m^*} \left(\frac{n\pi}{d}\right)^2. \quad (6.13)$$

The wave functions of the first three levels are shown in Fig. 6.2. Equation 6.13 describes an infinite ladder of levels with quantization energy increasing in proportion to n^2 in units of $(\hbar^2\pi^2/2m^*d^2)$. The ground state is the $n = 1$ level, and the levels of higher n are the excited states of the system.

The energies of the first two levels for an electron with $m^* = 0.1m_0$ in a 10 nm quantum well are 38 meV and 150 meV respectively. These values should be compared to the thermal energy $k_B T$, which is 25 meV at room temperature. It is clear that the quantization energy is greater than the thermal energy at room temperature, and thus that the quantum description of the motion is appropriate. The comparison of the quantization energies to the thermal energy gives a criterion to decide whether a particular quantum well will in fact exhibit quantum effects at a particular temperature. This criterion can be compared to the one based on the Heisenberg uncertainty principle given in eqn 6.4. It is easy to show that the two criteria predict a cross-over from classical to quantum behaviour at roughly the same value of d . In fact, the criterion based on eqn 6.4 is too strict by about a factor of five. (See Exercise 6.2.)

Although real semiconductor quantum wells have finite barriers, the infinite barrier model is a good starting point for a discussion of their properties. The accuracy of the model will be highest for states with small quantization energies in material combinations that give rise to high barriers at the interfaces. A few useful general points emerge from the analysis:

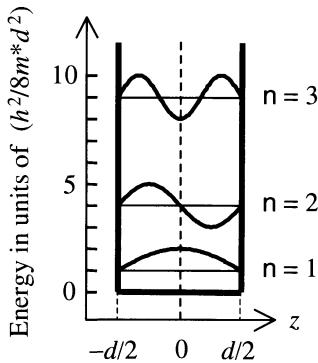


Fig. 6.2 The infinite one-dimensional potential well. The first three energy levels and corresponding wave functions are shown.

(1) The energy of the levels is inversely proportional to the effective mass and the square of the well width. This means that low mass particles in narrow quantum wells have the highest energies.

(2) Since the energy depends on the effective mass, the electrons, heavy holes and light holes will all have different quantization energies. In the valence band, the heavy holes will have the lowest energy, and are dominant in most situations because they form the ground state level.

(3) The wave functions can be identified by their number of nodes, i.e. the number of zero crossings within the well. It is evident from Fig. 6.2 that the n th level has $(n - 1)$ nodes.

(4) The states are also labelled by their parity with respect to inversion about the centre of the well, that is, whether $\varphi(-z) = +\varphi(z)$ (even parity) or $\varphi(-z) = -\varphi(z)$ (odd parity). States of odd n have even parity, and *vice versa*.

These points also apply to more realistic models of quantum wells in which the barriers at the interfaces are only of finite height. As we will see below, the infinite well model overestimates the quantization energy. In real quantum wells with finite barriers, the particles are able to tunnel into the barriers to some extent, and this allows the wave function to spread out further and thus reduces the confinement energy.

6.3.3 Finite potential wells

Figure 6.3 shows the band diagram of a more realistic quantum well which has a finite potential barrier of height V_0 at each interface. There are now only a finite number of bound states with energy $E < V_0$. These bound states are labelled by a quantum number n , and it can be shown that there is always at least one, no matter how small V_0 is: see Exercise 6.4.

The Schrödinger equation within the quantum well is the same as before (eqn 6.10). We will therefore have sine and cosine solutions of the form:

$$\varphi_w(z) = C \sin(kz) \quad (6.14)$$

and

$$\varphi_w(z) = C \cos(kz), \quad (6.15)$$

where

$$\frac{\hbar^2 k^2}{2m_w^*} = E. \quad (6.16)$$

Note that we have added a subscript w to the effective mass to clarify that it is the value for the semiconductor used for the quantum well layer. By comparison with eqn 6.11 and also Fig. 6.2, we see that the bound states with odd values of n have cosine solutions, while those with even n have sine solutions with a node at $z = 0$. The wave functions given in eqns 6.14 and 6.15 are valid for $-d/2 \leq z \leq +d/2$.

We now consider the extension of the wave functions into the barrier regions. This occurs because the finite potential discontinuity allows the electrons and holes to tunnel into the barriers. We therefore no longer have nodes at the interfaces. The Schrödinger equation in the barrier regions is given by:

$$-\frac{\hbar^2}{2m_b^*} \frac{d^2 \varphi(z)}{dz^2} + V_0 \varphi(z) = E \varphi(z), \quad (6.17)$$

We can reach the same conclusion from the uncertainty principle: see eqn 6.3.

In a bulk semiconductor like GaAs, the heavy and light hole states are degenerate at $k = 0$. This is a consequence of the high symmetry of the cubic lattice. The lifting of this degeneracy is caused by the difference in the effective masses, but it can also be seen as a consequence of the lower symmetry of the quantum well. The bulk crystals are isotropic, but quantum wells are not: the z direction is physically distinguishable from the other two. As explained in Section 1.5.1, we therefore expect certain degeneracies to be lifted, in the same way that a magnetic field splits the heavy and light hole states of bulk GaAs via their different magnetic quantum numbers.

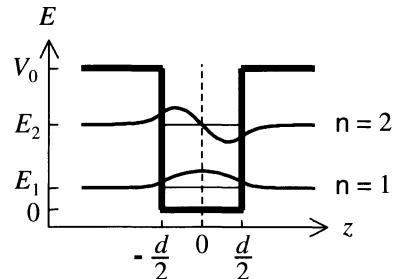


Fig. 6.3 First two bound states of a finite potential well of depth V_0 and width d .

where m_b^* is the effective mass of the barrier material. In general, m_b^* and m_w^* will not be the same because the materials that constitute the quantum well and barrier regions have different band structures. The solutions of eqn 6.17 are exponentials of the form:

$$\varphi_b(z) = C' e^{\pm \kappa z} \quad (6.18)$$

where κ satisfies:

$$\frac{\hbar^2 \kappa^2}{2m_b^*} = V_0 - E. \quad (6.19)$$

For bound states we require that the solutions decay in the barrier, so we choose $\varphi(z) = C' \exp(-\kappa z)$ for $z \geq +d/2$ and $\varphi(z) = C' \exp(-\kappa z)$ for $z \leq -d/2$.

The wave functions and energies of the bound states can be found by applying the appropriate boundary conditions at the interfaces. These tell us that both the wave function $\varphi(z)$ and the particle flux $(1/m^*)d\varphi/dz$ must be continuous at $\pm d/2$. Hence we must have that

$$\varphi_w(\pm d/2) = \varphi_b(\pm d/2), \quad (6.20)$$

and

$$\frac{1}{m_w^*} \left(\frac{d\varphi_w}{dz} \right)_{z=\pm d/2} = \frac{1}{m_b^*} \left(\frac{d\varphi_b}{dz} \right)_{z=\pm d/2}. \quad (6.21)$$

The wave functions must be symmetric about $z = 0$, so we just concentrate on $z = +d/2$. We consider the solutions with cosine solutions in the quantum wells first. The wave function continuity requires that

$$C \cos(kd/2) = C' \exp(-\kappa d/2), \quad (6.22)$$

while flux continuity requires that

$$-C \frac{k}{m_w^*} \sin(kd/2) = -C' \frac{\kappa}{m_b^*} \exp(-\kappa d/2). \quad (6.23)$$

On dividing eqn 6.23 by eqn 6.22, we find:

$$\tan(kd/2) = \frac{m_w^* \kappa}{m_b^* k}. \quad (6.24)$$

On following a similar procedure for the solutions with sine solutions in the quantum well, we find:

$$\tan(kd/2) = -\frac{m_b^* k}{m_w^* \kappa}. \quad (6.25)$$

On substituting the values of k and κ from eqns 6.16 and 6.19, we can now solve for the energy E of the bound states, and hence find the wave functions. Unfortunately, there is no analytic solution for E . The equations have to be solved numerically or graphically. An example of how this is done is given in Example 6.1.

It is useful to make a few general observations about the solutions, as we did for the case of the infinite well in the previous section.

(1) The spreading of the wave functions into the barrier by tunnelling reduces the quantum confinement energy compared to that of an infinite barrier well.

(2) The decay constant can be found by substituting E_n for E in eqn 6.19. This means that the levels near the top of the well with E_n close to V_0 tunnel more into the barrier regions because they have a smaller decay constant.

(3) The eigenstates can be identified by the number of nodes, just as for infinite wells. The n th bound state has $(n - 1)$ nodes. The potential energy has inversion symmetry about $z = 0$, and so the eigenstates have well-defined parities.

Figure 6.3 sketches the wave functions of a typical finite well with two bound states. The similarity between these wave functions and the first two states of the infinite well shown in Fig. 6.2 is apparent. The main difference is that the wave functions of the finite well spread out more by tunnelling into the barrier, whereas the wave functions of the infinite well stop abruptly at the interface.

It is useful to compare directly the predictions of the finite and infinite well models. Table 6.2 tabulates the energies of the bound states of a 10 nm GaAs quantum well with $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barriers for the two models. In all cases, the infinite well model overestimates the quantization energy. The discrepancy gets worse for the higher levels. Note that the quantization energies of the heavy holes are smaller than those of the electrons because of their heavier effective mass. Note also that the separation of the first two electron levels is more than three times the thermal energy at room temperature, where $k_B T \sim 25 \text{ meV}$. This confirms that we expect to observe two-dimensional physics for the electrons at 300 K. The quantum confinement of the heavy holes is less good, but is still acceptable since $E_2 - E_1$ is comparable to $k_B T$ at 300 K. Although the infinite well model overestimates the confinement energies, it is a useful starting point for the discussion of the physics because of its simplicity.

The possibility of tunnelling into the barriers gives rise to a whole series of electronic and optoelectronic quantum well tunnelling devices.

Table 6.2 Bound states of a 10 nm GaAs/ $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ quantum well calculated using the finite and infinite well models. The states are labelled by the particle type (e for electron, hh for heavy hole and lh for light hole) and by the quantum number n . All energies are in meV.

State	Finite well	Infinite well
e1	32	57
e2	120	227
e3	247	510
hh1	7	11
hh2	30	44
hh3	66	100
hh4	112	177
lh1	21	40
lh2	78	160

Example 6.1

Calculate the energy of the first electron bound state in a GaAs/AlGaAs quantum well with $d = 10 \text{ nm}$ and $V_0 = 0.3 \text{ eV}$. Take $m_w^* = 0.067m_0$ and $m_b^* = 0.092m_0$. Compare this value to the one calculated for an infinite quantum well.

Solution

The first bound state has a maximum at $z = 0$, and so we look for the solutions with cosine wave function in the well region. By making the substitution $x = kd/2$, we can use eqns 6.16 and 6.19 to recast eqn 6.24 in the form:

$$\tan x = \left(\frac{m_w^*}{m_b^*} \right)^{1/2} \frac{\sqrt{\xi - x^2}}{x}, \quad (6.26)$$

where

$$\xi = \frac{m_w^* d^2 V_0}{2\hbar^2}, \quad (6.27)$$

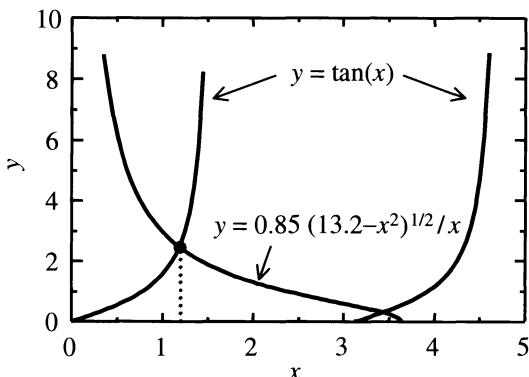


Fig. 6.4 Graphical solution of eqn 6.26 for the parameters given in Example 6.1.

and

$$E = \frac{2\hbar^2 x^2}{m_w^* d^2}. \quad (6.28)$$

Our task is thus to solve eqn 6.26, with

$$\left(\frac{m_w^*}{m_b^*}\right)^{1/2} = \left(\frac{0.067}{0.092}\right)^{1/2} = 0.85,$$

and

$$\xi = \frac{0.067 m_0 \times (10^{-8})^2 \times 0.30 \text{ eV}}{2\hbar^2} = 13.2.$$

Figure 6.4 plots the functions $y = \tan x$ and $y = 0.85\sqrt{13.2 - x^2}/x$ on the same scales, and shows that the first value of x where the two functions are the same is $x = 1.18$. Hence from eqn 6.28 we find the required bound state energy:

$$E = \frac{2\hbar^2 (1.18)^2}{0.067 m_0 \times (10^{-8})^2} = 31.5 \text{ meV}.$$

This value of E can be compared to that given by eqn 6.13 for an infinite well:

$$E_1 = \frac{\hbar^2 \pi^2 (1)^2}{2 \times 0.067 m_0 \times (10^{-8})^2} = 57 \text{ meV}.$$

The infinite well model thus overestimates the energy of the bound state by a factor of 1.8.

6.4 Optical absorption and excitons

In Sections 3.2 and 3.3 of Chapter 3 we used Fermi's golden rule to calculate the absorption spectrum of a bulk semiconductor. Then in Section 4.2 of Chapter 4 we studied how the spectrum is altered by excitonic effects. We now follow a similar approach for quantum wells, beginning with the selection rules and density of states for the optical transitions, and then moving on to consider the excitonic effects.

6.4.1 Selection rules

We consider a quantum well irradiated by light of angular frequency ω propagating in the z direction, as shown in Fig. 6.5. The photons are absorbed by exciting electrons from an initial state $|i\rangle$ at energy E_i in the valence band to a final state $|f\rangle$ at energy E_f in the conduction band. Conservation of energy requires that $E_f = (E_i + \hbar\omega)$.

Fermi's golden rule tells us that the absorption rate is determined by the density of states and the square of the electric dipole matrix element. (See Section B.2 in Appendix B.) The transition rate can be calculated by combining eqns 3.2, 3.3 and 3.6 to obtain:

$$W_{i \rightarrow f} = \frac{2\pi}{\hbar} |\langle f | -e\mathbf{r} \cdot \boldsymbol{\epsilon} | i \rangle|^2 g(\hbar\omega), \quad (6.29)$$

where \mathbf{r} is the position vector of the electron, $\boldsymbol{\epsilon}$ is the electric field amplitude of the light wave, and $g(\hbar\omega)$ is the density of states. We have simplified the form of the electric dipole perturbation here by setting the $e^{\pm i\mathbf{k}\cdot\mathbf{r}}$ factor of the light in eqn 3.6 equal to unity. As discussed in connection with eqn 3.12, this approximation is justified because the photon wave vector is negligible in comparison to that of the electron.

We first consider the matrix element for the transition. This will allow us to work out important selection rules. With photons incident in the z direction as shown in Fig. 6.5 the polarization vector of the light is in the x, y plane. We therefore have to evaluate matrix elements of the form

$$M = \langle f | x | i \rangle = \int \Psi_f^*(\mathbf{r}) x \Psi_i(\mathbf{r}) d^3\mathbf{r}. \quad (6.30)$$

When we considered matrix elements of this type in Section 3.2, it made no difference whether we evaluated $\langle f | x | i \rangle$ or $\langle f | y | i \rangle$ or $\langle f | z | i \rangle$. This was a consequence of the isotropy of the cubic semiconductors we were considering. In the case of the quantum well, however, the x and y directions are equivalent, but the z direction is physically different. Therefore, for quantum wells we will have:

$$\langle f | x | i \rangle = \langle f | y | i \rangle \neq \langle f | z | i \rangle. \quad (6.31)$$

In this section we will concentrate on x, y polarized light, which is the usual experimental arrangement.

We are interested in evaluating eqn 6.30 for transitions between bound quantum well states in the valence and conduction bands. Figure 6.6 illustrates the type of transition we are considering. The figure specifically shows a transition from an $n = 1$ hole level to an $n = 1$ electron level, and from an $n = 2$ hole level to an $n = 2$ electron level.

We consider a general transition from the n th hole state to the n' th electron state. In analogy to the Bloch functions of eqns 3.7 and 3.8, we can use eqns 6.5 and 6.7 to write the initial and final quantum well wave functions in the form:

$$\Psi_i \equiv |i\rangle = \frac{1}{\sqrt{V}} u_v(\mathbf{r}) \varphi_{hn}(z) e^{i\mathbf{k}_{xy} \cdot \mathbf{r}_{xy}} \quad (6.32)$$

$$\Psi_f \equiv |f\rangle = \frac{1}{\sqrt{V}} u_c(\mathbf{r}) \varphi_{en'}(z) e^{i\mathbf{k}'_{xy} \cdot \mathbf{r}_{xy}}. \quad (6.33)$$

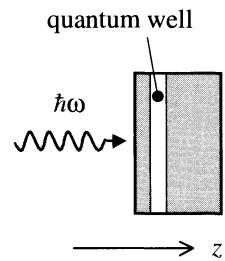


Fig. 6.5 Photons incident on a quantum well with light propagating in the z direction.

We are using Dirac notation here. See the margin comment on Section 3.2.

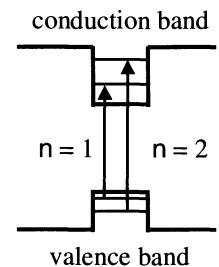


Fig. 6.6 Interband optical transitions in a quantum well. The $n = 1$ and $n = 2$ transitions are indicated.

The three factors in these wave functions denote the envelope function for the valence or conduction band as appropriate, the bound states of the quantum well in the z direction, and the plane waves for the free motion in the x , y plane. We have written explicit subscripts to show that the plane waves only span the 2-D x , y coordinates.

The momentum of the photon is very small in comparison to that of the electrons, and so conservation of momentum in the transition requires that $\mathbf{k}_{xy} = \mathbf{k}'_{xy}$. This is the two-dimensional equivalent of eqn 3.12 for 3-D bulk semiconductors. Therefore, on substituting eqns 6.32 and 6.33 into eqn 6.30, we see that the matrix element breaks into two factors:

$$M = M_{cv} M_{nn'} \quad (6.34)$$

where M_{cv} is the valence–conduction band dipole moment:

$$M_{cv} = \langle u_c | x | u_v \rangle = \int u_c^*(\mathbf{r}) x u_v(\mathbf{r}) d^3\mathbf{r}, \quad (6.35)$$

and $M_{nn'}$ is the electron–hole overlap given by

$$M_{nn'} = \langle e n' | h n \rangle = \int_{-\infty}^{+\infty} \varphi_{en'}^*(z) \varphi_{hn}(z) dz. \quad (6.36)$$

It will usually be the case that the constituent material of the quantum well (e.g. GaAs) has strongly allowed electric dipole transitions between the conduction and valence bands. We considered this point in Section 3.3.1. Therefore, we can assume that M_{cv} is non-zero. Hence the matrix element for the optical transitions is proportional to the overlap of the electron and hole states given by eqn 6.36. This allows us to work out some straightforward selection rules on $\Delta n = n' - n$.

Consider first an infinite quantum well with wave functions of the form given by eqn 6.11. The overlap factor is

$$M_{nn'} = \frac{2}{d} \int_{-d/2}^{+d/2} \sin\left(k_n z + \frac{n\pi}{2}\right) \sin\left(k_{n'} z + \frac{n'\pi}{2}\right) dz. \quad (6.37)$$

This is unity if $n = n'$ and zero otherwise. (See Exercise 6.6.) Hence we obtain the following selection rule for an infinite quantum well:

$$\Delta n = 0. \quad (6.38)$$

This is why we only showed $\Delta n = 0$ transitions in Fig. 6.6.

In finite quantum wells the electron and hole wave functions with differing quantum numbers are not necessarily orthogonal to each other because of the differing decay constants in the barrier regions. This means that there are small departures from the selection rule of eqn 6.38. However these $\Delta n \neq 0$ transitions are usually weak, and are strictly forbidden if Δn is an odd number, because the overlap of states with opposite parities is zero. (See Exercise 6.6.)

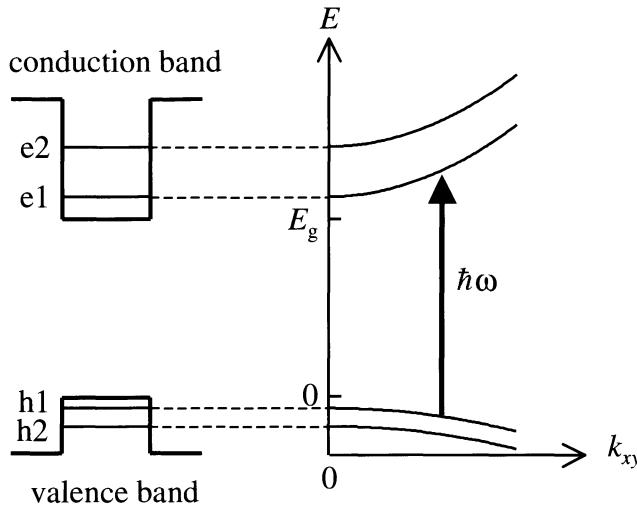


Fig. 6.7 The $n = 1$ interband optical transition in a quantum well at finite k_{xy} .

6.4.2 Two-dimensional absorption

The shape of the absorption spectrum in a quantum well can be understood by applying the selection rules we have just derived. If we increase the photon energy from zero, no transitions will be possible until we cross the threshold for exciting electrons from the ground state of the valence band (the $n = 1$ heavy hole level) to the lowest conduction band state (the $n' = 1$ electron level). This is a $\Delta n = 0$ transition and is therefore allowed. This threshold occurs at a photon energy given by

$$\hbar\omega = E_g + E_{hh1} + E_{e1}, \quad (6.39)$$

where E_g is the band gap of the quantum well material. This immediately gives us a very important result. The optical absorption edge of the quantum well has been shifted by $(E_{hh1} + E_{e1})$ compared to the bulk semiconductor. Since the confinement energies can be varied by choice of the well width, this gives us a way to tune the frequency of the absorption edge.

The right hand side of Fig. 6.7 shows the $E-k_{xy}$ diagram for the transition between the $n = 1$ levels. The bands have parabolic dispersions according to eqn 6.9. Conservation of momentum and the negligible \mathbf{k} vector of the photon imply that the electron and hole states have the same k_{xy} values. The energy of the transition shown by the vertical arrow is given by:

$$\begin{aligned} \hbar\omega &= E_g + \left(E_{hh1} + \frac{\hbar^2 k_{xy}^2}{2m_{hh}^*} \right) + \left(E_{e1} + \frac{\hbar^2 k_{xy}^2}{2m_e^*} \right) \\ &= E_g + E_{hh1} + E_{e1} + \frac{\hbar^2 k_{xy}^2}{2\mu}, \end{aligned} \quad (6.40)$$

where μ is the electron–hole reduced effective mass defined in eqn 3.22. This makes it clear that the transitions with $\hbar\omega = (E_g + E_{hh1} + E_{e1})$ occur at $k_{xy} = 0$.

Equation 6.40 can be compared directly to eqn 3.23 for the bulk semiconductor. We have already noted the shift of the absorption threshold from E_g

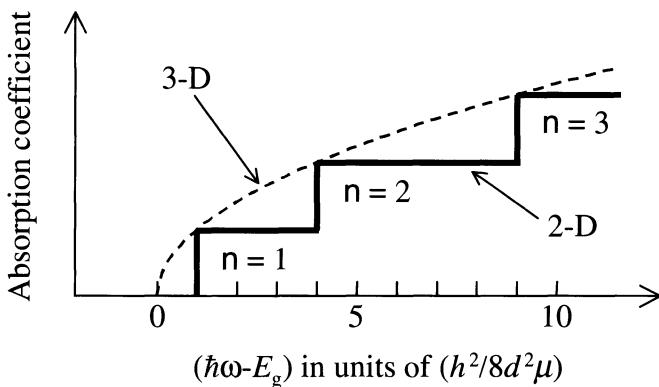


Fig. 6.8 The absorption coefficient for an infinite quantum well of width d compared to the equivalent bulk semiconductor. μ is the electron–hole reduced mass. Excitonic effects are ignored.

to $(E_g + E_{hh1} + E_{e1})$. The other crucial difference is that the wave vector in eqn 6.40 for the quantum well spans only the 2-D x, y coordinates, instead of the full 3-D x, y, z space. This has a very important consequence for the joint density of states factor that enters the transition rate in eqn 6.29. The 3-D bulk semiconductor had a parabolic density of states given by eqn 3.16, which led to the absorption edge given in eqn 3.25. By contrast, the joint density of states for a 2-D material is independent of energy and is given by (see Exercise 6.3):

$$g(E) = \frac{\mu}{\pi\hbar^2}. \quad (6.41)$$

This means that the absorption coefficient will have a step-like structure, being zero up to the threshold energy given in eqn 6.39, and then having a constant non-zero value for larger photon energies.

The argument above can be repeated for the other allowed optical transitions in the quantum well. The next strong $\Delta n = 0$ transition for the heavy hole states occurs at an energy of $(E_g + E_{hh2} + E_{e2})$ which corresponds to exciting an electron from the $n = 2$ heavy hole state to the $n' = 2$ electron level. Once the photon energy crosses this threshold, the absorption coefficient will show a new step. There will also be other steps corresponding to transitions from the light hole states to the conduction band.

The functional form of the absorption coefficient for an infinite quantum well is shown in Fig. 6.8. The confinement energies of the electron and hole states are given by eqn 6.13, and the $\Delta n = 0$ selection rule is strictly obeyed. The threshold energy for the n th transition is thus given by:

$$\hbar\omega = E_g + \frac{\hbar^2 n^2}{2\pi^2 m_e^* d^2} + \frac{\hbar^2 n^2}{2\pi^2 m_h^* d^2} = E_g + \frac{\hbar^2 n^2}{2\pi^2 \mu d^2}. \quad (6.42)$$

The spectrum therefore consists of a series of steps with threshold energies given by eqn 6.42. For comparison, the energy-dependence of the absorption coefficient for the equivalent bulk semiconductor is plotted on the same scale. The shift of the absorption edge by the confinement energy is evident, together with the change of shape from the parabola of the bulk semiconductor to the step-like structure for the quantum well. In essence, this simply reflects the change in the density of states on going from 3-D to 2-D.

Example 6.2

Estimate the difference in the wavelength of the absorption edge of a 20 nm GaAs quantum well and bulk GaAs at 300 K.

Solution

We see from eqn 6.39 that the absorption edge of a quantum well occurs at $E_g + E_{hh1} + E_{el}$. We can estimate the confinement energies by using the infinite potential well model. Using eqn 6.13 and the effective mass data for GaAs given in Table C.2, we find that $E_{hh1} = 2 \text{ meV}$ and $E_{el} = 14 \text{ meV}$. These energies are small compared to typical quantum well barrier heights, and so the infinite well approximation is going to be reasonably accurate. The band edge therefore shifts from 1.424 eV to $(1.424 + 0.002 + 0.014) = 1.440 \text{ eV}$. This corresponds to a blue shift of 10 nm.

6.4.3 Experimental data

Figure 6.9 shows the absorption spectrum of a high quality GaAs MQW structure containing 40 quantum wells of width 7.6 nm. The barriers were made of AlAs, and the sample temperature was 6 K. It is clear that the predicted step-like behaviour shown in Fig. 6.8 is well reproduced in the data, although the experimental spectrum is complicated by excitonic effects, which give rise to the strong peaks in the absorption at the edge of each step. These excitonic effects will be discussed further in Section 6.4.4, and we concentrate for now on the gross features of the absorption spectrum.

The most pronounced steps in the spectrum are due to the $\Delta n = 0$ transitions. The first of these occurs for the $n = 1$ heavy hole transition at 1.59 eV. This is closely followed by the step due to the $n = 1$ light hole transition at 1.61 eV. This should be compared with the low temperature band edge

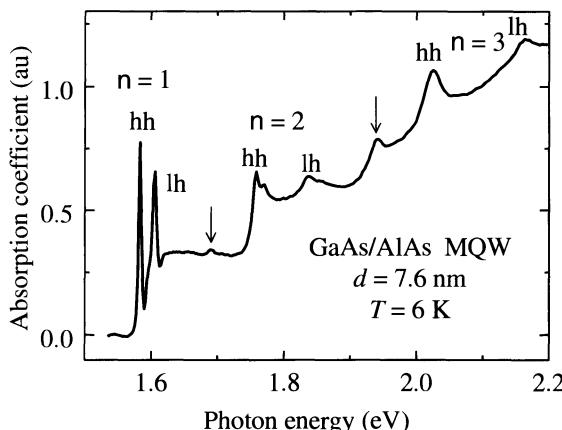


Fig. 6.9 Absorption coefficient of a 40 period GaAs/AlAs MQW structure with 7.6 nm quantum wells at 6 K. After [1], copyright 1996 Taylor & Francis Ltd., reprinted with permission.

The two weak peaks identified by arrows are caused by parity-conserving $\Delta n \neq 0$ transitions. The one at 1.69 eV is the $hh3 \rightarrow e1$ transition, while that at 1.94 eV is the $hh1 \rightarrow e3$ transition.

absorption spectra of bulk GaAs shown in Figs 4.3 and 4.4. We see that the band edge has been shifted in the quantum well by 0.07 eV.

The steps at the band edge are followed by a flat spectrum up to 1.74 eV in which the absorption is practically independent of energy. At 1.77 eV there is a further step in the spectrum due to the onset of the $n = 2$ heavy hole transition. This is followed by the step for the $n = 2$ light hole transition at 1.85 eV. Further steps due to the $n = 3$ heavy and light transitions are observed at 2.03 eV and 2.16 eV respectively.

6.4.4 Excitons in quantum wells

We now return to consider the excitonic effects that give rise to the sharp peaks that are very prominent in the experimental data shown in Fig. 6.9. As discussed in Chapter 4, excitons are bound electron–hole pairs held together by their mutual Coulomb attraction. Since the optical transition can be considered as the creation of an electron–hole pair, the Coulomb attraction increases the absorption rate because it enhances the probability of forming the electron–hole pair. Hence we observe peaks at the resonant energies for exciton formation. These peaks occur at the sum of the single particle energies less the binding energy of the bound pair. Detailed analysis of the data shown in Fig. 6.9 reveals that the binding energies of the quantum well excitons are about 10 meV. This is substantially higher than the value of 4.2 meV in bulk GaAs. (See Section 4.2.)

The enhancement of the excitonic binding energy in the quantum well is a consequence of the quantum confinement of the electrons and holes. This forces the electrons and holes to be closer together than they would be in a bulk semiconductor, and hence increases the attractive potential. It is possible to show that the binding energy of the ground state exciton in an ideal 2-D system is enhanced by a factor of four compared to the bulk material (see Exercise 6.9). This should be compared with the factor of ~ 2.5 deduced from the experimental data. Although we do not observe perfect 2-D enhancement of the binding energy, the increase is still substantial. The enhancement of the excitonic effects in quantum wells is very useful for device applications, as we will discuss further in the next section.

The reason why the exciton binding energy is not enhanced by a factor of four is that a real quantum well is not a perfect 2-D system. The quantum well has a finite width, and the wave functions actually extend into the barriers due to tunnelling.

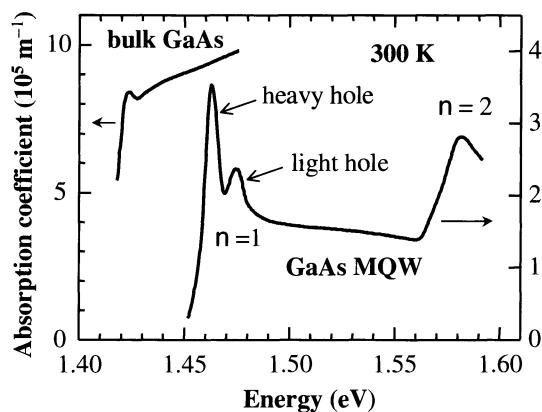


Fig. 6.10 Room temperature absorption spectrum of a GaAs/Al_{0.28}Ga_{0.72}As MQW structure containing 77 GaAs quantum wells of width 10 nm. The absorption spectrum of bulk GaAs at the same temperature is shown for comparison. After [2], copyright 1982 American Institute of Physics, reprinted with permission.

One of the most useful consequences of the enhancement of the exciton binding energy in quantum wells is that the excitons are still stable at room temperature. This contrasts with bulk GaAs, which only shows strong excitonic effects at low temperatures. This can be clearly seen in the data shown in Fig. 6.10, which compares the absorption coefficient of a GaAs MQW structure with 10 nm quantum wells to that of bulk GaAs at room temperature. The bulk sample merely shows a weak shoulder at the band edge, but the MQW shows strong peaks for both the heavy hole and the light hole excitons. The more or less flat absorption coefficient expected for quantum wells above these peaks is also evident.

A striking difference between the quantum well and bulk absorption shown in Fig. 6.10 is the lifting of the degeneracy of the heavy and light hole states. The bulk sample shows a single excitonic shoulder at the band edge, but the quantum well shows two separate peaks. As discussed in point (2) of Section 6.3, this follows from the different effective masses of the heavy and light holes, and it highlights the lower symmetry of the quantum well sample.

6.5 The quantum confined Stark effect

In Section 4.3.1 of Chapter 4 we considered the effects of a DC electric field on the excitons in bulk GaAs. We found that relatively small electric fields can ionize the excitons by pushing the electrons and holes in opposite directions. The situation in quantum wells is different if the field is applied along the z direction. The field still pushes the electrons and holes in opposite directions, but the barriers prevent the exciton from breaking apart. Hence the excitons are stable up to very high field strengths. These quantum confined excitons interact with the field and shift to lower energy. In analogy to the Stark effect in atomic physics, this shift is known as the **quantum confined Stark effect**.

The shift of the quantum well energy levels with the electric field can be calculated by perturbation theory. This can be done analytically in the case of quantum wells with infinite barriers. (See Exercise 6.12.) For small fields, the shift to the $n = 1$ level is given by:

$$\Delta E = -24 \left(\frac{2}{3\pi} \right)^6 \frac{e^2 \mathcal{E}_z^2 m^* d^4}{\hbar^2}, \quad (6.43)$$

where d is the well width, and \mathcal{E}_z is the component of the field in the z direction. This result is analogous to the quadratic Stark effect in atomic hydrogen: the levels shift to lower energy in proportion to $-\mathcal{E}_z^2$.

The quadratic red-shift of the levels can be understood as follows. The shift in energy is given by $\Delta E = -p_z \mathcal{E}_z$, where p_z is the electron dipole along the field direction. In the absence of the field, the $n = 1$ electron wave function is symmetric about the centre of the well at $z = 0$. Hence, the expectation value of z for the electron, namely $\langle z \rangle = \int \varphi^* z \varphi dz$, is zero. Thus $p_z = -e\langle z \rangle$ is also zero, and the field must induce a dipole to produce a shift in the levels. If the field is applied in the positive z direction, the electrons will be pushed towards negative z , and $\langle z \rangle$ acquires a negative value. This creates a positive dipole. At small fields, the magnitude of this induced dipole will be proportional to \mathcal{E}_z . Hence $p_z \propto +\mathcal{E}_z$ and $\Delta E \propto -\mathcal{E}_z^2$. The same argument can be applied to the hole states.

To a first approximation, the energy shift of the quantum well excitons will just be given by the sum of the shifts of the electron and hole levels. Equation 6.43 shows that it is the hole states that contribute the most, due to their larger effective mass.

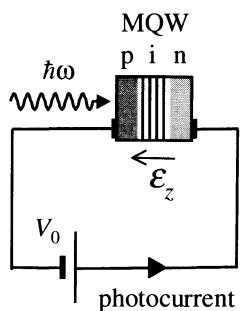


Fig. 6.11 Experimental set-up required to observe the quantum confined Stark effect. The quantum wells are grown in the i-region of a p-i-n diode, and the device is operated in reverse bias. This produces a strong electric field E_z across the quantum wells. The photocurrent generated when light is incident is determined by the absorption of the MQW layer.

The parity-forbidden transitions with Δn equal to an odd number become allowed when the field is applied due to the lowering of the symmetry of the system. (See Exercise 6.15.)

The quantum confined Stark effect can be observed by growing the quantum wells in the i-region of a p-i-n diode, as shown in Fig. 6.11. By operating the diode in reverse bias like a photodiode, strong DC electric fields can be applied in the growth direction. The magnitude of the field is given by eqn D.3 in Appendix D:

$$E_z = \frac{|V_{bi} - V_0|}{l_i}, \quad (6.44)$$

where V_{bi} is the built-in voltage of the diode, V_0 is the applied bias, and l_i is the i-region thickness. V_0 is negative in reverse bias, and thus the applied voltage augments the field due to the built-in voltage. As discussed in Section 4.3.1, the photocurrent generated in the device follows the frequency dependence of the absorption.

Figure 6.12 shows the photocurrent spectra of a GaAs/Al_{0.3}Ga_{0.7}As MQW p-i-n diode at bias voltages of 0 V and -10 V. The well width was 9.0 nm, and the temperature was 300 K. The i-region thickness was 1 μ m, and V_{bi} was 1.5 V. From eqn 6.44, we see that the voltages correspond to field strengths of $1.5 \times 10^6 \text{ V m}^{-1}$ and $1.15 \times 10^7 \text{ V m}^{-1}$ respectively. The $n = 1$ heavy and light hole exciton lines are clearly resolved at both field strengths, despite the fact that the temperature is 300 K and that E_z considerably exceeds the exciton ionization field of bulk GaAs, which is of order $6 \times 10^5 \text{ V m}^{-1}$.

The spectrum at -10 V shows a clear red-shift for both the heavy and light hole excitons. As expected, the shift is larger for the heavy hole excitons due to their larger mass. Since the electrons and holes shift in opposite directions, their overlap integral (eqn 6.36) decreases. This accounts for the reduction in the exciton absorption at the higher field strength. Two parity-forbidden transitions are clearly identified in Fig. 6.12, namely the hh2 \rightarrow e1 and hh1 \rightarrow e2 lines.

The ability to control the shape of the absorption spectrum by applying bias opens the possibility for making different types of optoelectronic devices. The

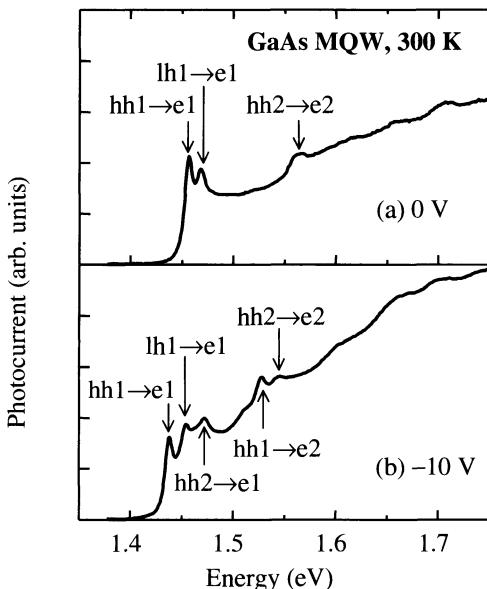


Fig. 6.12 Photocurrent spectra for a GaAs/Al_{0.3}Ga_{0.7}As MQW p-i-n diode with a 1 μ m thick i-region at room temperature. (a) $V_0 = 0$, (b) $V_0 = -10$ V. The quantum well thickness was 9.0 nm. The transitions are labelled by the electron and hole states that participate. After [1], copyright 1996 Taylor & Francis Ltd., reprinted with permission.

absorption at 1.44 eV (864 nm) in the quantum wells studied in Fig. 6.12 can be switched on and off by applying the bias. This allows us to make a voltage-tunable photodetector using the same arrangement as that shown in Fig. 6.11. The same device can also function as an intensity modulator by introducing a voltage-dependent loss on a 864 nm beam transmitted through the quantum wells. Further details of these optoelectronic devices may be found in the bibliography given in the further reading section.

6.6 Optical emission

The use of quantum well structures in electroluminescent devices is their main commercial application at present. As we will see, the insertion of quantum wells into the active region allows a greater range of emission wavelengths to be obtained, and also enhances the efficiency of the devices.

The general principles of light emission in semiconductors were discussed in Chapter 5. The light is generated when electrons in the conduction band recombine with holes in the valence band. We saw in Sections 5.2–5.4 that the luminescence spectrum generally consists of a peak at the band gap energy with a width determined by the carrier density and the temperature.

The physical processes responsible for light emission in quantum wells are essentially the same as those in bulk semiconductors. The electrons and holes injected electrically or optically rapidly relax to the bottom of their bands before emitting photons by radiative recombination. In a quantum well, the lowest levels available to the electrons and hole correspond to the $n = 1$ confined states. Hence the low intensity luminescence spectrum consists of a peak of spectral width $\sim k_B T$ at energy

$$h\nu = E_g + E_{hh1} + E_{e1}. \quad (6.45)$$

This shows that the emission peak is shifted by the quantum confinement of the electrons and holes to higher energy compared to the bulk semiconductor.

Figure 6.13 shows the photoluminescence spectrum of a 2.5 nm ZnCdSe quantum well with a cadmium concentration of 20 %. $Zn_{0.8}Cd_{0.2}Se$ is a II–VI alloy semiconductor with a direct band gap of 2.55 eV at 10 K, which corresponds to the blue/green spectral region. The barriers of the quantum

The shape of the emission spectrum is only slightly affected by reducing the dimensionality of the system from 3-D to 2-D. The low intensity emission spectrum of a bulk semiconductor is given by eqn 5.12. In a quantum well, the $(h\nu - E_g)^{1/2}$ factor from the 3-D density of states will be replaced by the unit step function derived from the 2-D density of states. In both 3-D and 2-D the net result is that we get a peak of width $\sim k_B T$ starting at the threshold energy for absorption.

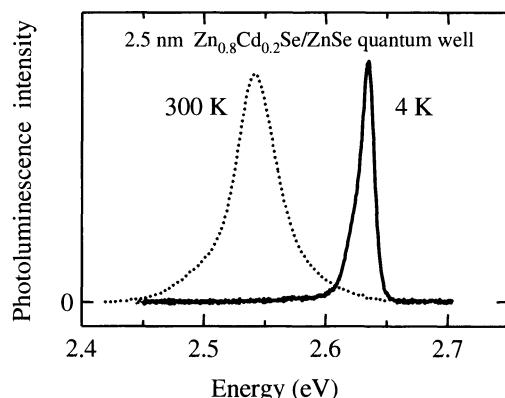


Fig. 6.13 Emission spectrum of a 2.5 nm $Zn_{0.8}Cd_{0.2}Se/ZnSe$ quantum well at 10 K and at room temperature. The spectra have been normalized so that their peak heights are the same. (After C.J. Stevens and R.A. Taylor, Personal communication)

wells are made from ZnSe, which has a band gap of 2.82 eV. The spectrum at 10 K peaks at 2.64 eV (470 nm) and has a full width at half maximum of 16 meV. The emission energy is about 0.1 eV larger than the band gap of the bulk material, and the line width is limited by the inevitable fluctuations in the well width that occur during the epitaxial growth. (See Exercise 6.16.) At room temperature the peak emission energy has shifted to 2.55 eV (486 nm) and the spectrum has broadened so that the line width is about 48 meV ($\sim 2k_B T$).

As mentioned at the start of this section, the use of quantum wells in light emitting devices is one of the main motivations for their development. Quantum wells offer three main advantages over the equivalent bulk materials:

- The shift of the luminescence peak by the confinement energy ($E_{e1} + E_{hh1}$) allows the wavelength of light emitting devices to be tuned by choice of the well width.
- The increased overlap between the electron and hole wave functions in the quantum well means that the emission probability is higher. This shortens the radiative lifetime, and the radiative recombination wins out over competing non-radiative decay mechanisms. The radiative efficiency is therefore higher in the quantum wells, which makes it easier to make bright light emitting devices.
- The total thickness of the quantum wells in an electroluminescent device is very small (~ 10 nm). This is well below the critical thickness for dislocation formation in non-lattice-matched epitaxial layers. This allows the use of non-lattice-matched combinations of materials, and hence gives even greater flexibility in emission wavelengths that can be obtained.

Electroluminescent devices can easily be made from quantum wells by incorporating them in the active region at the junction of a p-n diode, as discussed in Section 5.4 for bulk materials. The devices are operated in forward bias, and the light is emitted when the electrons and holes injected by the current recombine at the junction. GaAs quantum wells emitting around 800 nm are widely used as the lasers in compact disc players and printers. GaAs-based alloys are used to shift the wavelength into the red spectral region or into the infrared to match the optimal wavelengths for optical fibre systems at $1.3\ \mu\text{m}$ and $1.55\ \mu\text{m}$. (See Exercise 6.17.)

Example 6.3

Estimate the emission wavelength of a 15 nm GaAs quantum well laser at 300 K.

Solution

The emission wavelength is given by eqn 6.45. We estimate the confinement energies from eqn 6.13. Using the effective mass data given in Table C.2, we find $E_{hh1} = 3\ \text{meV}$ and $E_{e1} = 25\ \text{meV}$. The emission energy is therefore $1.424 + 0.003 + 0.025 = 1.452\ \text{eV}$, which corresponds to a wavelength of 854 nm.

6.7 Intersubband transitions

An **intersubband transition** is one in which we excite electrons and holes between the levels (or ‘subbands’) within the conduction or valence band. This contrasts with the interband transitions that we have been considering up till now in which the electrons move from the valence band to the conduction band and *vice versa*. Figure 6.14 illustrates a typical intersubband absorption transition in which an electron in the $n = 1$ level of a quantum well is excited to the $n = 2$ level by absorption of a photon.

A quick glance at Table 6.2 tells us that intersubband transitions occur at much lower photon energies than interband transitions. For example, the energy spacing between the $n = 1$ and $n = 2$ electron levels in a 10 nm GaAs/AlGaAs quantum well is of order 0.1 eV. This corresponds to an infrared wavelength of about 12 μm . We can therefore use intersubband transitions to make detectors and emitters for the infrared spectral region using GaAs quantum wells. This offers considerable advantages over narrow gap semiconductors because better performance can be obtained and it is also easier to fabricate the devices.

Intersubband transitions are excited by light polarized along the z direction. The matrix element for such a transition from the n th to the n' th subband is $\langle n|z|n' \rangle$, and the selection rule on $\Delta n = (n - n')$ is that Δn must be an odd number: see Exercise 6.18.

The requirement that the polarization must be along the z direction creates some technical difficulties. If the light is incident normal to the surface as shown in Fig. 6.5, there is no polarization component along the z direction. To create a z component, the light must be incident at an angle. However, it is only possible to couple about 10 % of the light to the intersubband transitions in this way, due to the high refractive index of semiconductors like GaAs. (See Exercise 6.19.) A better solution is to incorporate a metallic grating on the top of the sample. This can produce a substantial z component even for light incident normal to the surface.

Intersubband detectors have been under development since the late 1980s. They are made with n-type doped quantum wells so that there is a large population of electrons in the $n = 1$ level of the conduction band to absorb the light. In 1994 an intersubband laser called the ‘quantum cascade laser’ was reported. These intersubband devices work at wavelengths in the infrared, and are required for many important applications such as atmospheric sensing and pollution monitoring.

6.8 Bloch oscillators

In Section 6.2 we mentioned that Esaki and Tsu’s original paper considered the possibility of observing Bloch oscillations using superlattices. It is a striking predication of quantum mechanics that if a DC electric field \mathcal{E} is applied to a periodic structure with period L , the electrons oscillate with a frequency given by

$$\nu = \frac{e\mathcal{E}L}{h}. \quad (6.46)$$

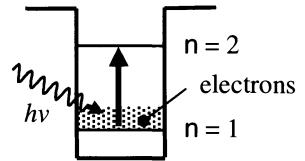


Fig. 6.14 The $e_1 \rightarrow e_2$ intersubband transition in an n-doped quantum well.

This phenomenon is known as **Bloch oscillations**. It has never been observed in normal crystals because the periodicity is too small.

Consider a typical bulk crystal with a lattice constant of 0.5 nm. Equation 6.46 predicts that the oscillation period is 8 ps at a high field strength of 10^6 V m^{-1} . This has to be compared to the time taken for an accelerating electron to scatter from a phonon, which is usually in the range 0.1–1 ps. We thus see that not even a small part of one period is completed.

Now consider a typical semiconductor superlattice, in which the repeated quantum wells form an artificial periodic structure with period $L = d + b$. The value of $(d + b)$ will be $\sim 10 \text{ nm}$, so that the oscillation period will be ~ 20 times smaller than the bulk crystals. This gives the electron a fighting chance of completing the oscillations before emitting a phonon.

In fact, it took more than 20 years of research to develop superlattices of sufficient purity to be able to observe Bloch oscillations. The breakthrough eventually came in 1993. The new Bloch oscillator devices have attracted much attention because the oscillating electrons act like dipole oscillators, and therefore emit electromagnetic waves at the Bloch frequency. The devices therefore offer the potential to develop a whole new class of high frequency solid state oscillators for the emission of photons in the far-infrared spectral region.

6.9 Quantum dots

The most obvious way to make quantum dots is by etching very small pillars on the surface of a quantum well wafer by lithography. The $\sim 10 \text{ nm}$ length scale required makes this approach very difficult and expensive with present technologies. The fact that some materials spontaneously form quantum dots bypasses the need to use these complicated lithographic methods.

We mentioned in Section 6.1 that other types of quantum confined semiconductor structures can be made in addition to the quantum wells that we have been concentrating on so far. With reference to Table 6.1, we see that if we confine the electrons in all three directions, we have a quantum dot structure. These quantum dot structures can be made by complicated lithographic techniques, but they also form spontaneously in certain materials. We will give two examples of this below.

A quantum dot structure may be considered as a three-dimensional quantum well, with no degrees of freedom at all. We thus have ‘0-dimensional’ properties, with quantized levels for all three directions of motion. If we have a rectangular dot with dimensions (d_x, d_y, d_z) , the energy levels of the system will be given by:

$$E(n_x, n_y, n_z) = \frac{\pi^2 \hbar^2}{2m^*} \left(\frac{n_x^2}{d_x^2} + \frac{n_y^2}{d_y^2} + \frac{n_z^2}{d_z^2} \right), \quad (6.47)$$

where the quantum numbers n_x , n_y and n_z specify the quantized levels in each direction. We have assumed here that we have infinite barriers in all three directions, so that we can use eqn 6.13 separately for each direction. The energy spectrum of the structures is therefore completely discrete, analogous to the energy spectrum of atoms. The difference is that we can ‘tune’ the position of the energy levels by altering the size of the quantum dot.

The change of the density of states of an electron in the conduction band on reducing the dimensionality of a system is illustrated in Fig. 6.15. The dashed line shows the $(E - E_g)^{1/2}$ dependence of the bulk material. The thin solid line corresponds to a quantum well with the characteristic step-like density of states

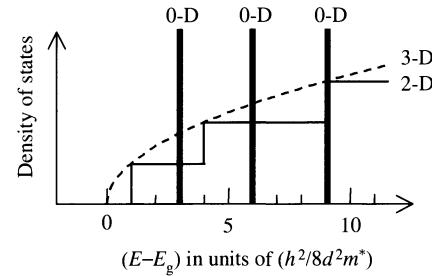


Fig. 6.15 Variation of the electron density of states with dimensionality. The dashed line is for a bulk semiconductor with a band gap of E_g . The thin solid line is for a quantum well of width d with infinite barriers. The thick solid lines are for a cubic quantum dot of dimension d with infinite barriers.

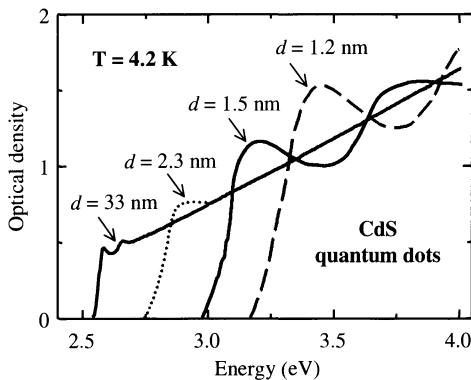


Fig. 6.16 Absorption spectra of glasses with CdS microcrystals of varying sizes at 4.2 K. The sample with $d = 33\text{ nm}$ effectively represents the properties of bulk CdS. After [3], reprinted with permission.

of 2-D materials. The thick solid line shows the density of states for the 0-D quantum dot. The density of states is discrete, with a series of delta functions at the energies described by eqn 6.47. (See Exercise 6.20.)

We have seen that the quantum confinement of the electrons and holes in 2-D quantum well structures gives rise to a host of interesting optical effects and improvements in the performance of optoelectronic devices. It is thus to be expected that the increased degree of confinement in quantum dots should give further benefits. By confining the carriers in all three dimensions, we increase the electron–hole overlap and thus increase the radiative quantum efficiency. The discrete nature of the density of states also reduces the thermal spread of the carriers within their bands. Both these aspects are expected to reduce the threshold current of laser diodes and increase their quantum efficiency.

An attractive aspect of quantum dots is that they form spontaneously in certain materials. The two best-known examples are discussed briefly below.

6.9.1 Semiconductor doped glasses

We introduced the idea of semiconductor doped glasses in Section 1.4.5. II–VI semiconductors such as CdS, CdSe, ZnS and ZnSe are introduced into the glass during the melt process, forming very small microcrystals within the glass matrix. The dimensions of the microcrystals depend on the way the glass is produced, and with careful preparation it is possible to make quantum dots with good size uniformity.

Figure 6.16 shows the absorption spectra of CdS quantum dots in a glass matrix at 4.2 K. Spectra are shown for four different sizes of the microcrystals. Quantum size effects are expected when the length d of the crystal is less than

about 3.5 nm. Hence the sample with $d = 33$ nm is not expected to show any quantum size effects, and it may therefore be taken to represent the properties of bulk CdS. This is apparent from the fact that the observed band edge occurs at 2.58 eV, which coincides with the band gap of bulk CdS at this temperature.

The other three samples with $d = 2.3$ nm, 1.5 nm and 1.2 nm respectively all satisfy the criterion for quantum size effects. They show an increasing shift of the absorption edge to higher energy with decreasing dot size, as expected from eqn 6.47. The spectra also show a broad peak at the edge which is caused by the enhanced excitonic effects in these 0-D structures. The magnitude of the quantum size effect is very large, with a shift of over 0.5 eV for the sample with $d = 1.2$ nm. The shift of the band edge with decreasing dot size is a very clear demonstration of the quantum size effects that we have been discussing throughout this chapter.

The shift of the absorption with dot size shown in Fig. 6.16 provides a way to tune the transmission edge of colour glass filters. By using different dots (e.g. CdSe with a band gap of 1.75 eV at room temperature), it is possible to move the transmission edge across the entire visible spectral region by altering the dot size.

6.9.2 Self-organized III–V quantum dots

The formation of quantum dots in III–V materials by self-organized growth has been discovered relatively recently. The dots are typically formed when we try to grow layers of InAs on a GaAs substrate. By reference to Fig. 5.11, we see that there is a large mismatch between the lattice constant of the epitaxial layer and the substrate. In the right conditions, it is energetically advantageous for the InAs to form small clusters rather than a uniformly strained layer. The surface physics determines that the dimensions of these clusters is of order 10 nm, which provides excellent quantum confinement of the electrons and holes in all three directions. This subject is still in the early stages of its development at present, but the fact that the dots are directly formed during the epitaxial growth makes them a very attractive option for optoelectronic applications.

Chapter summary

- Quantum confinement occurs when the dimensions of the structure are small enough that the confinement energy is greater than the thermal energy at that temperature.
- A structure with confinement in one dimension is called a quantum well. Structures with confinement in two or three dimension are called quantum wires and quantum dots respectively.
- Semiconductor quantum wells are made by epitaxial growth of very thin layers. The quantum confinement arises from the potential barriers at the interfaces between different semiconductors due to their different band gaps. The electrons and holes exhibit two-dimensional physics.

- A multiple quantum well is a crystal containing many quantum wells that are separated from each other. A superlattice is a similar structure but with thinner barriers, so that adjacent wells are coupled together by tunnelling through the barriers.
- The energies of the confined states can be calculated by modelling the system as a one-dimensional potential well with a depth determined by the difference in the band gaps of the constituent semiconductors.
- The quantum confinement shifts the absorption edge to higher energy compared to the bulk semiconductor. The absorption spectrum is mainly determined by the 2-D density of states of the quantum well and consists of a series of steps.
- Excitonic effects are enhanced in quantum wells. Exciton absorption peaks are readily observed at room temperature in the absorption spectra.
- The quantum confined Stark effect is the shift of the quantum well levels induced by an electric field, which causes a red shift in the band edge and exciton energies. The effect can be used to make optical modulators.
- The emission energy for luminescence is larger than in a bulk semiconductor due to the quantum confinement of the electrons and holes. Quantum wells make bright light-emitting devices, and the emission wavelength can be tuned by choice of the quantum well parameters.
- Intersubband transitions occur when electrons are excited between the subbands of a quantum well by absorption of a photon. The transitions occur at infrared wavelengths. Far-infrared emission can occur in a superlattice when a DC electric field induces Bloch oscillations.
- Quantum dot structures form spontaneously in certain semiconductor doped glasses. They are also formed in III–V semiconductors during self-organized growth of lattice mismatched materials.

Further reading

The seminal paper on semiconductor quantum wells is Esaki and Tsu (1970). Complementary introductory reading to the treatment given here may be found in Burns (1985) or Singleton (2001). The subject is treated at a more thorough level in Yu and Cardona (1996), or in a number of specific quantum well texts, such as Bastard (1990), Jaros (1989), Kelly (1995), Singh (1993), or Weisbuch and Vinter (1991).

The Stark effect in hydrogen is discussed in most quantum mechanics texts, for example Gasiorowicz (1996) or Schiff (1969), and also in atomic physics texts, such as Woodgate (1980). The quantum confined Stark effect and the applications of quantum wells in optoelectronic devices is discussed in Chuang (1995) or Fox (1996).

Blood (1999) gives a review of the use of quantum wells in visible-emitting diode lasers. The physics of intersubband transitions is discussed in Helm (2000), Liu and Capasso (2000a), and Liu and Capasso (2000b). Helm also discusses Bloch oscillations, as does Grahn (1995). The physics of quantum dots is described by Bimberg *et al.* (1999), Harrison (1999), and Woggan (1997).

References

- [1] Fox, A.M. (1996). *Contemporary Physics*, **37**, 111.
- [2] Miller, D.A.B., Chemla, D.S., Eilenberger, D.J., Smith, P.W., Gossard, A.C. and Tsang, W.T. (1982). *Appl. Phys. Lett.*, **41**, 679.
- [3] Ekimov, A.I. and Efros, Al. L. (1991). *Acta Physica Polonica A*, **79**, no. 1, 5.

Exercises

- (6.1) Estimate the temperature at which quantum size effects would be important for a semiconductor layer of thickness $1\ \mu\text{m}$ if the effective mass of the electrons is $0.1m_0$.
- (6.2) A particle of mass m^* is confined to move in a quantum well with infinite barriers of width d . Show that the energy separation of the first two levels is equal to $\frac{1}{2}k_B T$ when d is equal to $d = (3\hbar^2/4m^*k_B T)^{1/2}$. Evaluate d for electrons of effective mass m_0 and $0.1m_0$ at 300 K. Hence show that this value of d is smaller than the value of Δx given in eqn 6.4 by a factor of $\sqrt{3\pi^2}$.
- (6.3) Consider a gas of spin $\frac{1}{2}$ particles of mass m moving in a two-dimensional layer. Apply Born-von Karman periodic boundary conditions (i.e. $e^{ikx} = e^{ik(x+L)}$, etc, where L is a macroscopic length) to show that the density of states in k -space is $1/(2\pi)^2$. By considering the incremental area enclosed by two circles in k -space differing in radius by dk , show that the number of states with k vectors between k and $k+dk$ is given by $g(k)dk = (k/2\pi)dk$. Hence show that if the energy dispersion is given by $E(k) = \hbar^2 k^2/2m$, the density of states in energy space is given by

$$g(E)dE = \frac{m}{\pi\hbar^2} dE.$$

- (6.4) Explain, with reference to eqn 6.26, why a finite quantum well always has at least one bound state, no matter how small V_0 is.
- (6.5) Calculate the energy of the first heavy hole bound state of a GaAs/AlGaAs quantum well with $d = 10\ \text{nm}$. Take

$V_0 = 0.15\ \text{eV}$, $m_w^* = 0.34m_0$ and $m_b^* = 0.5m_0$. How does this energy compare to that of an equivalent well with infinite barriers?

- (6.6) Consider the electron-hole overlap integral $M_{nn'}$ for a quantum well given by:
- $$M'_{nn} = \int_{-\infty}^{+\infty} \varphi_{en'}^*(z) \varphi_{hn}(z) dz.$$
- (i) Show that M'_{nn} is unity if $n = n'$ and zero otherwise in a quantum well with infinite barriers.
 - (ii) Show that M'_{nn} is zero if $(n - n')$ is an odd number in a quantum well with finite barriers.
 - (6.7) Draw a sketch of the energy dependence of the absorption spectrum of a 5 nm GaAs quantum well at 300 K between 1.4 eV and 2.0 eV. Assume that the confining barriers are infinite and ignore excitonic effects. See Table C.2 for band structure data on GaAs.
 - (6.8) Discuss how the spectrum in Exercise 6.7 would change if (i) the barrier height is finite, and (ii) excitonic effects are included.
 - (6.9)* The variational technique introduced in Exercise 4.4 can be used to calculate the energy and radius of a 2-D exciton.¹ The Hamiltonian for the relative motion of an electron-hole pair in a 2-D material is given in polar coordinates by

$$\hat{H} = -\frac{\hbar^2}{2\mu} \left(\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \right) - \frac{e^2}{4\pi\epsilon_0\epsilon_r r},$$

¹ The 2-D exciton problem can also be solved exactly using Laguerre polynomials, but the variational approach is easier.

where $r^2 = (x^2 + y^2)$. As for the 3-D exciton considered in Exercise 4.4, we guess a trial wave function with a 1s-like radial dependence:

$$\Psi(r, \theta) = \left(\frac{2}{\pi\xi^2}\right)^{\frac{1}{2}} \exp\left(-\frac{r}{\xi}\right),$$

where ξ is the variational parameter.

- (i) Verify that the trial wave function is properly normalized.
- (ii) The variational energy $\langle E \rangle_{\text{var}}$ is given by:

$$\langle E \rangle_{\text{var}} = \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} \Psi^* \hat{H} \Psi r dr d\theta.$$

Show that $\langle E \rangle_{\text{var}}$ is given by:

$$\langle E \rangle_{\text{var}} = \frac{\hbar^2}{2\mu\xi^2} - \frac{e^2}{2\pi\epsilon_r\epsilon_0\xi}.$$

- (iii) Vary $\langle E \rangle_{\text{var}}$ with respect to ξ to obtain the best estimate for the energy. Show that this is four times larger than that of the equivalent bulk semiconductor.
- (iv) Show the Bohr radius of the 2-D exciton, namely the value of ξ that minimizes $\langle E \rangle_{\text{var}}$, is half that of the equivalent 3-D exciton.

(6.10)* Discuss qualitatively how you would expect the exciton binding energy in a GaAs/Al_{0.3}Ga_{0.7}As quantum well to vary with the quantum well thickness, given that the binding energy of the excitons in bulk GaAs and Al_{0.3}Ga_{0.7}As are 4 meV and 6 meV respectively.

(6.11) Figure 6.17 shows the absorption spectrum of a GaAs/AlAs quantum well at 4 K measured using the photoluminescence excitation technique.

- (i) Explain the principles of photoluminescence excitation spectroscopy.
- (ii) Account for the shape of the absorption spectrum.
- (iii) Estimate the width of the quantum wells by assuming that the wells behave like a perfect 2-D system with infinite barriers. Would you expect the true well width to be larger or smaller than the answer you have worked out this way?
- (iv) Deduce the binding energies of the heavy and light hole excitons, and comment on the values you obtain.

Band structure data for GaAs is given in Table C.2.

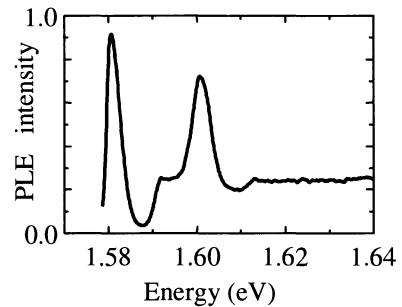


Fig. 6.17 Photoluminescence excitation (PLE) spectrum of a GaAs/AlAs quantum well at 4 K. (After R.A. Taylor, Personal communication)

(6.12)* The Stark shift of the confined levels in a quantum well can be calculated by using second-order perturbation theory. Consider the interaction between the electrons in a quantum well of width d and a DC electric field of strength \mathcal{E}_z applied along the z (growth) axis.

- (i) Explain why the perturbation to the energy of an electron is given by $H' = e\mathcal{E}_z z$.
- (ii) Explain why the first-order shift of the energy levels, given by:

$$\Delta E^{(1)} = \int_{-\infty}^{+\infty} \varphi(z)^* H' \varphi(z) dz,$$

is zero.

- (iii) The second-order energy shift of the $n = 1$ level is given by:

$$\Delta E^{(2)} = \sum_{n>1} \frac{|(1|H'|n\rangle|^2}{E_1 - E_n},$$

where

$$(1|H'|n\rangle = \int_{-\infty}^{+\infty} \varphi_1(z)^* H' \varphi_n(z) dz.$$

This can be evaluated exactly if we have infinite confining barriers. Within this approximation, show that the Stark shift is given approximately by:

$$\Delta E = -24 \left(\frac{2}{3\pi}\right)^6 \frac{e^2 \mathcal{E}_z^2 m^* d^4}{\hbar^2}.$$

(6.13) The magnitude of the $n = 1$ heavy hole exciton red shift is 10.5 nm at -10 V for the MQW p-i-n diode studied in Fig. 6.12. The sample has an i-region thickness of $1.0 \mu\text{m}$ and a built-in voltage of 1.5 V.

- (i) Estimate the magnitude of the red shift at -5 V.
- (ii) Estimate the average relative displacement of the electron and hole probability densities at -10 V.

* Exercises marked with an asterisk are more challenging.

- (6.14) Table 6.3 gives experimental data for the red shift of the $n = 1$ heavy hole transition due to the quantum confined Stark effect in two GaAs quantum well samples with well widths of 10 nm and 18 nm respectively. Compare the experimental data with the predictions of Exercise 6.12(iii), and account qualitatively for any major discrepancies. Band structure data for GaAs is given in Table C.2.

Table 6.3 Dependence of the $n = 1$ heavy hole transition on the electric field strength \mathcal{E}_z for two GaAs quantum well samples. Sample A had a well width of 10 nm, while sample B had a well width of 18 nm. The transition energies are given in eV.

\mathcal{E}_z (Vm $^{-1}$)	A	B
0	1.548	1.524
3×10^6	1.547	1.518
6×10^6	1.543	1.497
9×10^6	1.535	1.470

- (6.15) A DC electric field of magnitude \mathcal{E}_z is applied along the growth (z) axis of a quantum well. Use symmetry arguments to explain why transitions between confined electron and holes states with Δn equal to an odd number are forbidden at $\mathcal{E}_z = 0$, but not at finite \mathcal{E}_z .
- (6.16) By assuming that the confinement energy varies as d^{-2} , estimate the shift in the luminescence emission energy caused by a $\pm 5\%$ change in d for a 2.5 nm ZnCdSe quantum well, given that the total confinement energy for the electrons and holes at $d = 2.5$ nm is 0.1 eV. Compare this value to the measured linewidth of the 10 K data shown in Fig. 6.13, and comment on the answer. The unit cell size of the crystal is 0.28 nm, and the electron and

hole effective masses of $Zn_{0.8}Cd_{0.2}Se$ are $0.15m_0$ and $0.5m_0$ respectively.

- (6.17) A $Ga_{0.47}In_{0.53}As$ quantum well laser is designed to emit at $1.55 \mu m$ at room temperature. Estimate the width of the quantum wells within the device. ($E_g = 0.75$ eV, $m_e^* = 0.041m_0$, $m_{hh}^* = 0.47m_0$.)

- (6.18)* The matrix element for an intersubband transition between the n th and n' th subbands of a quantum well is given by:

$$\langle n|z|n' \rangle = \int_{-\infty}^{+\infty} \varphi_n^*(z) z \varphi_{n'}(z) dz.$$

- (i) By considering the parity of the states, prove that $\Delta n = (n - n')$ must be an odd number.
- (ii) Compare the relative strengths of the $1 \rightarrow 2$ and the $1 \rightarrow 4$ transitions in a 20 nm GaAs quantum well with infinite barriers. What is the wavelength of the $1 \rightarrow 2$ transition? ($m_e^* = 0.067m_0$.)

- (6.19) Linearly polarized light is incident on a quantum well sample at an angle θ to the normal (z) direction. The polarization direction lies within the plane of incidence. What is the maximum fraction of the power in the beam that can be absorbed by intersubband transitions if the refractive index of the crystal is 3.3?

- (6.20) Figure 6.15 shows the density of states for the conduction band electrons in a cubic quantum dot of dimension d . The barriers at the quantum dot interfaces are assumed to be infinite. Account for the energies of the quantized levels, and state the energies of the next four levels above $9(h^2/8m^*d^2)$.

Free electrons

7

In this chapter we will investigate the optical properties associated with free electrons. As the name suggests, these are electron systems that experience no restoring force from the medium when driven by the electric field of a light wave. The two main solid state systems that exhibit strong free electron effects are:

- **Metals.** Metals contain large densities of free electrons that originate from the valence electrons of the metal atoms.
- **Doped semiconductors.** n-type semiconductors contain free electrons, while p-type materials contain free holes. The free carrier density is determined by the concentration of impurities used for the doping.

We will begin our discussion of their optical properties by using the Drude–Lorentz model introduced in Section 2.1.3 of Chapter 2. This will enable us to explain the main optical property of metals that we mentioned in Section 1.4.3, namely that they reflect strongly in the visible spectral region. We will then apply our knowledge of interband transitions from Chapter 3 to obtain a better understanding of the detailed form of the reflectivity spectra of metals such as aluminium and copper. Next we will apply the Drude–Lorentz model to doped semiconductors to explain why doping causes infrared absorption. Finally, we will consider the collective oscillations of the whole free carrier gas. This will naturally lead us to the notion of plasmons, which are elementary excitations of the quantized plasma oscillations.

7.1 Plasma reflectivity

A neutral gas of heavy ions and light electrons is called a **plasma**. Metals and doped semiconductors can be treated as plasmas because they contain equal numbers of fixed positive ions and free electrons. The free electrons experience no restoring forces when they interact with electromagnetic waves. This contrasts with bound electrons that have natural resonant frequencies in the visible or ultraviolet spectral regions due to the restoring forces of the medium.

In this section we will derive a formula for the dielectric constant of an electron plasma using the classical oscillator model discussed in Section 2.2 of Chapter 2. As noted in Section 2.1.3, this approach combines the **Drude model** of free electron conductivity with the **Lorentz model** of dipole oscillators, and is therefore known as the **Drude–Lorentz model**.

We begin by considering the oscillations of a free electron induced by the AC electric field $\mathcal{E}(t)$ of an electromagnetic wave. The equation of motion for

7.1	Plasma reflectivity	143
7.2	Free carrier conductivity	145
7.3	Metals	147
7.4	Doped semiconductors	154
7.5	Plasmons	160

We have assumed here that the light is polarized along the x direction. The model is not affected by this arbitrary choice provided the medium is isotropic, as it usually will be for a free electron system.

the displacement x of the electron is:

$$m_0 \frac{d^2x}{dt^2} + m_0\gamma \frac{dx}{dt} = -e\mathcal{E}(t) = -e\mathcal{E}_0 e^{-i\omega t}. \quad (7.1)$$

where ω is the frequency of the light, and \mathcal{E}_0 is its amplitude. The first term represents the acceleration of the electron, while the second is the frictional damping force of the medium. The term on the right hand side is the driving force exerted by the light. Equation 7.1 is the same as the equation of motion for a bound oscillator given in eqn 2.5, except that there is no restoring force term because we are dealing with free electrons.

By substituting $x = x_0 e^{-i\omega t}$ into eqn 7.1, we obtain

$$x(t) = \frac{e\mathcal{E}(t)}{m_0(\omega^2 + i\gamma\omega)}. \quad (7.2)$$

The polarization P of the gas is equal to $-Nex$, where N is the number of electrons per unit volume. By recalling the definitions of the electric displacement D and the relative dielectric constant ϵ_r (cf. eqns A.2 and A.3), we can write:

$$\begin{aligned} D &= \epsilon_r \epsilon_0 \mathcal{E} \\ &= \epsilon_0 \mathcal{E} + P \\ &= \epsilon_0 \mathcal{E} - \frac{Ne^2 \mathcal{E}}{m_0(\omega^2 + i\gamma\omega)}. \end{aligned} \quad (7.3)$$

Therefore:

$$\epsilon_r(\omega) = 1 - \frac{Ne^2}{\epsilon_0 m_0} \frac{1}{(\omega^2 + i\gamma\omega)}. \quad (7.4)$$

This equation is identical to eqn 2.14 for the bound oscillator except that the resonant frequency ω_0 is zero and we have not yet considered the effects of background polarizability. Equation 7.4 is frequently written in the more concise form:

$$\epsilon_r(\omega) = 1 - \frac{\omega_p^2}{(\omega^2 + i\gamma\omega)}. \quad (7.5)$$

where

$$\omega_p = \left(\frac{Ne^2}{\epsilon_0 m_0} \right)^{\frac{1}{2}}. \quad (7.6)$$

ω_p is known as the **plasma frequency**.

Let us first consider a lightly damped system. In this case, we put $\gamma = 0$ in eqn 7.5 so that

$$\epsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2}. \quad (7.7)$$

The complex refractive index \tilde{n} of the medium is related to the complex dielectric constant by $\tilde{n} = \sqrt{\epsilon_r}$ (cf. eqn 1.19). This means that \tilde{n} is imaginary for $\omega < \omega_p$ and positive for $\omega > \omega_p$, with a value of zero precisely at $\omega = \omega_p$. The reflectivity R can be calculated from eqn 1.26:

$$R = \left| \frac{\tilde{n} - 1}{\tilde{n} + 1} \right|^2. \quad (7.8)$$

The fact that the refractive index is imaginary below ω_p means that the extinction coefficient κ is large, and hence that the medium is highly absorbing. This point will be explained further in Section 7.2. It is a general property of systems with high absorption coefficients that they also have high reflectivities.

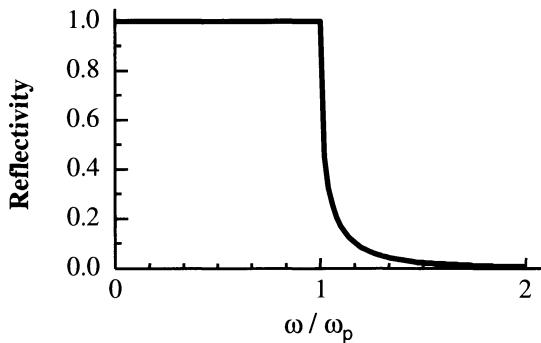


Fig. 7.1 Reflectivity of an undamped free carrier gas as a function of frequency.

By substituting the frequency dependence of \tilde{n} into this formula, we see that R is unity for $\omega \leq \omega_p$, and then decreases for $\omega > \omega_p$, approaching zero at $\omega = \infty$. This frequency dependence is plotted in Fig. 7.1.

The basic conclusion is that we expect the reflectivity of a gas of free electrons to be 100 % for frequencies up to ω_p . This result is very well confirmed by experimental data. In Sections 7.3 and 7.4 below we will see how the plasma reflectivity effect is observed in both metals and doped semiconductors.

One of the best examples of plasma reflectivity effects is the reflection of radio waves from the upper atmosphere. The atoms in the ionosphere are ionized by the ultraviolet light from the sun to produce a plasma of ions and free electrons. The plasma frequency is in the MHz range, and so the low frequency waves used for AM radio transmissions are reflected, but not the higher frequency waves used for FM radio or television. (See Exercise 7.2.)

Example 7.1

Zinc is a divalent metal with $6.6 \times 10^{28} \text{ m}^{-3}$ atoms per unit volume. Account for the shiny appearance of zinc.

Solution

Zinc has two electrons per atom, and so the free electron density N is $2 \times (6.6 \times 10^{28}) = 1.3 \times 10^{29} \text{ m}^{-3}$. We use this value of N in eqn 7.6 to find $\omega_p = 2.0 \times 10^{16} \text{ Hz}$. This corresponds to a wavelength of $2\pi c/\omega_p = 93 \text{ nm}$, which is in the ultraviolet spectral region. The free electrons in zinc will reflect all frequencies below ω_p . All wavelengths in the visible spectral region are therefore reflected, and zinc has a shiny mirror-like surface.

7.2 Free carrier conductivity

In deriving eqn 7.7, we neglected the damping of the free carrier oscillations. We can recast the equation of motion in a way that makes the physical significance of the damping term more apparent. To do this we note that \dot{x} is the electron velocity \mathbf{v} . Hence we can rewrite eqn 7.1 as:

$$m_0 \frac{d\mathbf{v}}{dt} + m_0 \gamma \mathbf{v} = -e\mathbf{E}. \quad (7.9)$$

Since $m_0 \mathbf{v}$ is the momentum \mathbf{p} , we see that:

$$\frac{d\mathbf{p}}{dt} = -\frac{\mathbf{p}}{\tau} - e\mathbf{E}, \quad (7.10)$$

where we have replaced the damping rate γ by the reciprocal of the damping time τ . This shows that the electron is being accelerated by the field, but loses its momentum in time τ . In other words, τ is the **momentum scattering time**.

In an AC field of the form $\mathcal{E}(t) = \mathcal{E}_0 e^{-i\omega t}$, we look for solutions to the equation of motion with $x = x_0 e^{-i\omega t}$. This implies that $|\mathbf{v}| = \dot{x}$ also has a time variation of the form $\mathbf{v} = \mathbf{v}_0 e^{-i\omega t}$. On substituting this into eqn 7.9, we obtain:

$$\mathbf{v}(t) = \frac{-e\tau}{m_0} \frac{1}{1 - i\omega\tau} \mathcal{E}(t). \quad (7.11)$$

The current density \mathbf{j} is related to the velocity and field through:

$$\mathbf{j} = -Ne\mathbf{v} = \sigma \mathcal{E}, \quad (7.12)$$

where σ is the electrical conductivity. On combining eqns 7.11 and 7.12, we obtain the **AC conductivity** $\sigma(\omega)$:

$$\sigma(\omega) = \frac{\sigma_0}{1 - i\omega\tau}, \quad (7.13)$$

where

$$\sigma_0 = \frac{Ne^2\tau}{m_0}. \quad (7.14)$$

σ_0 is the conductivity measured with DC electric fields.

In principle, the momentum scattering time can be deduced from eqn 7.13 by measuring the conductivity of the free carrier gas as a function of frequency. However, for a metal or doped semiconductor, the value of τ is typically in the range 10^{-14} – 10^{-13} s. This means that we have to go to near optical frequencies to notice any departure from the DC conductivity. Hence we must use optical measurements to obtain information about τ .

By comparing eqns 7.4 and 7.13, we see that the AC conductivity and the dielectric constant are related to each other through:

$$\epsilon_r(\omega) = 1 + \frac{i\sigma(\omega)}{\epsilon_0\omega}. \quad (7.15)$$

Thus optical measurements of $\epsilon_r(\omega)$ are equivalent to AC conductivity measurements of $\sigma(\omega)$.

At very low frequencies that satisfy $\omega \ll \tau^{-1}$, we can derive a useful relationship between the conductivity of the free carrier gas and the attenuation coefficient for electromagnetic waves. This can be achieved by first splitting $\epsilon_r(\omega)$ into its real and imaginary components in accordance with eqn 1.18. Equation 7.5 with $\gamma = \tau^{-1}$ gives:

$$\epsilon_1 = 1 - \frac{\omega_p^2\tau^2}{1 + \omega^2\tau^2} \quad (7.16)$$

$$\epsilon_2 = \frac{\omega_p^2\tau}{\omega(1 + \omega^2\tau^2)}. \quad (7.17)$$

We then work out n and κ , the real and imaginary parts of the complex refractive index, using eqns 1.22 and 1.23, and hence deduce the attenuation

Optical data on free carrier reflectivity is frequently discussed in terms of the conductivity rather than the dielectric constants.

coefficient α from κ . Since $\omega\tau \ll 1$ implies that $\epsilon_2 \gg \epsilon_1$, we can obtain solutions with $n \approx \kappa = (\epsilon_2/2)^{\frac{1}{2}}$. Using eqn 1.16 we therefore obtain:

$$\alpha = \frac{2\omega(\epsilon_2/2)^{\frac{1}{2}}}{c} = \left(\frac{2\omega_p^2 \tau \omega}{c^2} \right)^{\frac{1}{2}}. \quad (7.18)$$

We can put this equation in a more accessible form by noting from eqn 7.14 that $\omega_p^2 \tau = \sigma_0/\epsilon_0$ and from eqn A.28 that $c^2 = 1/\epsilon_0 \mu_0$. This gives:

$$\alpha = (2\sigma_0 \omega \mu_0)^{\frac{1}{2}}. \quad (7.19)$$

Hence we see that the attenuation coefficient is proportional to the square root of the DC conductivity and the frequency.

Equation 7.19 implies that AC electric fields can only penetrate a short distance into a conductor such as a metal. This well-known phenomenon is called the **skin effect**. If the field strength varies as $\exp(-z/\delta)$ with the distance z from the surface, then the power falls off as $\exp(-2z/\delta)$. By comparison with the definition of α in eqn 1.4, we see that:

$$\delta = \frac{2}{\alpha} = \left(\frac{2}{\sigma_0 \omega \mu_0} \right)^{\frac{1}{2}}. \quad (7.20)$$

δ is known as the **skin depth**.

At higher frequencies the relationship given in eqn 7.19 breaks down because the approximation $\omega\tau \ll 1$ is no longer valid. In this case we can derive a different frequency dependence for the attenuation coefficient. This will be discussed when we consider the absorption due to free carriers in doped semiconductors in Section 7.4.1

Example 7.2

The DC electrical conductivity of copper is $6.5 \times 10^7 \Omega^{-1} \text{m}^{-1}$ at room temperature. Calculate the skin depth at 50 Hz and 100 MHz.

Solution

The skin depth is given by eqn 7.20. At 50 Hz we have $\omega = 2\pi \times 50 = 314 \text{ Hz}$. Inserting this value of ω into eqn 7.20 with $\sigma_0 = 6.5 \times 10^7 \Omega^{-1} \text{m}^{-1}$ gives $\delta = 8.8 \text{ mm}$. At 100 MHz, $\omega = 6.28 \times 10^8 \text{ Hz}$, and the skin depth δ is only $6.2 \mu\text{m}$.

7.3 Metals

The free electron model of metals was proposed by P. Drude in 1900. The model provides a basic explanation for why metals are good conductors of

heat and electricity, and is the starting point for more sophisticated theories. As we will see here, it is also successful in explaining a number of important optical properties, such as the fact that metals tend to be good reflectors. On the other hand, we will need to use band theory to explain the details of the optical spectra, and also to account for the coloured appearance of some metals, for example, copper and gold.

7.3.1 The Drude model

The Drude free electron model of metals considers the valence electrons of the atoms to be free. When an electric field is applied, the free electrons accelerate and then undergo collisions with the characteristic scattering time τ introduced in eqn 7.10. The electrical conductivity is therefore limited by the scattering, and measurements of σ allow the value of τ to be determined through eqn 7.14.

The free electron density N in the Drude model is equal to the density of metal atoms multiplied by their valency. Table 7.1 lists the Drude free electron densities for a number of common metals. The values of N are in the range 10^{28} – 10^{29} m^{-3} . These very large free electron densities explain why metals have high electrical and thermal conductivities. The plasma frequencies calculated using eqn 7.6 are also tabulated in Table 7.1, together with the wavelength λ_p that corresponds to ω_p . It is apparent that the very large values of N lead to plasma frequencies in the ultraviolet spectral region.

In the visible spectral region where $\omega/2\pi \sim 10^{15} \text{ Hz}$, we are usually in a situation with $\omega \gg \gamma$. This is because $\tau = \gamma^{-1}$ is typically of order 10^{-14} s . Therefore the simplification of eqn 7.5 to eqn 7.7 is a good approximation. With ω_p in the ultraviolet, the visible photons have frequencies below ω_p and thus ϵ_r is negative. As discussed in Section 7.1, this means that the reflectivity is expected to be 100 % up to ω_p . This explains the first and most obvious optical property of metals, namely that they tend to be good reflectors at visible frequencies.

Table 7.1 Free electron density and plasma properties of some metals. The figures are for room temperature unless stated otherwise. The electron densities are based on data taken from reference [1]. The plasma frequency ω_p is calculated from eqn 7.6, and λ_p is the wavelength corresponding to this frequency.

Metal	Valency	N (10^{28} m^{-3})	$\omega_p/2\pi$ (10^{15} Hz)	λ_p (nm)
Li (77 K)	1	4.70	1.95	154
Na (5 K)	1	2.65	1.46	205
K (5 K)	1	1.40	1.06	282
Rb (5 K)	1	1.15	0.96	312
Cs (5 K)	1	0.91	0.86	350
Cu	1	8.47	2.61	115
Ag	1	5.86	2.17	138
Au	1	5.90	2.18	138
Be	2	24.7	4.46	67
Mg	2	8.61	2.63	114
Ca	2	4.61	1.93	156
Al	3	18.1	3.82	79

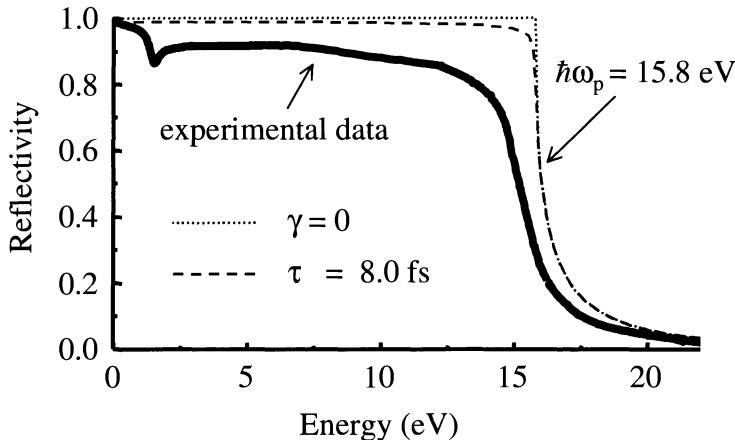


Fig. 7.2 Experimental reflectivity of aluminium as a function of photon energy. The experimental data is compared to predictions of the free electron model with $\hbar\omega_p = 15.8$ eV. The dotted curve is calculated with no damping. The dashed line is calculated with $\tau = 8.0 \times 10^{-15}$ s, which is the value deduced from DC conductivity. The experimental data is taken from reference [3], copyright 1963 American Institute of Physics, reprinted with permission.

A striking implication of the free carrier model is that the dielectric constant changes from being negative to positive as we go through the plasma frequency. This means that the reflectivity ceases to be 100 % above ω_p (see Fig. 7.1) and some of the light can be transmitted through the metal. Thus we expect that all metals will eventually become transmitting if we go far enough into the ultraviolet so that $\omega > \omega_p$. This phenomenon is known as the **ultraviolet transparency of metals**.

In order to observe the ultraviolet transmission threshold at the plasma frequency, it is necessary that there should be no other absorption processes occurring at ω_p . This condition is best satisfied in the alkali metals. Table 7.2 lists the wavelengths of the ultraviolet transmission edges observed in the alkalies. The experimental wavelengths can be compared with those predicted from the calculated plasma frequency tabulated in Table 7.1. The experimental results are in reasonable agreement with the predictions, and show the correct trend on descending the periodic table. The discrepancies can be explained to a large extent by replacing the free electron mass with the electron effective mass derived from the band structure of the metal. (See Exercise 7.4.)

Figure 7.2 shows the measured reflectivity of aluminium as a function of photon energy from the infrared to the ultraviolet spectral region. We see that the reflectivity is over 80 % up to ~ 15 eV, and then drops off to zero at higher frequencies. Thus aluminium shows the characteristic ultraviolet transparency edge predicted by the Drude model. The relatively featureless reflectivity at visible frequencies is exploited in commercial mirrors.

The plasma frequency listed in Table 7.1 corresponds to a photon energy of 15.8 eV. The dotted line in Fig. 7.2 gives the reflectivity predicted from eqn 7.7 with $\hbar\omega_p = 15.8$ eV. On comparing the experimental and theoretical results, we see that the model accounts for the general shape of the spectrum, but there are some important details that are not explained.

An improved attempt to model the experimental data can be made by computing the dielectric constant with the damping term included using eqn 7.5. Example 3.3 explains how this is done. The reflectivity calculated for the value of τ deduced from the DC conductivity, namely 8.0×10^{-15} s, is plotted as the dashed line in Fig. 7.2. The main difference between the two calculated curves

Table 7.2 Ultraviolet transmission threshold wavelength λ_{UV} for the alkali metals. After [2].

Metal	λ_{UV} (nm)
Li	205
Na	210
K	315
Rb	360
Cs	440

We will see in Section 7.5 that the plasma frequency can be determined directly using electron energy loss spectroscopy.

is that the damping causes the reflectivity to be less than unity below ω_p , and the ultraviolet transmission edge is slightly broadened. However, this is only a relatively small effect because $\omega_p \gg \tau^{-1}$.

The inclusion of damping makes a small improvement in the fit to the data, but there are two important features that are still not explained. Firstly, the reflectivity is significantly lower than predicted, and secondly, there is a dip around 1.5 eV, where we would have expected a featureless curve. Both of these points can be explained by considering the interband absorption rates. These are discussed in the next section.

Example 7.3

The conductivity of aluminium at room temperature is $4.1 \times 10^7 \Omega^{-1}\text{m}^{-1}$. Calculate the reflectivity at 500 nm.

Solution

We first work out the damping time τ from the conductivity using eqn 7.14. Taking the value of $N = 1.81 \times 10^{29} \text{ m}^{-3}$ from Table 7.1, we find:

$$\tau = \frac{m_0 \sigma_0}{N e^2} = 8.0 \times 10^{-15} \text{ s}.$$

Table 7.1 also gives us the value of the plasma frequency, namely $\omega_p = 2.4 \times 10^{16} \text{ Hz}$. The wavelength of 500 nm corresponds to an angular frequency $\omega = 2\pi c/\lambda = 3.8 \times 10^{15} \text{ Hz}$. We use these frequencies in eqns 7.16 and 7.17 to calculate the real and imaginary parts of the complex dielectric constant:

$$\epsilon_1 = 1 - \frac{\omega_p^2 \tau^2}{1 + \omega^2 \tau^2} = -39,$$

and

$$\epsilon_2 = \frac{\omega_p^2 \tau}{\omega(1 + \omega^2 \tau^2)} = 1.3.$$

We then work out the real and imaginary parts of the complex refractive index using eqns 1.22 and 1.23. This gives:

$$n = \frac{1}{\sqrt{2}} \left(-39 + [(-39)^2 + (1.3)^2]^{\frac{1}{2}} \right)^{\frac{1}{2}} = 0.10,$$

and

$$\kappa = \frac{1}{\sqrt{2}} \left(+39 + [(-39)^2 + (1.3)^2]^{\frac{1}{2}} \right)^{\frac{1}{2}} = 6.2.$$

We finally obtain the reflectivity from eqn 1.26:

$$R = \frac{(n-1)^2 + \kappa^2}{(n+1)^2 + \kappa^2} = \frac{(-0.9)^2 + (6.2)^2}{(1.1)^2 + (6.2)^2} = 99 \text{ \%}.$$

This shows that the inclusion of the damping only reduces the reflectivity by 1 % in this case.

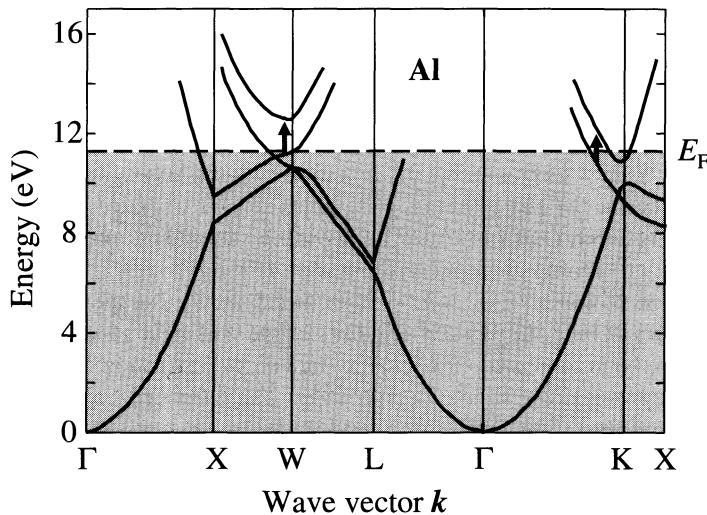


Fig. 7.3 Band diagram of aluminium. The transitions at the W and K points that are responsible for the reflectivity dip at 1.5 eV are labelled. After [4], copyright 1961 American Institute of Physics, reprinted with permission.

7.3.2 Interband transitions in metals

The absorption of light by direct interband transitions was discussed in detail in Chapter 3. Direct transitions involve the promotion of electrons to a higher band by absorption of a photon with the correct energy. The electron does not change its k vector significantly because of the very small momentum of the photon. Thus the transitions appear as vertical arrows on the E - k band diagram of the solid.

Interband absorption is important in metals because the electromagnetic waves penetrate a short distance into the surface, and if there is a significant probability for interband absorption, the reflectivity will be reduced from the free carrier value. The interband absorption spectra of metals are determined by their complicated band structures and Fermi surfaces. Furthermore, we need to consider transitions at frequencies in which the free carrier properties are also important. For this reason we will only consider two examples to illustrate the general principles, namely aluminium and copper, and then make some general comments about other metals such as silver and gold.

Aluminium

The band diagram of aluminium is shown in Fig. 7.3. Aluminium has an electronic configuration of $[\text{Ne}]3s^23p^1$ with three valence electrons. The crystal structure is face-centred cubic, which has a body-centred cubic (bcc) reciprocal lattice, as shown in Fig. C.5 of Appendix C. The first Brillouin zone is completely full, and the valence electrons spread into the second, third and slightly into the fourth zones. The band structure appears quite complex due to the irregular shape of the bcc Brillouin zone. However, the bands are actually very close to the free electron model, with significant departures only in the vicinity of the Brillouin zone boundaries. The bands are filled up to the Fermi energy E_F , which is marked on the diagram. Direct transitions can take place from any of the states below the Fermi level to unoccupied bands directly above them on the E - k diagram.

We came across a similar example of parallel bands when we discussed the absorption rate at the critical points in the band structure of silicon in Section 3.5.

Fermi's golden rule given in eqn 3.2 tells us that the absorption rate is proportional to the density of states for the transition. The dip in the reflectivity at 1.5 eV which is apparent in Fig. 7.2 is a consequence of the 'parallel-band' effect. This occurs when there is a band above the Fermi level that is approximately parallel to another band below E_F . In this case, the interband transitions from a large number of occupied k states below the Fermi level will all occur at the same energy. Hence the density of states at the energy difference between the two parallel bands will be very high, which will result in a particularly strong absorption at this photon energy.

Inspection of the band diagram of aluminium shows that the parallel-band effect occurs at both the W- and K-points of the Brillouin zone. These transitions have been identified on Fig. 7.3. The energy separation of the parallel bands is approximately 1.5 eV in both cases. The enhanced transition rate at this photon energy thus explains the reflectivity dip observed at 1.5 eV in the experimental data. Moreover, we can see from the band diagram that there will be further transitions between bands below the Fermi level to unoccupied bands above E_F at a whole range of photon energies greater than 1.5 eV. The density of states for these transitions will be lower than at 1.5 eV because the bands are not parallel. However, the absorption rate is still significant, and accounts for the reduction of the reflectivity to a value below that predicted by the Drude model in the visible and ultraviolet spectral regions.

Copper

Copper has an electronic configuration of $[\text{Ar}]3d^{10}4s^1$. The outer $4s$ bands approximate reasonably well to free electron states with a dispersion given by $E = \hbar^2 k^2 / 2m_0$. They therefore form a broad band covering a wide range of energies. The $3d$ bands, on the other hand, are more tightly bound and are relatively dispersionless, occupying only a narrow range of energies. The density of states for the two bands is illustrated schematically in Fig. 7.4. The narrow $3d$ bands can hold ten electrons, and therefore their density of states is sharply peaked. The $4s$ band, which can hold two electrons, is much broader, with a smaller maximum. The 11 valence electrons of copper fill up the $3d$ band, and half fill the $4s$ band. The Fermi energy therefore lies within the $4s$ band above the $3d$ band. Interband transitions are possible from the filled $3d$ bands to unoccupied states in the $4s$ band above E_F , as illustrated in Fig. 7.4. This implies that there will be a well-defined threshold for interband transitions from the $3d$ bands to the $4s$ band.

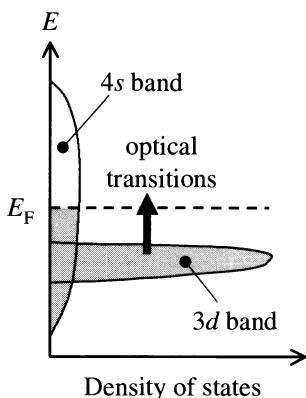


Fig. 7.4 Schematic density of states for the $3d$ and $4s$ bands of a transition metal such as copper.

Figure 7.5 shows the actual band structure and density of states of copper. The general features indicated in Fig. 7.4 are apparent in the calculated curves. The $3d$ electrons lie in relatively narrow bands with very high densities of states, while the $4s$ bands are much broader with a lower density of states. The Fermi energy lies in the middle of the $4s$ band above the $3d$ band. Interband transitions are possible from the $3d$ bands below E_F to unoccupied levels in the $4s$ band above E_F . The lowest energy transitions are marked on the band diagram in Fig. 7.5. The transition energy is 2.2 eV which corresponds to a wavelength of 560 nm.

Figure 7.6 shows the measured reflectivity of copper from the infrared to the ultraviolet spectral region. Based on the plasma frequency given in Table 7.1,

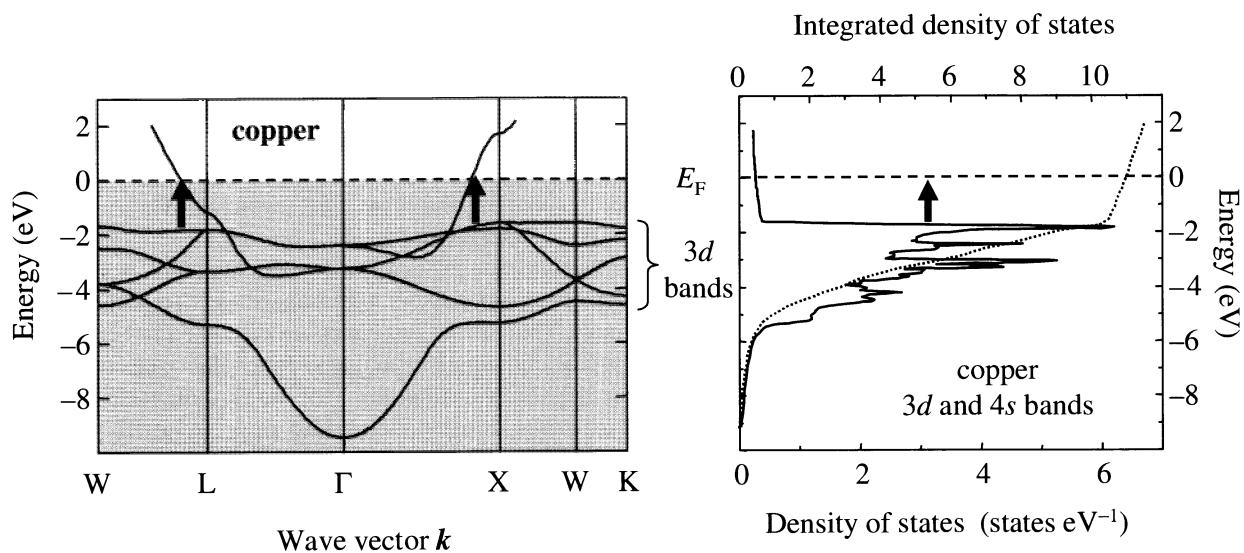


Fig. 7.5 Calculated band structure of copper. The transitions from the 3d bands responsible for the interband transitions around 2 eV are identified. The right hand side of the figure shows the density of states calculated from the band structure. The strongly peaked features between about -2 eV and -5 eV are due to the 3d bands. The dotted line is the integrated density of states. The Fermi level corresponds to the energy where the integrated density of states is equal to 11. After [5].

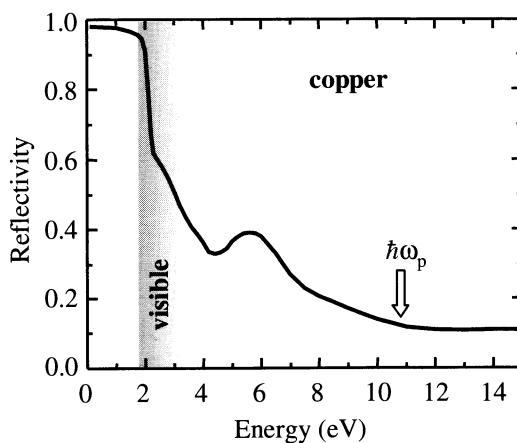


Fig. 7.6 Reflectivity of copper from the infrared to the ultraviolet spectral region. The reflectivity drops sharply above 2 eV due to interband transitions. After [6].

we would expect near-100 % reflectivity for photon energies below 10.8 eV, which corresponds to an ultraviolet wavelength of 115 nm. However, the experimental reflectivity falls off sharply above 2 eV due to the interband absorption edge discussed above. This explains why copper has a reddish colour.

Silver and gold

The arguments used for copper can be applied to the other noble metals. The important parameter is the energy gap between the d -bands and the Fermi energy, as shown in Fig. 7.4. In gold the interband absorption threshold occurs at a slightly higher energy than copper, which explains why it has a yellowish

colour. In silver, on the other hand, the interband absorption edge is around 4 eV. This frequency is in the ultraviolet, and so the reflectivity remains high throughout the whole visible spectrum. (See Fig. 1.5.) This explains why silver does not have any particular colour, and also why it is so widely used for making mirrors. Gold is also used for mirrors, but only at infrared wavelengths.

7.4 Doped semiconductors

Silicon and germanium originate from group IV of the periodic table, and are therefore obviously four-valent. Compound semiconductors such as the III–Vs and II–VIIs also behave as if they have a valency of four, because they have a total of eight electrons for every two atoms, and these electrons are shared in a polar covalent bond.

The controlled doping of semiconductors with impurities is an essential part of solid state technology. The general principles are discussed in Section C.1 of Appendix C. The atoms of the pure crystal have four valence electrons. The introduction of **donor** impurities from group V of the periodic table provides an excess of electrons, while **acceptor** impurities from group III lead to a deficit of electrons, which is equivalent to an excess of holes. Doping that produces excess electrons is called **n-type**, while doping that produces excess holes is called **p-type**.

Experimental measurements on doped semiconductors show that the presence of impurities give rise to new absorption mechanisms and also to a free carrier plasma reflectivity edge. Our aim here is to explain these effects by applying a suitably modified version of the free carrier model and by considering the quantized levels created by the impurity atoms. In the two subsections that follow, we will first consider the free carrier effects, and then move on to discuss the absorption associated with the impurity levels.

7.4.1 Free carrier reflectivity and absorption

The free electron model developed in Sections 7.1 and 7.2 can be applied to doped semiconductors if we make two appropriate modifications. Firstly, we must account for the fact that the electrons and holes are moving in the conduction or valence band of a semiconductor. This is easily achieved by assuming that the carriers behave as particles with an effective mass m^* rather than the free electron mass m_0 . Secondly, we must remember that there are other mechanisms that can contribute to the dielectric constant as well as the free carrier effects. The main extra effect that we will need to consider is the contribution to the polarization due to the optical response of the bound electrons.

The two modifications mentioned above can be handled if we rewrite eqn 7.3 in the following form:

$$\begin{aligned} D &= \epsilon_r \epsilon_0 \mathcal{E} \\ &= \epsilon_0 \mathcal{E} + P_{\text{other}} + P_{\text{free carrier}} \\ &= \epsilon_{\text{opt}} \epsilon_0 \mathcal{E} - \frac{Ne^2 \mathcal{E}}{m^*(\omega^2 + i\gamma\omega)}. \end{aligned} \quad (7.21)$$

The term P_{other} accounts for the polarizability of the bound electrons, while the effective mass m^* accounts for the band structure of the semiconductor. The carrier density N that appears in this equation is the density of free electrons or holes generated by the doping process. Note that the sign of the charge cancels,

and so the only difference between electrons and holes in this treatment is in the effective mass that is used.

The free carrier effects due to doping are most noticeable in the spectral region 5–30 μm , where we would normally expect the semiconductor to be completely transparent. Hence the value of ϵ_{opt} that we use in eqn 7.21 is the one measured in the transparent spectral region below the interband absorption edge. This value is known from the refractive index of the undoped semiconductor: $\epsilon_{\text{opt}} = n^2$. (See eqn 1.24 with $\kappa = 0$ below the band edge.)

Equation 7.21 tells us that the frequency dependence of the dielectric constant is given by:

$$\epsilon_r(\omega) = \epsilon_{\text{opt}} - \frac{Ne^2}{m^*\epsilon_0} \frac{1}{(\omega^2 + i\gamma\omega)}. \quad (7.22)$$

This can be rewritten as:

$$\epsilon_r(\omega) = \epsilon_{\text{opt}} \left(1 - \frac{\omega_p^2}{(\omega^2 + i\gamma\omega)} \right), \quad (7.23)$$

where the plasma frequency ω_p is now given by

$$\omega_p^2 = \frac{Ne^2}{\epsilon_{\text{opt}}\epsilon_0 m^*}. \quad (7.24)$$

We have written the dielectric constant in this way to make the link to the Drude model apparent. The difference between the plasma frequency for the semiconductor given in eqn 7.24 and that given in eqn 7.6 is that we have replaced m_0 by m^* , and we have included a non-resonant dielectric constant to account for the background polarizability of the bound electrons.

If we assume that the system is lightly damped, then we can ignore the damping term in eqn 7.23. This then implies that ϵ_r is negative below ω_p and positive at higher frequencies. We thus expect to observe a plasma reflectivity edge at ω_p just as we did in metals. Since the carrier density is much smaller than in metals, the plasma edge occurs at frequencies in the infrared spectral range. This prediction is very well borne out by infrared reflectivity data.

Figure 7.7 shows the measured reflectivity of n-type InSb as a function of the electron density. The fundamental absorption edge at the band gap of InSb occurs at 6 μm , while the phonon absorption band lies around 50 μm . Thus we would expect pure InSb to be transparent in the wavelength range shown and have a featureless reflectivity spectrum. Instead, the data shows a well-defined reflectivity edge, which shifts to shorter wavelengths as the electron density increases, in accordance with eqn 7.24.

One very striking feature of the data is the zero in the reflectivity at wavelengths just below the plasma edge. This occurs at a frequency given by (see Exercise 7.8):

$$\omega^2 = \frac{\epsilon_{\text{opt}}}{\epsilon_{\text{opt}} - 1} \omega_p^2. \quad (7.25)$$

By fitting this formula to the data, the effective mass of InSb can be determined. (See Exercise 7.9.)

At frequencies above ω_p , the presence of free carriers leads to the absorption of light. This effect is called **free carrier absorption**, and can be observed in

As explained in Section 2.2.2 of Chapter 2, solids have a number of resonant frequencies, each of which can be modelled by dipole oscillators. There are resonant frequencies in the infrared due to the phonons, and others in the near-infrared, visible or ultraviolet due to the bound electrons. The phonon absorption bands will be discussed in detail in Chapter 10, and occur in the range 30–100 μm for a typical III–V semiconductor.

We will see in Section 7.5 that ω_p can also be determined directly by Raman scattering measurements.

The data shown in Fig. 7.7 demonstrates the phenomenon of the plasma reflectivity edge more clearly than many of the equivalent results obtained for metals. This is because it is not possible to vary the electron density in metals. Moreover, in metals the plasma frequencies are much higher, and the reflectivity edge is frequently obscured by interband transitions.

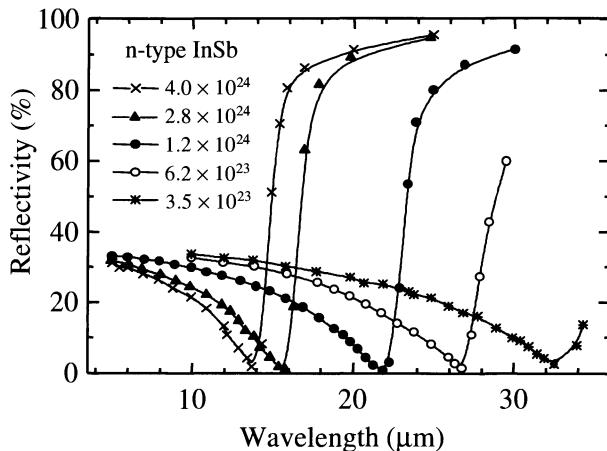


Fig. 7.7 Free carrier reflectivity of n-type InSb at room temperature as a function of the free electron density. After [7], copyright 1957 American Institute of Physics, reprinted with permission.

the near-infrared spectral region below the fundamental absorption edge at the band gap, where the semiconductor would normally be transparent. To see how this effect arises, we split the dielectric constant given in eqn 7.23 into its real and imaginary parts. This gives:

$$\epsilon_1 = \epsilon_{\text{opt}} \left(1 - \frac{\omega_p^2 \tau^2}{1 + \omega^2 \tau^2} \right) \quad (7.26)$$

$$\epsilon_2 = \frac{\epsilon_{\text{opt}} \omega_p^2 \tau}{\omega (1 + \omega^2 \tau^2)} \quad (7.27)$$

where we have made the usual substitution of τ^{-1} for γ . In a typical semiconductor, with $\tau \sim 10^{-13}$ s at room temperature, it is safe to make the approximation $\omega \tau \gg 1$ at frequencies in the near-infrared. Furthermore, the free carrier term in ϵ_r will be small. Therefore we can assume $\epsilon_1 \approx \epsilon_{\text{opt}}$, and that $\epsilon_2 \ll \epsilon_1$. In these conditions we find solutions to eqns 1.22 and 1.23 with $n = \sqrt{\epsilon_{\text{opt}}}$ and $\kappa = \epsilon_2/2n$. This allows us to deduce the absorption coefficient using eqn 1.16. The result is:

$$\alpha_{\text{free carrier}} = \frac{\epsilon_{\text{opt}} \omega_p^2}{nc\omega^2 \tau} = \frac{Ne^2}{m^* \epsilon_0 n c \tau} \frac{1}{\omega^2}. \quad (7.28)$$

This shows that the free carrier absorption is proportional to the carrier density and should vary with frequency as ω^{-2} .

Experimental data on a number of n-doped samples leads to the conclusion that $\alpha_{\text{free carrier}} \propto \omega^{-\beta}$, where β is in the range 2–3. The departure of β from the predicted value of 2 is caused by the failure of our assumption that τ is independent of ω . To see why this is important, we illustrate the physical processes that are occurring during free carrier absorption in Fig. 7.8. The figure shows the conduction band of an n-type semiconductor, which is filled up to the Fermi level determined by the free carrier density. Absorption of a photon excites an electron from an occupied state below the Fermi level to an unoccupied level above E_F . The photon only has a very small momentum compared to the electron, and therefore cannot change the electron's momentum significantly. It is obvious from Fig. 7.8 that a scattering event must occur to

The skin effect considered in Section 7.2 may also be considered as free carrier absorption. In the skin effect, however, we are considering the absorption at frequencies below ω_p where the material is highly reflective. We are now considering absorption above ω_p where the material should be transparent.

The approximation that τ is independent of ω effectively says that the relaxation time of the electrons does not depend on their initial energy. This is equivalent to the energy independent relaxation time approximation of the Boltzmann equation used in electron transport theory. It is well known that this approximation is only valid in a limited range of conditions. See Ashcroft and Mermin (1976) for further details.

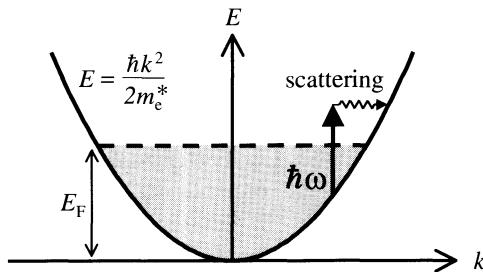


Fig. 7.8 A free carrier transition in a doped semiconductor.

conserve momentum in the process. Hence the absorption must be proportional to the scattering rate $1/\tau$, in accordance with the prediction of eqn 7.28.

The mechanisms that can contribute to the momentum conserving process in free carrier absorption include phonon scattering and scattering from the ionized impurities left behind by the release of the free electrons from their dopants. It is a sweeping oversimplification to characterize all the possible scattering processes with a single frequency-independent scattering time τ deduced from the DC conductivity. Thus it is hardly surprising that the experimental data does not exactly show an ω^{-2} dependence.

The free carrier reflectivity and absorption of p-type semiconductors can be modelled by a similar treatment to the one developed here for n-type samples. The only change that has to be made is in the effective mass that is used in the calculation. Thus we would expect that all the main results will hold, provided we take account of the fact that the scattering time for holes is not necessarily the same as that for electrons. However, p-type samples also show another effect in addition to those related with the free carriers, which is discussed below.

Figure 7.9 shows the valence band of a p-type III–V semiconductor. This is a larger scale version of the band structure diagram given previously in Fig. 3.5, except that now there are unfilled states near $k = 0$ due to the p-type doping. Optical transitions can take place in which an electron is promoted from an occupied state below E_F in the light hole (lh) band to an empty one in the heavy hole (hh) band above E_F . This is called **intervalence band absorption**. Other intervalence band transitions are possible in which an electron is promoted from the split-off (SO) band to either the lh or hh band. The range of energies over which these transitions occur can be calculated from the effective masses, the doping density and the split-off band energy (see Exercise 7.12). The absorption occurs in the infrared, and can be a strong process because no scattering events are required to conserve momentum.

Measurements of the intervalence band absorption give a value for the energy Δ of the split-off band relative to the top of the valence band, and also the ratio of the effective masses.

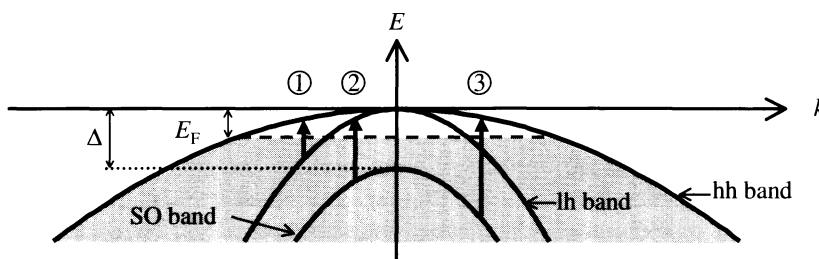


Fig. 7.9 Intervalence band absorption in a p-type semiconductor. E_F is the Fermi energy determined by the doping density. The labelled arrows indicate: (1) transitions from the light hole (lh) band to the heavy hole (hh) band; (2) transitions from the split-off (SO) band to the lh band; and (3) transitions from the SO band to the hh band.

7.4.2 Impurity absorption

The n-type doping of a semiconductor with donor atoms introduces a series of hydrogenic levels just below the conduction band. These quantized states are called **donor levels**, and are illustrated schematically in Fig. 7.10. The impurity levels give rise to two new absorption mechanisms, in addition to any of the free carrier effects discussed in the previous section. If the donor states are occupied, it will be possible to absorb photons by exciting electrons between the levels as illustrated in Fig. 7.10(a). On the other hand, if the states are empty, then it will be possible to absorb light by exciting electrons from the valence band to the donor states as illustrated in Fig. 7.10(b).

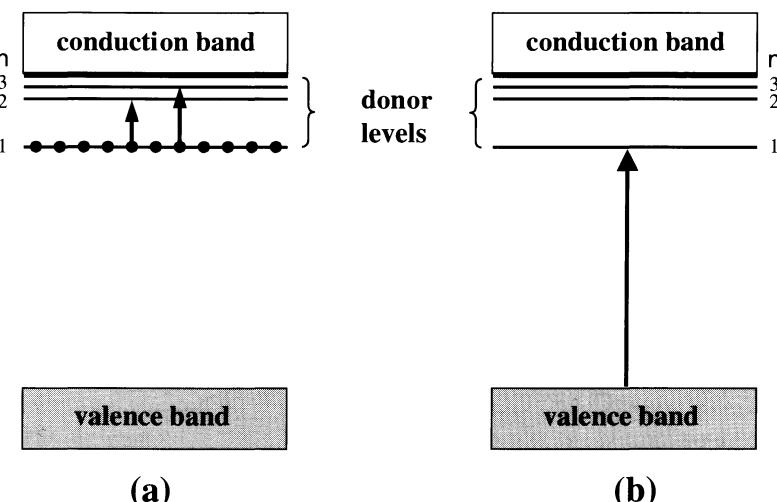
We consider first the transitions between the donor levels illustrated in Fig. 7.10(a). For such a process to occur, the donor levels must be occupied. This will be the case at low temperatures, when there is insufficient thermal energy to promote the electrons from the donor levels into the conduction band.

The frequencies of the donor level transitions can be calculated if the energies of the impurity states are known. In the simplest model, we assume that the electron is released into the crystal, and is then attracted back towards the positively charged impurity atom. The electron and the ionized impurity then form a hydrogenic system bound together by their mutual Coulomb attraction. As a first approximation, we can use the Bohr formula, provided we use the effective mass m_e^* instead of the free electron mass m_0 , and also include the dielectric constant ϵ_r for the semiconductor. Hence the energy of the donor levels E_n^D will be given by:

$$E_n^D = - \frac{m_e^*}{m_0} \frac{1}{\epsilon_r^2} \frac{R_H}{n^2}, \quad (7.29)$$

where R_H is the hydrogen Rydberg (13.6 eV) and n is an integer.

At low temperatures we can assume that all the electrons from the donors will be in the ground state $n = 1$ impurity level. Optical transitions can then take place in which the electrons are promoted to higher donor levels or into



Equation 7.29 is very similar to eqn 4.1 for the exciton binding energy except that the electron effective mass appears instead of the reduced electron–hole mass. This is because we are now considering the attraction of an electron to a heavy ion which is bound in the lattice, instead of that between a free electron and a free hole.

Fig. 7.10 Impurity absorption mechanisms in an n-type semiconductor: (a) transitions between donor levels; (b) transitions from the valence band to empty donor levels. The magnitude of the donor level energies has been exaggerated in this diagram to make the mechanisms clearer.

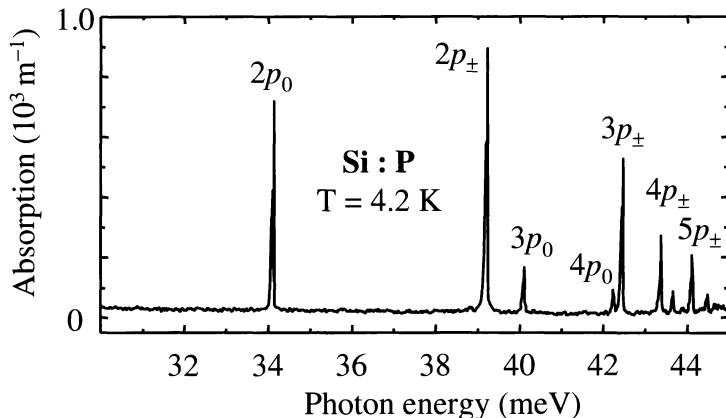


Fig. 7.11 Infrared absorption spectrum of n-type silicon doped with phosphorous at a density of $1.2 \times 10^{20} \text{ m}^{-3}$. The temperature was 4.2 K. After [8], copyright 1981 American Institute of Physics, reprinted with permission.

the conduction band by absorption of a photon. Figure 7.10(a) illustrates two possible transitions of this type, in which the electron is promoted to either the $n = 2$ or the $n = 3$ donor level. These transitions give rise to absorption lines analogous to the hydrogen Lyman series with frequencies given by:

$$h\nu = \frac{m_e^*}{m_0} \frac{R_H}{\epsilon_r^2} \left(1 - \frac{1}{n^2}\right), \quad (7.30)$$

where n is the quantum number of the final impurity level. If we insert typical values into eqn 7.30 we find that the photon energies are in the range 0.01–0.1 eV. This means that the transitions occur in the infrared spectral region.

Figure 7.11 shows the absorption spectrum of n-type silicon at liquid helium temperatures. The sample was doped with phosphorous at a density of $1.2 \times 10^{20} \text{ m}^{-3}$. The absorption lines correspond to transitions exciting electrons from the $n = 1$ shell to higher shells. In the language of atomic physics, these are $1s \rightarrow np$ transitions. These transitions converge at high n to the donor ionization energy of phosphorous in silicon, which is 45 meV.

The spectrum shown in Fig. 7.11 is actually more complicated than eqn 7.30 would suggest. It actually consists of two series of transitions, which are labelled as either np_0 or np_{\pm} . The np_0 series obey eqn 7.30 very well, but the np_{\pm} transitions have a different frequency dependence. This complexity is caused by the anisotropy of the effective mass of silicon. The frequency dependence of the two series can be modelled by assigning different effective Rydbergs for the ‘0’ and ‘ \pm ’ states. (See Exercise 7.13.)

We now consider briefly the absorption mechanism shown in Fig. 7.10(b). These transitions can be observed at temperatures when the donor levels are partly unoccupied due to the thermal excitation of the electrons into the conduction band. Absorption processes can then occur in which electrons are excited from the top of the valence band to the empty donor levels.

The valence band \rightarrow donor level transitions occur at photon energies just below the band gap E_g , with a threshold given by $E_g - E_1^D$. However, the transitions tend to be broadened into a continuum both by the thermal effects and by the fact that transitions can take place from a whole range of states within the valence band. Hence the impurity transitions cause a smearing of the abrupt absorption edge at the band gap found in pure semiconductors.

The absolute value of the absorption coefficient for the impurity transitions is around 10^3 m^{-1} . This is much smaller than for interband transitions which typically have values in the range 10^6 – 10^8 m^{-1} . However, if we were to assume that the absorption strength is simply proportional to the number of atoms that contribute, we would expect the impurity absorption to be weaker than the interband absorption by about a factor of $\sim 10^{-9}$. The measured ratio is much larger because the impurity lines are very sharp, whereas the interband transitions spread out into bands.

The absorption strength will always be weak compared to the interband and excitonic transitions due to the relatively small number of impurity atoms compared to the density of states within the conduction band. On the other hand, the transitions occur in the spectral region just below the band gap where we would normally expect no absorption at all. Hence these transitions do have an effect on the fundamental absorption edge, and make precise determinations of E_g from the absorption spectra at room temperature more difficult.

We have restricted our attention here exclusively to n-type semiconductors for the sake of simplicity. The same effects can of course occur in p-type materials.

7.5 Plasmons

Equation 7.7 tells us that the dielectric constant of a lightly damped gas of free electrons is expected to be zero at ω_p . This suggests that something unusual might happen at this frequency. This is indeed the case, as we discuss here.

A plasma system such as a metal consists of a fixed lattice of positive ions together with a gas of free electrons which exactly cancels the charge of the ions. Figure 7.12(a) illustrates a slab of the ions and electrons from within the bulk of the crystal. Consider a displacement of the whole free electron gas as illustrated in Fig. 7.12(b). The fixed lattice of positive ions will exert a restoring force to oppose this displacement of the electrons. This may cause the electrons to overshoot as shown in Fig. 7.12(c), which would then cause a restoring force in the opposite direction. The net result is that the whole electron gas can oscillate backwards and forwards with respect to the fixed lattice of positive ions. These oscillations are called **plasma oscillations**.

The frequency of the plasma oscillations can be calculated as follows. The displacement of the electrons by $\pm u$ illustrated in Figs. 7.12(b) and (c) gives rise to surface charges at either side of the slab. The surface charge per unit area is $-Neu$, where N is the number of electrons per unit volume in the metal. The unbalanced ions give an identical positive surface charge at the other end of the slab. Gauss's law tells us that the electric field \mathcal{E} is equal to Neu/ϵ_0 . The direction of this electric field is such as to oppose the displacement of the

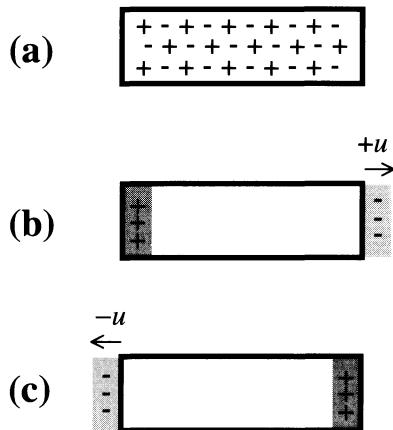


Fig. 7.12 Longitudinal plasma oscillations in a slab from within the bulk of a metal. At equilibrium (a) the charges of the positive ions and electrons cancel and the metal is neutral. Displacements of the electron gas as a whole in either direction are shown in (b) and (c). This gives rise to the positive and negative surface charges shown by dark and light shading respectively. The displacements lead to restoring forces that oppose the displacement and sustain oscillations at the plasma frequency.

electrons. The equation of motion for a unit volume of the electron gas is thus:

$$Nm_0 \frac{d^2u}{dt^2} = -Ne\mathcal{E} = -\left(\frac{N^2e^2}{\epsilon_0}\right)u. \quad (7.31)$$

This can be recast in the more familiar form:

$$\frac{d^2u}{dt^2} + \left(\frac{Ne^2}{\epsilon_0 m_0}\right)u = 0. \quad (7.32)$$

This describes the harmonic oscillations of the gas at an angular frequency given by $(Ne^2/\epsilon_0 m_0)^{\frac{1}{2}}$, that is, at the plasma frequency ω_p defined in eqn 7.6.

This simple analysis confirms that the plasma has a natural resonant frequency at ω_p . The polarization P (i.e. the dipole moment per unit volume) is equal to Neu and is in the opposite direction to the field. Hence the electric displacement $D = \epsilon_0\mathcal{E} + P$ is zero, which implies that the dielectric constant ϵ_r is zero. Thus the plasma oscillations correspond to oscillations in the conditions where $\epsilon_r = 0$.

Quantized plasma oscillations are called **plasmons**. These can be observed in metals by electron energy loss spectroscopy. A beam of electrons with energy $E_{in} \sim 2$ keV is fired at a thin sample. The electrons can excite plasmons as they pass through the metal. The energy E_{out} of the transmitted electrons will therefore be given by:

$$E_{out} = E_{in} - n\hbar\omega_p, \quad (7.33)$$

where n is the number of plasmons emitted during the passage through the sample. Thus by measuring the energy of the electrons emerging, the plasma frequency can be determined.

Plasmons can also be observed directly in doped semiconductors. Since the plasma frequencies are much lower, it is possible to use Raman scattering techniques to measure the plasmon energies. The general principles of Raman scattering will be discussed in Section 10.5. The basic point is that the photon energy changes as it traverses the sample through inelastic light scattering processes with the plasmons in the medium. Conservation of energy requires that the energy $\hbar\omega_{out}$ of the outgoing photon must satisfy:

$$\hbar\omega_{out} = \hbar\omega_{in} \pm \hbar\omega_p, \quad (7.34)$$

where $\hbar\omega_{in}$ is the energy of the incoming photon. The + sign corresponds to plasmon absorption and the - sign to plasmon emission.

Figure 7.13 shows the results of a Raman scattering experiment on n-type GaAs at 300 K. The doping density was $1.75 \times 10^{23} \text{ m}^{-3}$. The data is plotted as a function of the frequency shift of the light in wave number units. The data shows two clear peaks shifted by $\pm 130 \text{ cm}^{-1}$ relative to the incoming laser beam due to plasmon emission and absorption. The electron effective mass of GaAs is $0.067m_0$ and ϵ_{opt} is 10.6. Hence from eqn 7.24 we find $\omega_p = 2.8 \times 10^{13} \text{ Hz}$, which is equivalent to 150 cm^{-1} . The experimental data is thus in reasonably good agreement with the model.

Note that the plasma oscillations are *longitudinal*: the direction of the electric field is the same as the displacement of the wave. This is a rather surprising result, because Maxwell's equations usually imply that electric field waves have to be *transverse*. We will come across another situation like this when we discuss the optical properties of phonons in Section 10.2.2. We will see there that longitudinal electric field oscillations are only allowed at the frequency where the dielectric constant falls to zero.

In a III–V semiconductor such as GaAs, we would normally expect to observe Raman signals from optical phonons in the same energy range as the plasmon peaks. The two weak peaks in Fig. 7.13 at $\pm 272 \text{ cm}^{-1}$ and $\pm 296 \text{ cm}^{-1}$ are in fact caused by optical phonons. The Raman signals from the phonons are linearly polarized, and have been strongly suppressed in the data by use of orthogonal polarizers in front of the detectors.

It is very common to use wave number units in Raman spectroscopy. The wave number \bar{v} is equal to the reciprocal of the wavelength: $\bar{v} = 1/\lambda$. It is effectively a unit of energy with $1 \text{ cm}^{-1} \equiv 0.124 \text{ meV}$.

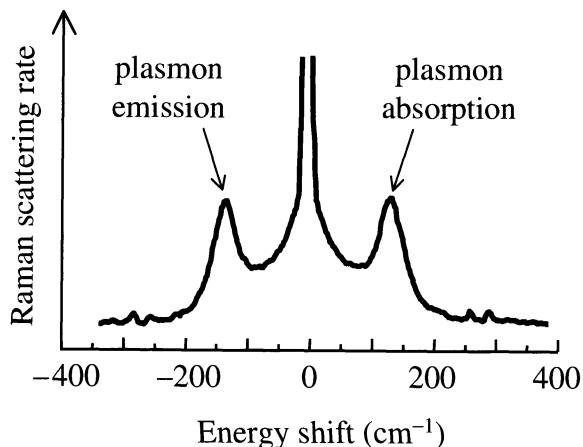


Fig. 7.13 Raman scattering measurements on n-type GaAs at 300 K. The doping density was $1.75 \times 10^{23} \text{ m}^{-3}$. The data is displayed as a function of the energy shift of the outgoing photons relative to the incoming ones in wave number units. After [9], copyright 1972 Excerpta Medica Inc., reprinted with permission.

Chapter summary

- Free electron effects are observed in metals and doped semiconductors. They can be modelled by the classical dipole oscillator model with no restoring force term. This approach is called the Drude–Lorentz model.
- The free electron plasma reflects strongly up to a specific frequency called the plasma frequency, which depends on the electron density. The damping rate of the oscillations is determined by the momentum scattering time deduced from electrical conductivity measurements.
- Metals reflect strongly due to the plasma reflectivity effect. At frequencies above the plasma frequency, the metals become transparent. This effect is called the ultraviolet transparency of metals.
- Interband transitions are possible in metals from states below the Fermi energy to empty levels above it. The interband absorption can reduce the reflectivity from the value predicted by the Drude–Lorentz model, and must therefore be considered to obtain a good fit to experimental reflectivity data.
- Doped semiconductors reflect at frequencies in the infrared due to the plasma reflectivity of the free electrons and holes generated by the doping process. Free carrier absorption effects can be observed at frequencies above the plasma frequency but below the fundamental absorption edge at the band gap.
- P-type semiconductors show an additional absorption mechanism in the infrared due to intervalence band transitions.
- Doped semiconductors show sharp infrared absorption lines due to impurity transitions at low temperatures. At room temperature, the impurity states broaden the fundamental absorption edge.
- Longitudinal plasma oscillations occur at the plasma frequency. The quantized oscillations are called plasmons. These can be observed by electron energy loss spectroscopy in metals, or by Raman scattering in doped semiconductors.

Further reading

The properties of electromagnetic waves in a conducting medium are covered in many electromagnetism and optics textbooks, for example Bleaney and Bleaney (1976), Born and Wolf (1999) or Hecht (1998).

The free carrier model of metals is covered in Singleton (2001). It is also covered in Ashcroft and Mermin (1976), Burns (1985) or Kittel (1996). Kittel gives a good discussion of plasmons in metals.

Free carrier reflectivity and absorption in semiconductors has been reviewed by Pidgeon (1980), and is also covered by Yu and Cardona (1996). Yu and Cardona give further details about intervalence band and impurity absorption.

References

- [1] Wyckoff, R.W.G. (1963). *Crystal structures* (2nd edn). Interscience, New York.
- [2] Givens, M. P. (1958). *Solid state physics*, Vol. 6 (ed. F. Seitz and D. Turnbull). Academic Press, New York, p. 313.
- [3] Ehrenreich, H., Philipp, H.R. and Segall, B. (1962). *Phys. Rev.*, **132**, 1918.
- [4] Segall, B. (1961). *Phys. Rev.*, **124**, 1797.
- [5] Moruzzi, V.L., Janak, J.F. and Williams, A.R. (1978). *Calculated electronic properties of metals*. Pergamon Press, New York.
- [6] Lide, D.R. (1996). *CRC handbook of chemistry and physics* (77th edn). CRC Press, Boca Raton.
- [7] Spitzer, W.G. and Fan, H.Y. (1957). *Phys. Rev.*, **106**, 882.
- [8] Jagannath, C., Grabowski, Z.W. and Ramdas, A.K. (1981). *Phys. Rev. B*, **23**, 2023.
- [9] Mooradian, A. (1972). In *Laser handbook* Vol. II (ed. F.T. Arecchi and E.O. Schulz-duBois). North Holland, Amsterdam, p. 1409.

Exercises

- (7.1) Derive a relationship between the Fermi energy E_F of a metal and its plasma frequency ω_p .
- (7.2) The ionosphere reflects radio waves with frequencies up to about 3 MHz, but transmits waves with higher frequencies. Estimate the free electron density in the ionosphere.
- (7.3) Estimate the skin depth of radio waves of frequency 200 kHz in sea water, which has an average electrical conductivity of about $4\Omega^{-1}\text{m}^{-1}$. Hence discuss the difficulties you might encounter when attempting to communicate with a submerged submarine using radio waves.
- (7.4) Cesium metal is found to be transparent to electromagnetic radiation of wavelengths below 440 nm. Calculate a value for the electron effective mass using the data given in Table 7.1.
- (7.5) The momentum scattering time of silver is 4.0×10^{-14} s at room temperature. Calculate the dielectric constant at 500 nm, neglecting interband absorption effects. Hence estimate the reflectivity of a silver mirror at this wavelength. See Table 7.1 for the plasma frequency of silver.
- (7.6) Estimate the fraction of light with wavelength 1 μm that is transmitted through a 20 nm thick gold film at 77 K, where the DC electrical conductivity is $2 \times 10^8 \Omega^{-1}\text{m}^{-1}$. The plasma frequency and electron density of gold are given in Table 7.1.

- (7.7) Figure 7.14 shows the measured reflectivity of gold in the wavelength range 100–1000 nm. Account qualitatively for the shape of the spectrum, and deduce the energy gap between the *d* bands and the Fermi energy. Use the data to explain the characteristic colour of gold.

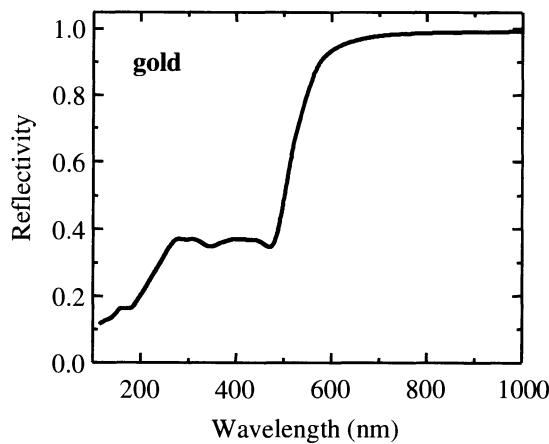


Fig. 7.14 Reflectivity of gold in the wavelength range 100–1000 nm. After [6].

- (7.8) What is the value of the dielectric constant of a medium that has zero reflectivity? Use eqn 7.22 to show that the reflectivity of a lightly damped doped semiconductor is zero at the angular frequency given in eqn 7.25.
- (7.9) Use the data shown in Fig. 7.7 to deduce the value of the electron effective mass of InSb at each carrier density. Take $\epsilon_{\text{opt}} = 15.6$.
- (7.10) The absorption coefficient at room temperature of an n-type sample of InAs with a doping level of $1.4 \times 10^{23} \text{ m}^{-3}$ is found to be 500 m^{-1} at $10 \mu\text{m}$. Estimate the momentum scattering time, given that the electron effective mass is $0.023m_0$ and the refractive index is 3.5.
- (7.11)* A laser beam operating at 632.8 nm with an intensity of 10^6 W m^{-2} is incident on a sample of pure InP at room temperature. The absorption coefficient at this wavelength is $6 \times 10^6 \text{ m}^{-1}$, and the carrier lifetime is 1 ns. Estimate the free carrier absorption coefficient at the wavelength of a CO₂ laser (10.6 μm), where the refractive index is 3.1. The effective mass and momentum scattering time for the electrons are $0.08m_0$ and $2 \times 10^{-13} \text{ s}$, while the equivalent values for the holes are $0.6m_0$ and $5 \times 10^{-14} \text{ s}$.
- (7.12)* Consider the intervalence band processes illustrated in Fig. 7.9 for a heavily doped p-type sample of GaAs

containing $1 \times 10^{25} \text{ m}^{-3}$ acceptors. The valence band parameters for GaAs are given in Table C.2.

- (i) Work out the Fermi energy in the valence band on the assumption that the holes are degenerate. What are the wave vectors of the heavy and light holes at the Fermi energy?
- (ii) Calculate the upper and lower limits of the photon energies for the three absorption processes labelled (1), (2) and (3) in Fig. 7.9, namely the $\text{lh} \rightarrow \text{hh}$, the $\text{SO} \rightarrow \text{lh}$ and the $\text{SO} \rightarrow \text{hh}$ transitions.
- (7.13) Figure 7.11 shows the infrared absorption spectrum of n-type silicon, which has a dielectric constant of 16. Two series of lines labelled $n p_0$ and $n p_{\pm}$ are identified in the data.
- (i) Show that the $n p_0$ series is consistent with eqn 7.30, and deduce a value for the electron effective mass for these transitions.
- (ii) Show that the $n p_{\pm}$ follows the following formula:
- $$h\nu = \frac{R_0^*}{1^2} - \frac{R_{\pm}^*}{n^2},$$
- stating the values of R_0^* and R_{\pm}^* you find from the data.
- (7.14) It is found that the infrared absorption spectrum of a lightly doped n-type semiconductor with a relative dielectric constant of 15.2 consists of a series of sharp lines at low temperatures. The energies of the lines are given by:
- $$E(n) = R^*(1 - 1/n^2)$$
- where R^* is 2.1 meV and n is an integer greater than 1. Explain why the energies of the absorption lines are almost independent of the type of impurity atoms used for the doping, and deduce a value for the electron effective mass.
- (7.15) Explain why the fundamental absorption edge of InSb at 10 K shifts from $5.26 \mu\text{m}$ to $5.44 \mu\text{m}$ when doped with zinc acceptors, and deduce a value for the ground state acceptor level energy relative to the valence band.
- (7.16) An argon ion laser operating at 514.5 nm is incident on an n-type GaAs sample. A peak is observed in the intensity of the scattered light at 534.3 nm. Explain this observation, and estimate the electron density, given that $m_e^* = 0.067m_0$ and $n = 3.3$.
- (7.17) Calculate the doping density at which the plasmons in n-type GaAs have the same wave number as the longitudinal optic phonon at 297 cm^{-1} . Take $m_e^* = 0.067m_0$ and $n = 3.3$.

* Exercises marked with an asterisk are more challenging.

8

Molecular materials

In this chapter we will be considering the optical properties of solid state materials based on organic molecules. This subject was introduced in Section 1.4.4, and was also briefly mentioned when we considered Frenkel excitons in organic crystals in Section 4.5.3. It is a topic which has come to the fore during the 1990s, mainly as a result of technological advances that have led to the development of a whole range of organic electroluminescent devices.

The optical physics of organic materials is different to that of the inorganic crystals that we have been considering up to this point in two significant ways.

- (1) The electronic and vibrational states tend to be localized near individual molecules, which means that the properties of the solids are quite similar to those of the constituent molecules. The solid state mainly provides a convenient way to incorporate large densities of light-emitting molecules into an electroluminescent device, without necessarily leading to substantially new physical effects.
- (2) The optical transitions are vibronic in character: the transition simultaneously changes the electronic and vibrational state of the molecule. This is equally true for isolated molecules and for the solids that we will be considering here. In the solids, the vibronic aspect of the transitions is somewhat affected by the increased number of vibrational modes that are present compared to isolated molecules.

Both of these points make it clear that we need to understand the physics of electronic transitions in isolated molecules first before we can properly explain the optical properties of molecular solids. Therefore, after first giving an overview of the types of material that we will be considering here, we start the chapter with a brief review of the electronic states in conjugated molecules. We then consider the basic principles of vibronic transitions in simple isolated molecules. This will provide us with a good introduction to the physics of the molecular solids that are considered in the rest of the chapter, and leads us to our goal of explaining the physical principles that underpin the burgeoning subject of molecular optoelectronics. The physics of the vibronic transitions studied here will also serve as a useful introduction for the other types of vibronic systems that we will be considering in Chapter 9.

8.1	Introduction to molecular materials	166
8.2	Electronic states in conjugated molecules	167
8.3	Optical spectra of molecules	169
8.4	Aromatic hydrocarbons	177
8.5	Conjugated polymers	179
8.6	Organic optoelectronics	181

8.1 Introduction to molecular materials

The difference between saturated and conjugated organic molecules will be explained in Section 8.2 below.

The name ‘aromatic’ originates from the fact that the liquids and gases tend to have strong aromas.

The terminology ‘monomer’ and ‘polymer’ will be explained in Section 8.5 when we consider the properties of conjugated polymers.

The molecular solids that we will be considering in this chapter are all based on conjugated organic compounds. The reason for this restriction of the scope was explained previously in Section 1.4.4: solids based on inorganic compounds are considered elsewhere in this book, while the optical physics of saturated organic compounds is not significantly different to that of insulators and glasses.

By restricting our attention to conjugated organic compounds, we will be concentrating on materials with delocalized π orbitals. The two main examples that we will consider are:

- aromatic hydrocarbons;
- conjugated polymers.

The aromatic hydrocarbons are carbon–hydrogen compounds containing benzene rings in their structure. Conjugated polymers, on the other hand, are long chain molecules made by bonding many individual individual molecular units (‘monomers’) together. The aromatic hydrocarbons have cyclic conjugation, while the polymers have linear conjugation. This distinction, and the importance of π electrons in conjugated systems, will be explained when we consider the electronic states of molecules in Section 8.2 below.

Single crystals of some conjugated organic materials have been prepared, most notably for the aromatic hydrocarbons and simple polymers such as polydiacetylene. In many other compounds, however, the solid phase is amorphous, and the samples are made by spin-coating thin films onto glass substrates from the solution. Thus the molecular solids that we will be considering here will sometimes fall within the category of traditional solid state physics based on crystals, but more usually we will be within the realms of ‘soft’ condensed matter.

The molecular solids are formed by condensation of neutral organic compounds and are held together by van der Waals type interactions. These interactions are relatively weak compared to the interactions within the molecule itself, which originate from the strong covalent bonds between the atoms. This is exemplified by the low melting point of organic solids and their generally soft structure. It means that the electronic states remain tightly bound to the constituent molecules. Therefore, we will be dealing mainly with localized electronic states. This contrasts with the delocalized band states that we have been considering in Chapters 3–7.

The same conclusion can be reached for the vibrational states. The most important phonons of molecular solids are localized modes with discrete frequencies. These are essentially just the vibrational modes of the constituent molecules, perhaps with their frequencies slightly altered in the condensed phase. There will, of course, be delocalized phonon bands present in molecular crystals, but these play only a secondary role in the optical properties in most instances.

As we have already noted in the preamble, the localized nature of the electronic and vibrational states means that the optical properties of the solid are rather similar to those of the individual molecules. We must therefore study the electronic states and optical properties of individual molecules before we can understand molecular solids in any detail. This is the subject of the next two sections.

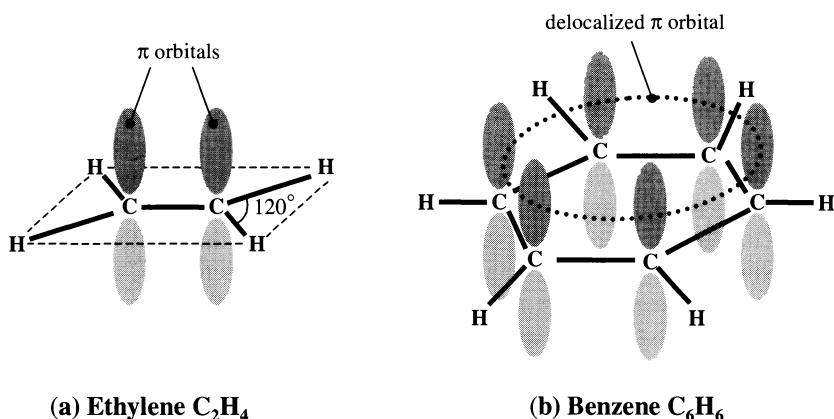


Fig. 8.1 (a) The ethylene molecule (C_2H_4). The carbon and hydrogen atoms lie within a plane defined by the σ bonds shown by the thick black lines. The π orbitals lie above and below this plane. (b) The benzene molecule (C_6H_6). The π electrons form a delocalized ring orbital above and below the plane of the hexagon defined by the six carbon atoms.

8.2 Electronic states in conjugated molecules

In this section we will give a brief summary of the electronic states that determine the optical physics of conjugated molecular solids. The reader is referred to the bibliography listed in the Further Reading section for a more thorough exposition of this subject.

The chemistry of organic molecules is based on the covalent bonds between the carbon atoms. Carbon has an atomic configuration of $1s^2 2s^2 2p^2$, with four valence electrons in the $n = 2$ atomic shell. In elemental carbon (diamond), each carbon atom forms single covalent bonds with four adjacent carbon atoms. In organic compounds there may be single, double, or triple bonds between adjacent carbon atoms. In molecules with double or triple bonds, the valence electrons are divided between the σ bonds and the π bonds. It is easiest to explain how this works with the help of specific examples.

Consider the ethylene ($H_2C=CH_2$) molecule shown in Fig. 8.1(a). Each carbon atom is bonded to two hydrogen atoms, and has a double bond with the other carbon atom. The two $2s$ electrons hybridize with one of the $2p$ electrons to form three sp^2 bonds. These are the σ bonds and are arranged at an angle of about 120° to each other. This means that the carbon and hydrogen atoms lie within a plane. The other $2p$ electron forms a π orbital derived from the $2p_z$ atomic orbital. The wave functions have lobes above and below the plane defined by the nuclei. The electrons in these π orbitals are called **π electrons**. The overlap of the π orbitals produces the second bond between the two carbon atoms.

Consider now the benzene molecule (C_6H_6) shown in Fig. 8.1(b). This is also a planar molecule, with the six carbon atoms arranged as a hexagon. Each carbon atom forms sp^2 σ bonds with one hydrogen atom and its two adjacent carbon atoms. The π electrons now form a ring orbital above and below the plane of the hexagon.

The chemical structure of benzene is traditionally drawn as a hexagon with alternating double and single bonds between the carbon atoms. (See for example, Fig. 8.11.) We know in reality that this does not occur, and the π electrons are shared equally between the two bonds on either side of each carbon atom. Organic molecules like benzene with alternating double and single bonds are

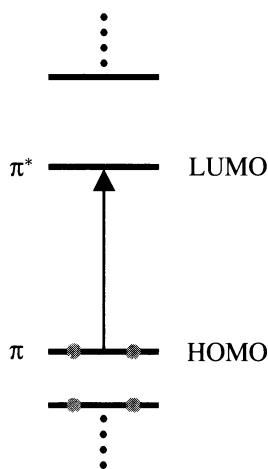


Fig. 8.2 Schematic energy level diagram for a conjugated molecule. The lowest energy electronic transition takes place between the HOMO (highest occupied molecular orbital) and the LUMO (lowest unoccupied molecular orbital) states. This is a $\pi \rightarrow \pi^*$ transition.

Transitions to the σ^* excited states of the σ electrons occur at much higher energies, because it takes a large amount of energy to break a σ bond.

said to be **conjugated**. These contrast with other large molecules with only single bonds between the carbons, which are called **saturated**.

The π electrons in conjugated compounds are able to spread out in large delocalized orbitals. For example, the formation of the ring orbital in benzene allows the π electrons to spread out much more than the electrons in the σ bonds. If we apply the ideas of quantum confinement developed in Section 6.1, we can understand that the spreading of the π electron wave function leads to a reduction in the energy. This is apparent from the fact that the lowest electronic transition of ethylene occurs at around 6.9 eV, whereas the equivalent transition in benzene occurs at 4.6 eV. If we use larger molecules with more delocalized π electrons, we can reduce the confinement energy even more, and push the transition energies down into the visible spectral region. This is why the transitions of the π electrons in large conjugated molecules are particularly interesting.

The benzene ring is an example of a **cyclic conjugated** molecule. The name follows from the fact that the electron wave functions must have cyclic periodicity around the closed ring. It is also possible to form **linear conjugated** molecules, in which the π electrons delocalize along a chain rather than into a ring. The conjugated polymers that we will consider in Section 8.5 are good examples. It is easy to apply the ideas of quantum confinement to these linear molecules and thus to obtain a rough estimate of the electron energy. This approach shows that we need about seven repeat units to reduce the transition energies down into the visible spectral region. (See Exercise 8.2.)

The electronic states of the molecule can be arranged in order of increasing energy in much the same way as we do for atoms. Figure 8.2 gives a schematic diagram of how this looks for a typical conjugated molecule. The electrons from the constituent atoms of the molecule fill up the molecular orbitals until they are all paired off in bonds. The highest filled energy level is called the **HOMO** level (highest occupied molecular orbital). In the case of a conjugated molecule, this will be a π orbital, because the electrons in the σ bonds are very tightly bound. The first energy level above the HOMO state is called the **LUMO** level (lowest unoccupied molecular orbital). This will be an excited configuration of the π orbitals, and is labelled a π^* state. The lowest energy transition therefore involves the promotion of a π electron to a π^* state, and is thus labelled a $\pi \rightarrow \pi^*$ transition.

The electrons in the ground state of a molecule are all paired off in bonds with their spins antiparallel. This means that the ground state HOMO level has a spin quantum number S equal to 0. The excited states, however, can either have $S = 0$ or $S = 1$. This is because the excitation process puts an unpaired electron in the excited state and leaves an unpaired electron in the HOMO state. According to the rules of addition of angular momenta, the two spin $\frac{1}{2}$ electrons can combine to give a total spin of either 0 or 1. This point is illustrated in Fig. 8.3. The multiplicity of the spin states is equal to $(2S + 1)$, because there are $(2S + 1)$ degenerate m_s levels in each state. Hence the $S = 0$ states are known as **singlets**, while the $S = 1$ states are called **triplets**. Triplets tend to have lower energies than the singlet counterparts.

The separation of the electronic levels into singlet and triplet states has very important consequences on the optical spectra. Each molecule will have a series of singlet excited states labelled S_1, S_2, S_3, \dots , in addition to its

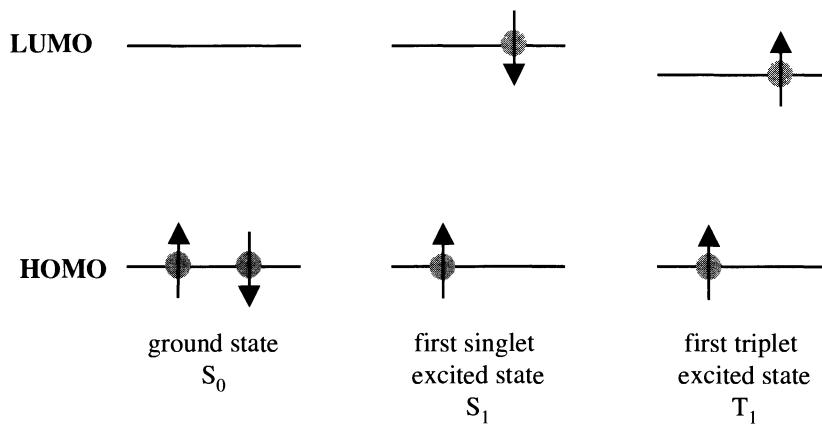


Fig. 8.3 The unpaired electrons of an excited molecule can either have their spins antiparallel or parallel. The states with antiparallel spins have spin quantum number $S = 0$ and are called singlets after their multiplicity. The states with parallel spins have $S = 1$ and are called triplets. The ground state is always a singlet.

singlet ground state which is labelled S_0 . There will be a similar series of triplet excited states labelled T_1 , T_2 , T_3 , ... Since photons carry no spin, they can only excite transitions between electronic states of the same spin. Therefore, transitions from the S_0 ground state to the triplet excited state are not allowed. The main optical absorption edge therefore corresponds to the $S_0 \rightarrow S_1$ singlet–singlet transition. The emission spectrum is likewise dominated by the $S_1 \rightarrow S_0$ transition.

The singlet excited states have short lifetimes of order 1–10 ns due to the dipole-allowed transitions to the S_0 ground state. The lowest triplet state, on the other hand, has a long radiative lifetime because of the low probability for the $T_1 \rightarrow S_0$ transition. The different time scales for the singlet–singlet and triplet–singlet transitions are conveniently distinguished by describing the emission processes as **fluorescence** and **phosphorescence**, respectively. As mentioned in Section B.3, this distinction is based on whether the emission is fast or slow, with the dividing line drawn at around 10^{-8} s. A schematic diagram of the two types of emission processes is given in Fig. 8.12.

We might have expected the transition probability to be exactly zero due to the spin selection rule. However, spin–orbit coupling can mix a small amount of singlet character into the triplet states and allows some probability for triplet–singlet transitions.

8.3 Optical spectra of molecules

The optical properties of molecules are generally divided into three spectral regions:

- The far-infrared spectra: wavelength $\lambda > 100 \mu\text{m}$.
- The infrared spectra: $\lambda \sim 1 - 100 \mu\text{m}$.
- The visible and ultraviolet spectra: $\lambda < 1 \mu\text{m}$.

These three spectral regions correspond respectively to transitions between the rotational, vibrational and electronic states of the molecule. In this chapter we will be concerned only with the visible/ultraviolet spectra related to the electronic transitions. As we will see, the molecule can also change its vibrational state during the transition, and thus we must understand the concept of **vibrational–electronic** or **vibronic** transitions. The molecule can, of course, also change its rotational state during an electronic transition, but since the scale of the rotational energies is so much smaller than that of the electronic energies, we will not need to be concerned with that much detail here.

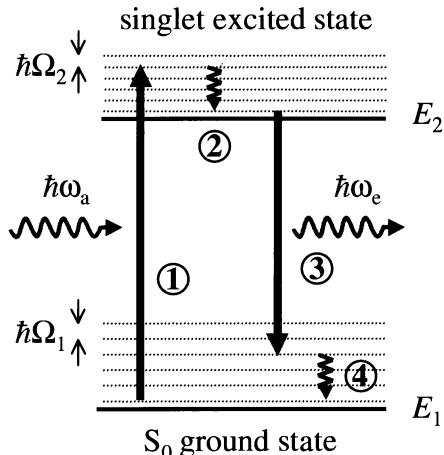


Fig. 8.4 Schematic diagram of the vibrational-electronic transitions in a molecule. The four processes indicated are respectively: (1) absorption; (2) non-radiative relaxation; (3) emission and (4) non-radiative relaxation.

8.3.1 Electronic-vibrational transitions

The basic physics of the vibrational-electronic spectra of molecules can be understood with reference to Fig. 8.4. The diagram shows absorption and emission transitions between two electronic states of a molecule with energies of E_1 and E_2 respectively. For simplicity, we assume that the lower state is the S_0 ground state, and the upper state is a singlet excited state with allowed electric-dipole transitions from S_0 .

The atoms in a molecule can vibrate about their bonds, which gives the molecule vibrational energy in addition to its electronic energy. Hence we must associate a series of vibrational levels with each electronic state, as shown in Fig. 8.4. Quantum mechanics tells us that the energy of a vibrational oscillation of angular frequency Ω is equal to $(n + \frac{1}{2})\hbar\Omega$, where n is the number of quanta excited. Thus the energy of the molecule in the ground state level when n_1 quanta of frequency Ω_1 are excited will be given by:

$$E = E_1 + (n_1 + \frac{1}{2})\hbar\Omega_1. \quad (8.1)$$

In the same way, the energy of the molecule in the excited electronic state with n_2 quanta of frequency Ω_2 excited will be given by:

$$E = E_2 + (n_2 + \frac{1}{2})\hbar\Omega_2. \quad (8.2)$$

The subscripts allow for the possibility that the vibrational frequencies may be different for the two electronic states.

We consider an optical transition in which an electron is promoted from the ground state to the excited state by absorbing a photon. We assume that the molecule is initially in the $n_1 = 0$ vibrational level of the ground state. The transition is indicated by process (1) in Fig. 8.4. Conservation of energy requires that:

$$\begin{aligned} \hbar\omega_a &= (E_2 + (n_2 + \frac{1}{2})\hbar\Omega_2) - (E_1 + \frac{1}{2}\hbar\Omega_1) \\ &= \hbar\omega_0 + n_2\hbar\Omega_2, \end{aligned} \quad (8.3)$$

The energy of the vibrational quanta of a typical molecule are of order ~ 0.1 eV. Thus at room temperature, where $k_B T \sim 0.025$ eV, there will be relatively few thermally excited vibrational quanta. (See, for example, Exercise 8.3.)

where ω_a is the angular frequency of the absorbed photon and

$$\hbar\omega_0 = (E_2 - E_1 + \frac{1}{2}\hbar(\Omega_2 - \Omega_1)). \quad (8.4)$$

This vibrational–electronic process causes the electron to jump to the excited electronic state, and simultaneously creates vibrational quanta. Since n_2 can only take integer values, the absorption spectrum will in principle consist of a series of discrete lines with energy given by eqn 8.3.

The absorption transition leaves the molecule in the excited electronic state with a large amount of vibrational energy. This excess vibrational energy is rapidly lost in radiationless relaxation processes, as indicated by the wiggly arrow labelled (2) in Fig. 8.4. The relaxation process occurs by spreading the vibrational energy of the individual excited molecule throughout the rest of the system. (See the discussion in Section 8.3.3 below.) Ultimately, the excess vibrational energy ends up as heat.

Once the molecule has relaxed to the bottom of the excited state, it then returns to the ground state by emitting a photon at energy $\hbar\omega_e$, as shown by process (3) in Fig. 8.4. This leaves the molecule in an excited vibrational level of the ground state. The frequency of the photon is given by:

$$\begin{aligned} \hbar\omega_e &= (E_2 + \frac{1}{2}\hbar\Omega_2) - (E_1 + (n_1 + \frac{1}{2})\hbar\Omega_1) \\ &= \hbar\omega_0 - n_1\hbar\Omega_1. \end{aligned} \quad (8.5)$$

Thus the emission spectrum consists of a series of vibrational–electronic lines with frequency given by eqn 8.5. The molecule finally returns to the $n_1 = 0$ level of the ground state by losing the excess vibrational quanta in further radiationless relaxation processes, as shown by step (4) in Fig. 8.4.

On comparing eqns 8.3 and 8.5, we see that the absorption occurs at a higher energy than the emission, except for the cases when no vibrational quanta are excited during the electronic transitions. This is a very common phenomenon, and should be contrasted with atomic transitions, in which the absorption and emission frequencies coincide. The difference in energy between the maximum absorption and the maximum emission is called the *Stokes shift*.

8.3.2 Molecular configuration diagrams

The vibrational–electronic spectra of molecules can be understood in more detail with the aid of **configuration diagrams**. These are diagrams which show the electronic energy of a molecule as a function of the **configuration coordinates**. In order to understand how these diagrams work, we will first consider the simplest type of molecule, namely a diatomic molecule.

The electronic energy of a diatomic molecule is usually calculated by applying the **Born–Oppenheimer approximation**. This says that the electronic and nuclear motions are independent, and means that we can draw graphs of the electronic energy as a function of the internuclear separation. The approximation is valid because the nuclei are much heavier than the electrons, and therefore move on a far slower time scale.

Figure 8.5 shows a schematic configuration diagram of a typical diatomic molecule. The diagram shows the energy of the ground state and one of the excited states as a function of the separation r between the two nuclei. If the

The discrete lines described by eqn 8.3 are frequently broadened into a continuum. See the discussion of the experimental data in Section 8.3.4.

In diatomic molecules there is only one vibrational mode possible, namely the stretching of the bond between the two atoms. The vibrational configuration of the molecule can therefore be given a direct physical interpretation as the internuclear separation.

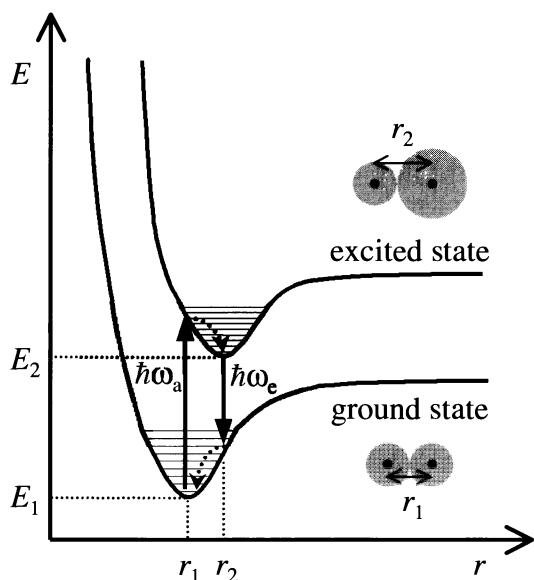


Fig. 8.5 Energy level diagram for the ground state and an excited state of a simple diatomic molecule, as a function of the separation r between the two nuclei. Vibrational-electronic absorption and emission transitions at energies of $\hbar\omega_a$ and $\hbar\omega_e$ are indicated. The schematic ‘dumb-bell’ diagrams of the molecule, with the radius of one of the atoms increasing in the excited state, illustrate the point that the equilibrium separations of the nuclei in the two electronic states are different.

The validity of the Born–Oppenheimer approximation is demonstrated experimentally by the fact that the vibrational oscillations of the molecule (which are caused by displacements of the nuclei from their equilibrium positions) occur at frequencies in the infrared spectral range ($\sim 10^{13}$ Hz), whereas the electronic transitions occur in the visible and ultraviolet spectral regions (10^{14} – 10^{15} Hz).

states are bound, there must be a minimum energy for some value of r . The position of the minimum in the ground state is labelled r_1 and corresponds to the equilibrium separation of the nuclei in the unexcited molecule. The minimum at r_2 is the mean separation of the nuclei when the molecule is in the excited electronic state. In general, r_1 and r_2 are not the same.

We can understand why the minima occur at different positions by discussing the behaviour of the simplest diatomic molecule, namely hydrogen. Consider the ground state of the H_2 molecule. When $r = \infty$, the atoms are independent of each other, and the ground state energy is that of two separate hydrogen atoms each in the $1s$ level. As r decreases from ∞ , the total energy of the system must decrease due to the cohesive energy of the H–H covalent bond. However, if r becomes too small, the energy will increase again due to the electron–electron and proton–proton repulsion. Therefore, the energy of the system must go through a minimum at some value of r , labelled r_1 , and then increase strongly for smaller r . r_1 is the equilibrium separation of the nuclei in the ground state. For the H_2 molecule, $r_1 = 0.074$ nm.

Now consider the energy of the first excited bound electronic state of the H_2 molecule. At $r = \infty$, this corresponds to one atom being in the $1s$ level and the other in the $2p$ level. The energy of the system will at first decrease with r due to the attractive forces between the atoms. We then go through another minimum labelled r_2 in Fig. 8.5 as the repulsive forces for small r become significant. In general, r_2 will not be equal to r_1 because the minimum energy is obtained when we maximize the cross-attractions between the electron of one atom and the proton of the other, while minimizing the sum of the proton–proton and electron–electron repulsions. This process obviously depends on the overlap of the electronic wave functions, which will be different for the orbitals of the $1s^2$ ground state and those of the $1s\ 2p$ excited state. In the case of the first excited state of H_2 , the energy minimum occurs at $r_2 = 0.13$ nm, which is substantially greater than r_1 .

Optical transitions from the ground state to the $1s\ 2p$ excited state of H_2 occur at 11.3 eV, which is slightly larger than the Lyman α line in atomic hydrogen: see Exercise 8.4.

The difference between r_1 and r_2 can be given a very simple interpretation with the aid of the schematic ‘dumb-bell’ pictures of the molecules which are included in Fig. 8.5. The equilibrium separation of the nuclei corresponds to the point where the atomic orbitals begin to touch and bond together. In the ground state, both atoms have the same radius, but in the excited state one of them has a larger radius. It is thus obvious that r_2 will be greater than r_1 .

The dependence of the electronic energy on the position coordinates shown in Fig. 8.5 is typical of other molecules, and can be used as a starting point for the discussion of the vibrational modes. The vibrational motion of the molecule will be determined by the shape of the $E(r)$ curve. Although the detailed functional form of $E(r)$ is complicated, it can be shown that for small displacements from the minimum position, the curve can always be approximated by a parabola. (See, for example, Exercise 8.5.) Therefore, near the minimum we can write:

$$E(x) = E_{\min} + \frac{1}{2}\mu\Omega^2x^2 \quad (8.6)$$

where μ is the reduced mass of the molecule. In the case of the ground state, $x = r - r_1$, while for the excited state, $x = r - r_2$. Equation 8.6 describes a simple harmonic oscillator of frequency Ω . The quantized vibrations about the equilibrium position at $x = 0$ associates a series of uniformly spaced energy levels with each electronic state as shown in Fig. 8.5. The separation of the vibrational levels will be different for each electronic state because the value of Ω depends on the curvature of the $E(r)$ curve (see Exercise 8.5), which will differ from state to state.

8.3.3 The Franck–Condon principle

The optical transitions between the coupled vibrational–electronic levels of a molecule can be understood by invoking the Franck–Condon principle. This says that the electronic transitions take place so rapidly that the nuclei do not move significantly during the transition. The Franck–Condon principle follows from the Born–Oppenheimer approximation, and is a consequence of the fact that electrons are much lighter than the nuclei.

The steps that take place when photons are absorbed and re-emitted by a molecule according to the Franck–Condon principle are illustrated schematically in Fig. 8.6. The molecule starts in the ground state with a mean nuclear separation of r_1 . The absorption of the photon promotes an electron to the excited state without altering r . The transition thus leaves the molecule in the excited state with a mean nuclear separation of r_1 instead of the equilibrium separation r_2 . The separation of the nuclei rapidly relaxes to r_2 before re-emitting a photon. This leaves the molecule in the ground state with a mean nuclear separation of r_2 . Further rapid relaxation processes occur to complete the cycle and bring the molecule back to its equilibrium separation in the ground state. These four steps correspond to the four processes indicated in the simplified energy level diagram shown in Fig. 8.4.

The ‘rapid relaxation’ processes that accompany the optical transitions need some clarification. If we think of the nuclear vibrations as analogous to those of a spring, we can see that the transition leaves the molecular spring in a compressed or extended stationary state at time $t = 0$. We know that in this

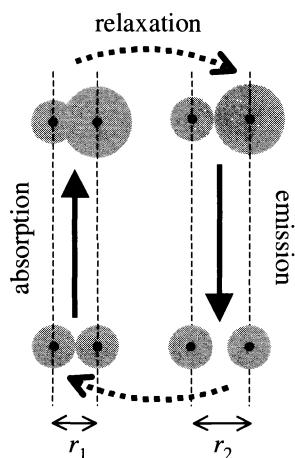


Fig. 8.6 Schematic representation of the processes that occur during the absorption and emission of photons by vibronic transitions in a molecule. The Franck–Condon principle states that the nuclei do not move during the optical transitions. r_1 and r_2 are the equilibrium separations of the nuclei in the ground and excited states respectively. One of the atoms has a larger radius in the excited molecule because the atom itself is in an excited state.

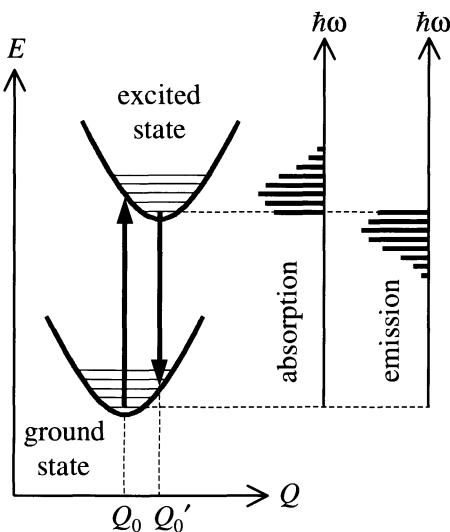


Fig. 8.7 Configuration diagram for two electronic states in a molecule. Vibrational-electronic transitions are indicated by the vertical arrows, together with a schematic representation of the absorption and emission spectrum. The probability amplitudes for the relevant vibrational levels for the absorption transition are shown in Fig. 8.8.

situation the spring will immediately begin to vibrate at its own natural frequency for $t > 0$. This is equivalent to instigating the oscillations of a specific vibrational mode in one particular molecule. However, the molecule may have other vibrational modes, and it can also interact with the other molecules that surround it. The relaxation of the vibrational energy created during the transition thus involves the spreading of the localized vibrational energy of a particular molecule to the other modes of the molecule and to the other molecules. This is a more technical way of saying that the excess energy ends up as heat. The vibrational relaxation typically occurs in less than 1 ps in a solid, which is much faster than the ~ 1 ns taken to re-emit a photon.

The Franck–Condon principle implies that we represent the optical transitions by vertical arrows in configuration diagrams, as shown in Fig. 8.5. The absorption of a photon puts the molecule in an excited vibrational state as well as an excited electronic state. The excess vibrational energy is lost very rapidly through non-radiative relaxation processes, as indicated by the dotted lines in Fig. 8.5. The frequencies of the photons absorbed and emitted are given by eqns 8.3 and 8.5. These describe a series of sharp lines with equal energy spacing.

In more complicated molecules with many degrees of freedom, the vibrational motion is described in terms of the normal modes of the coupled system. These vibrational modes are usually represented by a generalized coordinate Q , which has the dimensions of length. The Born–Oppenheimer approximation allows us to produce configuration diagrams in which we plot the electronic energy as a function of Q . Figure 8.7 is an example of such a configuration diagram. In general, the ground state and excited state have approximately parabolic minima at different values of the configuration coordinate. The optical transitions are indicated by vertical arrows, as prescribed by the Franck–Condon principle. The absorption and emission spectra consist of a series of lines with frequencies given by eqns 8.3 and 8.5, as shown in the right-hand side of the figure.

The relative intensities of the manifold of vibronic transitions can be calculated in the Franck–Condon approximation. The matrix element for an electric-dipole transition from an initial state Ψ_1 to a final state Ψ_2 is given by (c.f. eqn B.25) :

$$M_{12} = \langle 2 | e\mathbf{r} \cdot \boldsymbol{\epsilon}_0 | 1 \rangle \equiv \int d\xi_1 \cdots \int d\xi_N \Psi_2^* e\mathbf{r} \cdot \boldsymbol{\epsilon} \Psi_1, \quad (8.7)$$

where \mathbf{r} is the position vector of the electron, $\boldsymbol{\epsilon}$ is the electric field of the light wave, and ξ_1, \dots, ξ_N represent the coordinates for all the relevant internal degrees of freedom of the molecule. For the coupled vibrational–electronic states that we are considering here, the total wave function will be a product of an electronic wave function that depends only on the electron coordinate \mathbf{r} , and a vibrational wave function that depends only on the configuration coordinate Q . We thus write the vibronic wave function for an electronic state i and a vibrational level n as:

$$\Psi_{i,n}(\mathbf{r}, Q) = \psi_i(\mathbf{r}) \varphi_n(Q - Q_0). \quad (8.8)$$

The vibrational wave function $\varphi_n(Q - Q_0)$ is just the wave function of a simple harmonic oscillator centred at Q_0 , the equilibrium configuration for the i th electronic state. On inserting the vibronic wave functions from eqn 8.8 into the matrix element given in eqn 8.7, we find:

$$M_{12} \propto \int \int \psi_2^*(\mathbf{r}) \varphi_{n_2}^*(Q - Q'_0) x \psi_1(\mathbf{r}) \varphi_{n_1}(Q - Q_0) d^3\mathbf{r} dQ, \quad (8.9)$$

where we have arbitrarily taken the light to be polarized along the x axis. Note that the equilibrium positions are different because they correspond to different electronic states. The matrix element can be separated into two parts:

$$M_{12} \propto \int \psi_2^*(\mathbf{r}) x \psi_1(\mathbf{r}) d^3\mathbf{r} \times \int \varphi_{n_2}^*(Q - Q'_0) \varphi_{n_1}(Q - Q_0) dQ. \quad (8.10)$$

The first factor is the usual electric-dipole moment for the electronic transition, which we are assuming to be non-zero. The second is the overlap of the initial and final vibrational wave functions. From Fermi's golden rule (eqn B.13), we know that the transition rate is proportional to the square of the matrix element. Hence the intensity of a specific vibronic transition will be proportional to the Franck–Condon factor:

$$I_{n_1, n_2} \propto \left| \int_0^\infty \varphi_{n_2}^*(Q - Q'_0) \varphi_{n_1}(Q - Q_0) dQ \right|^2. \quad (8.11)$$

The Franck–Condon factor basically states that the intensity of the transition is proportional to the overlap of the initial and final vibrational wave functions.

To see how the Franck–Condon factor works in practice, we need to look at the probability amplitudes (i.e. the square of the wave functions) for the vibrational levels involved in the transition. In the absorption transition shown in Fig. 8.7, the molecule starts in the $n_1 = 0$ level of the ground state, which has its equilibrium position at Q_0 , and finishes in the n_2 th level of the excited state, which is centred at Q'_0 . The relevant wave functions are shown in Fig. 8.8. We see that that we have a good overlap for several transitions, with the largest

In an electronic transition, the x that appears in M_{12} is the electron coordinate, because we are specifically considering the interaction with the dipole moment of the electron. The light will, of course, also interact with the dipole moment of the nucleus, but these transitions occur at much lower frequencies in the infrared, and can be ignored in the Franck–Condon approximation.

Harmonic oscillator wave functions peak near the classical turning points at the edge of the potential well as n gets larger. Therefore we can give an approximate rule of thumb that the Franck–Condon factor will be largest for the levels with the edge of their classical potential well close to Q_0 .

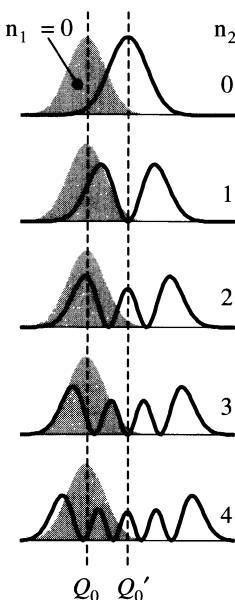


Fig. 8.8 Probability amplitudes for the vibrational levels involved in the absorption transition shown in Fig. 8.7. The initial wave function has been shaded. The molecule starts in the $n = 0$ level of the ground state, and finishes in the n_2 th level of the excited state. Q_0 and Q'_0 are the equilibrium positions for the ground and excited states respectively.

We do not show spectra for the simplest molecules like H_2 here because the first electronic excited state is above the ionization limit of the molecule, and the vibronic lines are not well resolved. Ammonia shows discrete vibronic lines because it is a small, rigid molecule with very well defined vibrational frequencies.

Franck–Condon factor for $n_2 \sim 2$. Hence we would expect the intensity to be largest for the $n_2 = 2$ level, as indicated by the schematic absorption spectrum shown in Fig. 8.7.

The reverse argument can be applied to the emission transitions. This leads to the schematic emission spectrum shown on the right of Fig. 8.7. In the simple model, we expect the emission spectrum to be the ‘mirror’ of the absorption spectrum when reflected about the centre frequency $\hbar\omega_0$. This is called the mirror symmetry rule.

8.3.4 Experimental spectra

The electronic transitions for small molecules usually occur in the ultraviolet spectral region. Figure 8.9 shows the absorption spectrum of ammonia (NH_3). Ammonia is a colourless gas at room temperature, with an absorption edge at 5.7 eV (217 nm). For photon energies above 5.7 eV a series of discrete lines are observed which exhibit the general behaviour shown schematically in Fig. 8.7. The spacing of the lines corresponds to the out-of-plane bending vibrational mode of the molecule with a quantized energy of approximately 0.114 eV. The progression peaks at $n_2 \sim 6$. Further progressions of vibrational–electronic transitions can be discerned in the data starting at 7.3 eV and 8.6 eV. These correspond to electronic transitions to higher singlet excited states of the molecule. It is evident that the data agree very well with the general predictions of the configuration diagram model, although the detailed interpretation can be quite complicated due to the overlapping of the different electronic–vibrational bands.

Figure 8.10 shows the absorption and emission spectra of the laser dye, pyrromethene 567, in benzene solution. Pyrromethene 567 is a large organic molecule with strong electronic transitions in the visible spectral region. The smaller transition energy compared to ammonia is, to a first approximation, just a consequence of the fact that the molecule is larger, and hence that the quantum confinement of the electrons is smaller. The spectra demonstrate reasonably well the mirror symmetry between the absorption and emission, with the emission occurring to the red of the absorption.

No discrete vibronic lines are observed for the laser dye because the large molecule possesses many vibrational modes of differing frequencies, which

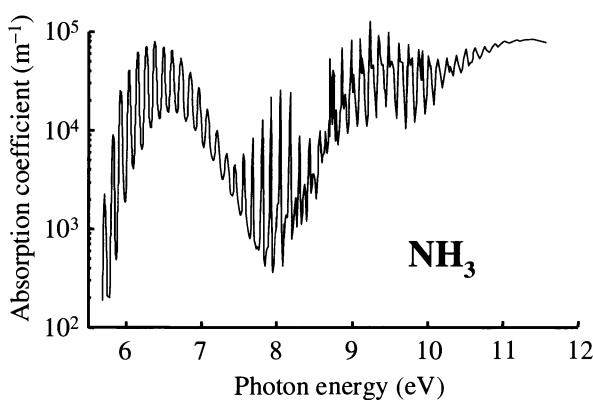


Fig. 8.9 Absorption spectrum of ammonia (NH_3) in the ultraviolet spectral region. After [1], copyright 1954 American Institute of Physics, reprinted with permission.

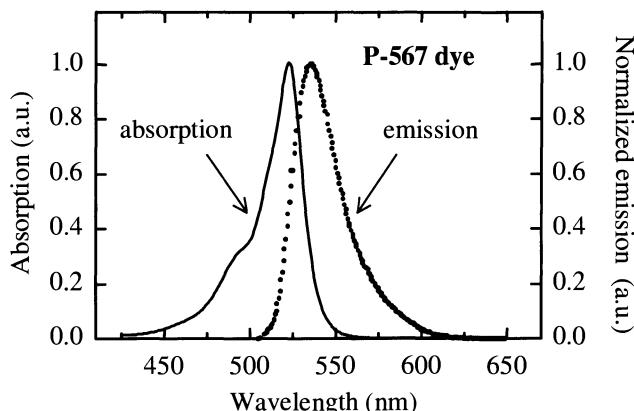


Fig. 8.10 Absorption and emission spectrum of pyromethene 567 in benzene solution. After [2], copyright 2000 Excerpta Medica Inc., reprinted with permission.

generate overlapping progressions of lines that fill out into a continuum. Furthermore, the thermal motion of the molecules and also collisions with the benzene solvent broaden the transitions so that the individual lines cannot be resolved. We therefore obtain continuous absorption and emission bands that follow the envelope of the vibrational–electronic progressions.

Having discussed the spectra of isolated molecules, we can now apply this knowledge to conjugated organic solids. We start with the aromatic hydrocarbons in Section 8.4 below, and then consider conjugated polymers in Section 8.5.

8.4 Aromatic hydrocarbons

Aromatic hydrocarbons are carbon–hydrogen compounds containing benzene rings in their structure. As explained in Section 8.2, their optical properties are determined by the large delocalized π orbitals around the benzene rings. $\pi \rightarrow \pi^*$ transitions are possible from the S_0 singlet ground state to the singlet excited states. In this section we will concentrate on one of the most widely studied aromatic hydrocarbons, namely anthracene ($C_{14}H_{10}$). This has three benzene rings, as shown in Fig. 8.11. The excitonic absorption spectrum of another aromatic hydrocarbon, namely pyrene, was discussed previously in Section 4.5.3 of Chapter 4.

Anthracene forms crystals of high optical quality with a melting point of 217°C . The relatively low melting point of anthracene is an indication of the weak nature of the van der Waals forces that bind the crystal together. Ionic or covalent crystals, by comparison, melt at much higher temperatures: e.g. 801°C for NaCl, 937°C for germanium. The weak nature of the van der Waals forces means that the covalent bonding within the molecule is much stronger than the interactions between the adjacent molecules in the crystal. We therefore expect the electronic states to be strongly localized, and the spectra of the crystals to be fairly similar to those of anthracene in solution.

Figure 8.12 gives a simplified level diagram for the first three electronic states of the anthracene molecule. As explained in Section 8.2, the states are classified by their spin quantum number S . The ground state is a singlet and is labelled S_0 . The first singlet excited state (S_1) occurs at 3.3 eV , which is 1.5 eV

We have already noted that the transition energies are determined, to a large extent, by the size of the molecule. We need several benzene rings to make the molecule large enough that its transitions occur in the visible spectral region. The absorption edge of benzene itself occurs well into the ultraviolet at 4.6 eV (267 nm).

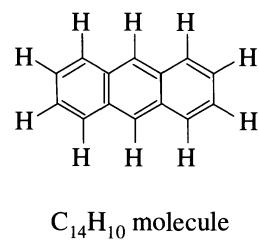


Fig. 8.11 Chemical structure of the anthracene molecule ($C_{14}H_{10}$).

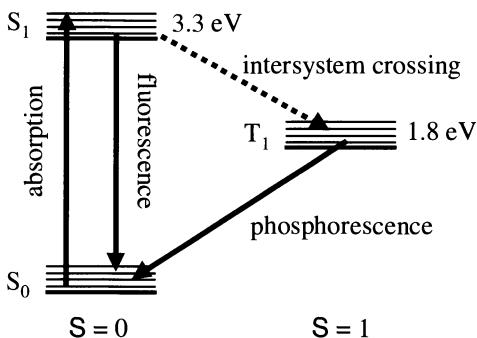


Fig. 8.12 Level diagram for the anthracene molecule. Absorption and fluorescence transitions can occur between the S_0 ground state and the S_1 excited state. Phosphorescent transitions from the T_1 state to the ground state are spin-forbidden and occur on a slow time scale. Electrons in the S_1 state have a small probability of transferring non-radiatively to the T_1 state by intersystem crossing.

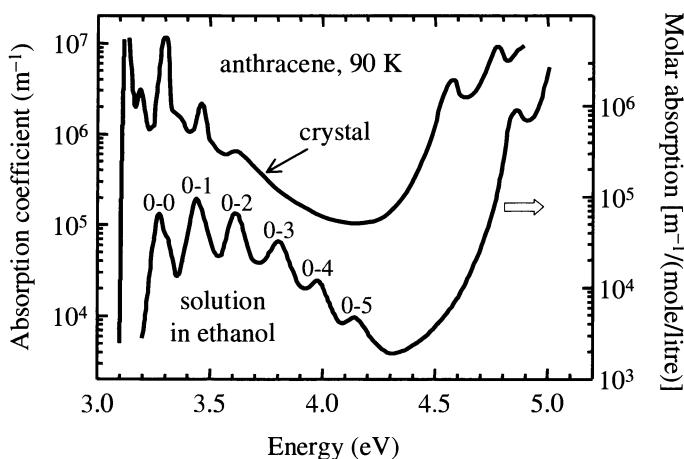


Fig. 8.13 Absorption spectrum of anthracene at 90 K. The absorption spectrum of single crystals is compared to that of a dilute molecular solution in ethanol. The vibronic transitions of the solution are labelled by the vibrational quantum numbers of the ground state and excited state. After [3], reprinted with permission.

above the first triplet excited state (T_1) at 1.8 eV. In Section 8.2 we explained how singlet-triplet transitions have a very low probability due to the spin selection rule. The absorption and emission spectra are therefore dominated by $S_0 \leftrightarrow S_1$ transitions. Phosphorescent $T_1 \rightarrow S_0$ transitions are only observed using special techniques. (See below.)

Figure 8.13 compares the absorption spectrum of a dilute solution of molecular anthracene in ethanol with that of anthracene crystals at 90 K. At higher temperatures, the vibrational structure is less well resolved due to the thermal broadening of the vibronic lines. The absorption edge of the solution occurs at the energy for the $S_0 \rightarrow S_1$ transition at 3.3 eV. About six vibrational-electronic lines with approximately equal spacing are resolved, with the largest intensity for the $n_2 = 1$ line. Similar vibrational structure is observed in the absorption spectrum of the crystal. Three main peaks with approximately the same spacing as in the solution are resolved. These correspond to transitions involving localized vibrations of the anthracene molecules. Additional lines are also observed due to coupling to new vibrational modes present in the crystal. Furthermore, the absorption edge occurs at a slightly lower energy than in the solution.

Note that the absorption strength is very high with values approaching 10^7 m^{-1} . The increase in absorption above 4.2 eV is caused by the onset of $S_0 \rightarrow S_2$ transitions.

The S_1 excited state has a lifetime of 27 ns due to the dipole-allowed $S_1 \rightarrow S_0$ fluorescent transition. The fluorescence spectrum of the crystal consists of a broad vibronic band running from about 3.2 eV (390 nm) to 2.3 eV (530 nm).

Prominent vibronic peaks occur at 3.05 eV, 2.93 eV, 2.76 eV and 2.61 eV. These compare very favourably with the energies expected by applying the mirror symmetry rule to the absorption spectrum shown in Fig. 8.13. (See Exercise 8.11.)

Phosphorescent transitions from the T_1 state at 1.8 eV can be observed by techniques of **delayed fluorescence**. This involves exciting the crystal with a laser pulse, and then observing the emission at long times afterwards. Only the S_1 states are populated by the laser pulse due to the spin selection rule. Most of the molecules return directly to the ground state by radiative transitions, thereby generating prompt fluorescent emission within the first 27 ns after the pulse has arrived. However, there is a small probability for non-radiative $S_1 \rightarrow T_1$ intersystem crossing. The T_1 state has a long lifetime of 24 ms due to the low probability for radiative emission. The weak phosphorescent emission from the T_1 states thus persists as an ‘afterglow’ for 24 ms after the pulse has excited the sample.

The effects described for anthracene are typical of many molecular materials. The absorption and emission spectra are determined by transitions to singlet excited states. The triplet states are only observed in emission experiments as weak phosphorescence. Each electronic transition shows a manifold of vibronic peaks with energy separation determined by the vibrational frequencies of the system.

There has been much development work on light emitting devices based on aromatic materials like anthracene. However, the technological emphasis at present is focussed more on the conjugated polymers described in Section 8.6. This is due to the simpler fabrication techniques required for amorphous thin film polymer devices.

Triplet-singlet transitions can only occur by way of the spin-orbit coupling which causes a weak mixing between the singlet and triplet states. In anthracene, the phosphorescence is always very weak because of the competition with non-radiative recombination mechanisms. (See the discussion in Section 5.1.)

An organic injection laser was demonstrated for the first time in 2000, using tetracene single crystals. The tetracene molecule ($C_{18}H_{12}$) has four benzene rings in a row. The emission wavelength was 576 nm. See Schön *et al.* (2000).

8.5 Conjugated polymers

Polymers are long-chain molecules composed of repeated sequences of monomer units based on carbon–carbon bonds. The name **polymer** is the logical progression of the sequence starting from the **monomer** (single molecule), then to the **dimer** (double molecule), and on to **polymer** (many molecule). This progression is illustrated for ethylene (C_2H_4) in Fig. 8.14. The dimer is $[C_2H_4]_2$, which is cyclobutane (C_4H_8). The polymer is polyethylene, or polythene for short, with a formula of $[CH_2]_n$, where n is a large number.

Polymers can be subdivided into two generic types, namely conjugated or saturated. As explained in Section 8.2, this division is based on whether there are alternating single–double bonds along the polymer backbone or not. The polythene structure shown in Fig. 8.14 is an example of a saturated polymer. In saturated polymers like polythene, all the electrons are incorporated into σ bonds, and are therefore very tightly bound. Their transitions are at high energies in the ultraviolet spectral region, and are not of particular interest here.

Figure 8.15 shows the chemical structure of two conjugated polymers. Figure 8.15(a) shows the simplest conjugated polymer, namely polyacetylene. This polymer is formed by combining many acetylene molecules (C_2H_2 , bonding $HC \equiv CH$) into a long chain with alternating single and double bonds

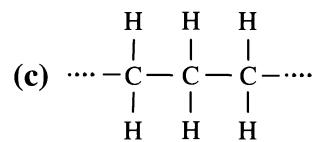
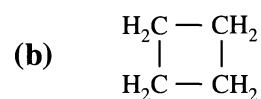
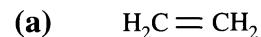


Fig. 8.14 (a) Ethylene (C_2H_4) monomer. (b) Ethylene dimer: $[C_2H_4]_2 \equiv C_4H_8$ (cyclobutane). (c) Poly-ethylene (polythene): $[CH_2]_n$.

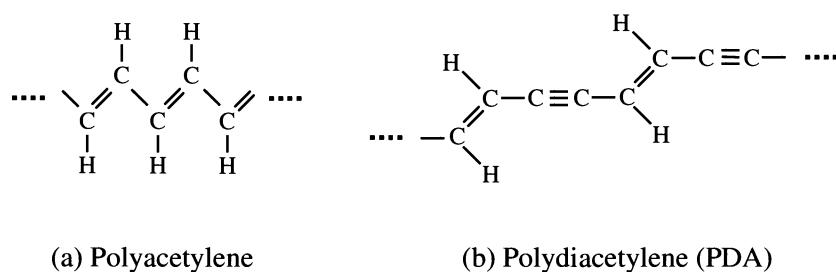


Fig. 8.15 Chemical structure of two conjugated polymers. (a) Polyacetylene. (b) Polydiacetylene (PDA).

The depiction of alternating single and double bonds for the conjugated polymers in Fig. 8.15 is obviously only schematic. In reality, the spare electron of the double bond is shared equally between both bonds in a delocalized π orbital.

between the carbon atoms:



Figure 8.15(b) shows the chemical structure of a slightly more complicated conjugated polymer, namely polydiacetylene (PDA), which incorporates both double and triple bond π electrons. PDA is one of the most widely studied conjugated polymers because it is able to form high quality crystals at room temperature. The $\pi \rightarrow \pi^*$ transitions occur in the visible spectral region for many of these conjugated polymers.

The absorption spectrum of polydiacetylene (PDA) single crystals is shown in Fig. 8.16. The spectrum shows a broad $S_0 \rightarrow S_1$ absorption band starting at 1.8 eV. The band shows clear substructure, with two well-resolved vibronic peaks at 1.9 eV and 2.1 eV. The peaks around 3.6 eV are caused by $S_0 \rightarrow S_2$ transitions.

The optical spectra of conjugated polymers like PDA are strongly affected by excitonic effects. In Chapter 4 we discussed how the Coulomb interaction between the electrons and holes in a semiconductor has a strong effect on the optical spectra through their tendency to bind together. Excitonic effects are present in molecular materials because the unpaired electron left in the ground state by an optical transition can be regarded as a hole state, in the same way that unfilled states in the otherwise full valence band of a semiconductor are considered as holes. This hole acts like a positive charge because it represents the absence of a negative electron.

We have already mentioned that the electronic states of molecular materials are strongly localized. Therefore, the excitons formed are of the tightly bound,

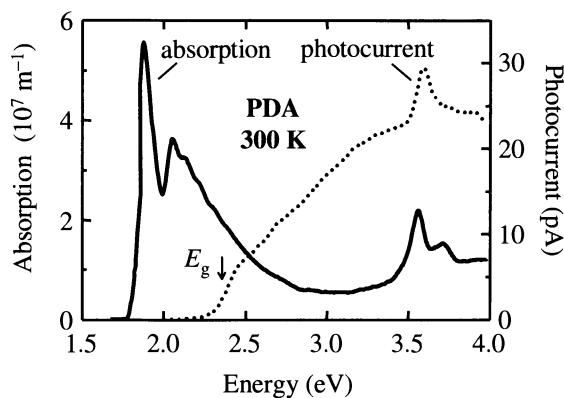


Fig. 8.16 Absorption spectrum of polydiacetylene (PDA) crystals at room temperature. The photocurrent spectrum of the same crystal is shown for comparison. After [4], copyright 1999 Excerpta Medica Inc., reprinted with permission.

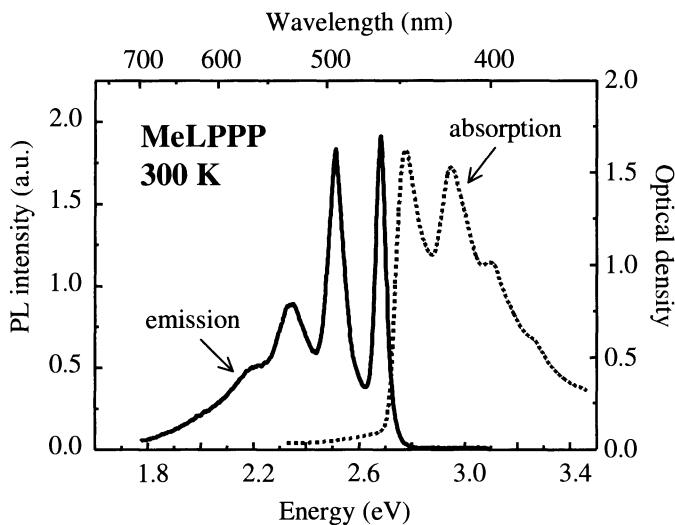


Fig. 8.17 Absorption and emission spectra of thin films of the ladder polymer MeLPPP at room temperature. MeLPPP is a methyl-substituted ladder-type polymer based on poly(paraphenylene). After [5], copyright 1999 American Institute of Physics, reprinted with permission.

or Frenkel, type. (See Section 4.5, especially Section 4.5.3.) The excitonic effects create bound states below the band edge and enhance the probability for radiative transitions.

The binding energy of the excitons in PDA can be deduced by comparing the absorption and photocurrent spectra of the same crystal. The photocurrent spectrum of PDA included in Fig. 8.16 shows a threshold at 2.4 eV, which is about 0.5 eV above the absorption edge. This is a clear indication that the absorption line at 1.9 eV is excitonic in character. The creation of the neutral tightly bound excitons does not affect the conductivity, and the photon energy must exceed the HOMO–LUMO band gap at 2.4 eV before free electrons and holes are available to produce a photocurrent. The measurements therefore indicate that the binding energy of the exciton is 0.5 eV.

Figure 8.17 shows the absorption and emission spectrum of a thin film sample of the conjugated polymer called ‘MeLPPP’ at room temperature. The $S_0 \rightarrow S_1$ transition lies in the green-blue spectral region, and is thus important for optoelectronic applications, as will be described in the next section. The approximate mirror symmetry between the absorption and emission is evident in the data, with four vibrational lines clearly distinguishable. The clarity of the vibrational structure in this material is very striking. This is particularly so if we compare the spectra of MeLPPP with those of the dye molecules in solution shown in Fig. 8.10. The improved resolution of the vibronic structure in the solid is a consequence of the greatly reduced thermal motion of the molecules compared to a solution.

This is a different situation to the photocurrent spectra of free excitons shown in Figs 4.5 and 6.12. The weakly bound free excitons in GaAs can easily dissociate after formation to produce free electrons and holes. The Frenkel excitons in molecular materials, by contrast, do not easily dissociate into free electrons and holes due to their much larger binding energy.

The data in Fig. 8.17 shows that the vibronic structure is more clearly resolved in emission than in absorption. This is because of the disorder in the thin film. The absorption averages over all the states, whereas the emission occurs from excitons that have had time to migrate to the lowest energy states of a particular vibronic level within the sample.

8.6 Organic optoelectronics

It is clear from Figs 8.16 and 8.17 that the electronic transition frequencies in conjugated polymers such as PDA and MeLPPP lie in the visible spectral region (~ 1.7 – 3 eV). The emission band of the polyfluorene-based polymer ‘F8’ mentioned in Section 1.4.4 also occurs in the visible spectral region.

The absorption band of F8 shown in Fig. 1.6 peaks in the ultraviolet at 380 nm, and the emission is Stokes shifted into the blue spectral region.

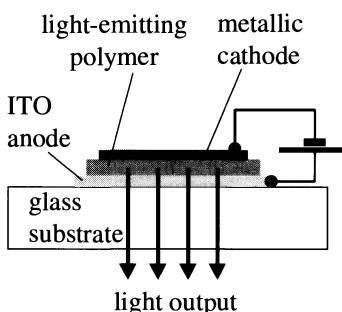


Fig. 8.18 Layer sequence for a typical polymer LED. The hole-injecting anode is usually made from indium-tin oxide (ITO) alloy, while metals such as aluminium, magnesium or calcium are commonly used for the electron-injecting cathodes.

One of the main technological challenges that must be overcome before polymer light diodes become widely-used is the optimization of the injection and transport of the electrons and holes into the active region. This is not as easy as for inorganic semiconductor devices because of the absence of delocalized band states. The electrons and holes must move by hopping between electronic states that are localized on particular polymer molecules rather than by band transport.

These visible-emitting polymers give rise to the possibility of a new technology for light-emitting devices. The subset of conjugated polymers used in this application are generally known as luminescent or light-emitting polymers.

The fundamental operating principle of a polymer light-emitting diode (LED) is much the same as that of its inorganic semiconductor counterpart. Electrons and holes are injected from opposite sides of the diode, and they recombine by emitting light in the active region in the centre of the structure. (cf. Fig. 5.10(a).) Figure 8.18 shows the basic layer sequence used in a typical polymer LED. The light-emitting polymer layer is sandwiched between a cathode for injecting the electrons, and an anode for injecting the holes. The cathode is typically made of aluminium, magnesium or calcium, while the anode is usually made from the alloy indium–tin oxide (ITO), which has the advantage of being both an efficient hole injector and also transparent at visible wavelengths. The whole structure is grown on a glass substrate, and the light generated in the polymer is emitted through the ITO layer and the substrate.

The first successful demonstration of a polymer light-emitting diode was reported in 1990, and used poly-phenylenevinylene (PPV) as the active layer. This gave bright emission in the green-yellow spectral region when a voltage of about 15 V was applied. Since then, much research has been done to develop new polymers that can emit over the whole range of wavelengths in the visible spectral region. Figure 8.19 shows the emission spectra for three copolymers of PPV. It can be seen that the whole visible spectrum can be covered by appropriate choice of the polymer structure, which opens the possibility for making full-colour organic displays. This technology is under active development by several optoelectronics companies.

One of the reasons why organic light-emitting polymers are attracting so much industrial interest is because of the low intrinsic cost of the active materials and also because of the ease of fabrication of the devices. In contrast to inorganic optoelectronic devices, polymer light-emitting diodes do not need to be crystalline. Instead, it is often sufficient to spin-coat thin film layers of amorphous materials onto the substrate. We have already noted that the optical properties of molecular materials are not markedly different from those of the

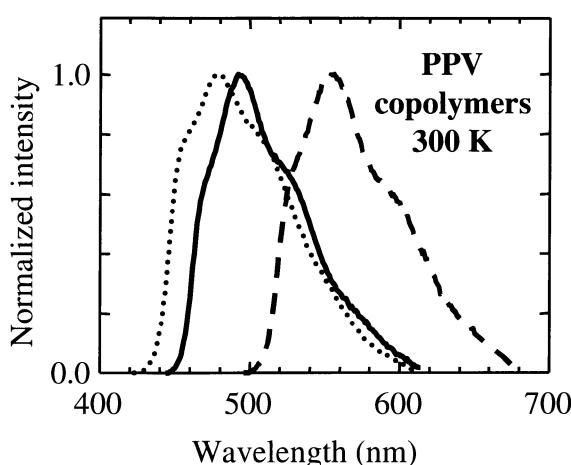


Fig. 8.19 Emission spectra of three poly-phenylenevinylene (PPV) based copolymers at room temperature. The shift in the emission wavelength between the three polymers is achieved by incorporating different substitutional units to the PPV backbone. After [6], copyright 1996 Excerpta Medica Inc., reprinted with permission.

individual molecules. Therefore, the absence of long range order has only a small effect on the main features of the optical spectra. This makes the preparation of the materials much easier and cheaper, and make them an attractive alternative to crystalline inorganic semiconductor LEDs for applications in displays.

Chapter summary

- The optical spectra of molecular solids are determined by localized electronic and vibrational states that are closely related to the states of the isolated molecules.
- Transitions between electronic states are vibronic in character. Vibrational quanta are simultaneously excited during the transition.
- The vibronic spectra can be understood by using configuration diagrams. The Franck–Condon principle says that the configuration coordinates do not change during a vibronic transition, and so the transitions are represented by vertical arrows on configuration diagrams.
- Radiationless relaxation occurs within the vibronic bands after the transition has occurred. The emission spectra are red-shifted with respect to the absorption spectra.
- The excited states of molecules can be divided into singlets and triplets. The ground state is always a singlet. Transitions to singlet excited states are spin-allowed, and dominate the absorption and emission spectra. Singlet–triplet transitions are spin-forbidden and have very low probabilities.
- Emission from singlet states is called fluorescence, while triplet–singlet emission is called phosphorescence.
- Conjugated organic molecules have an alternating sequence of single and double bonds between the carbon atoms. The π electrons from the double bonds form large delocalized orbitals. Larger molecules have smaller transition energies due to the reduced confinement energies of the π electrons.
- The most widely studied examples of conjugated molecules are the aromatic hydrocarbons and conjugated polymers. Many of these have transitions in the visible spectral region. Their absorption spectra show well-resolved vibronic transitions, and the electrons and holes bind together strongly to form Frenkel excitons.
- Conjugated polymers can be used to make electroluminescent devices operating in the visible spectral region.

Further reading

The electronic states of simple molecules are discussed in many introductory quantum mechanics texts, for example, Gasiorowicz (1996). More detailed discussions of molecular spectra can be found in Banwell and McCash (1994) or Haken and Wolf (1995). Klessinger and Michl (1995) give an advanced treatment of the photophysics of organic molecules.

An introductory treatment of molecular crystals is given in Wright (1995), while Pope and Swenborg (1999) give an authoritative treatment of the optical properties of both molecular crystals and polymers.

For review articles on light-emitting polymers, see Bradley (1996), Friend *et al.* (1999), or Mueller (2000). Schön *et al.* (2000) describe an organic injection laser based on a tetracene crystal.

References

- [1] Watanabe, K. (1954). *J. Chem. Phys.*, **22**, 1564.
- [2] Gorman, A.A., Hamblett, I., King, T.A. and Rahn, M.D. (2000). *Journal of Photochemistry and Photobiology A: Chemistry*, **130**, 127.
- [3] Wolf, H.C. (1958). *Z. Natürforsch.*, **A13**, 414.
- [4] Möller, S. and Weiser, G. (1999). *Chemical Physics*, **246**, 483.
- [5] Hertel, D., Bässler, H., Scherf, U. and Hörold, H.H. (1999). *J. Chem. Phys.*, **110**, 9214.
- [6] Bradley, D. (1996). *Current Opinion in Solid State and Materials Science*, **1**, 789.

Exercises

- (8.1) The Schrödinger equation for a one-dimensional harmonic oscillator is given by:

$$-\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + \frac{1}{2}m\Omega^2x^2\Psi = E\Psi.$$

Show that the following wave functions are solutions, stating the value of a and the energies of the three states.

$$\begin{aligned}\Psi_1 &= C_1 e^{-x^2/2a^2}, \\ \Psi_2 &= C_2 x e^{-x^2/2a^2}, \\ \Psi_3 &= C_3 \left(1 - \frac{2x^2}{a^2}\right) e^{-x^2/2a^2}.\end{aligned}$$

- (8.2) Consider the π electrons along a conjugated polymer as a one-dimensional system of length d defined by the total length of the molecule. This allows us to use the infinite potential well model described in Section 6.3.2 to estimate the electron energy. Use this approximation to find the value of d required to give the lower energy transition at 500 nm. Hence estimate the number of repeat units within a polymer that emits at this wavelength, given that the C–C bond length is about 0.1 nm.

- (8.3) The three vibrational modes of the carbon dioxide molecule have frequencies of 2×10^{13} Hz, 4×10^{13} Hz

and 7×10^{13} Hz. Calculate the ratio of the number of molecules with one vibrational quantum excited to those with none for the three modes when the temperature is 300 K.

- (8.4) Calculate the energy difference between two pairs of isolated hydrogen atoms, one of which has both atoms in the $1s$ state, and the other has one atom in the $1s$ state and the other in the $2p$ state. Account qualitatively for the difference between this value and the measured transition energy of 11.3 eV between the ground state and the first electronic excited state of the H_2 molecule.
- (8.5) The potential energy $U(r)$ of two neutral molecules separated by a distance r is sometimes described by the Lennard-Jones potential:

$$U(r) = \frac{A}{r^{12}} - \frac{B}{r^6}, \quad (8.12)$$

where A and B are positive fitting constants.

- (i) Justify the r^{-6} dependence of the attractive part of the potential.
- (ii) Sketch the form of $U(r)$ and show that the energy has a minimum at $r = r_0 = (2A/B)^{1/6}$.

- (iii)* Write down the Taylor expansion of $U(r)$ for small displacements about r_0 and hence show that the form of the potential is parabolic near the minimum. Calculate the angular frequency for harmonic oscillations about this minimum in terms of A , B and the reduced mass μ of the molecule.
- (8.6) The absorption spectrum of benzene is found to consist of a series of lines with wavelengths given by 267 nm, 261 nm, 254 nm, 248 nm, 243 nm, 238 nm and 233 nm. Estimate the energy of the S_1 excited state, and the dominant vibrational frequency of the molecule.
- (8.7) Use the data shown in Fig. 8.9 to draw a schematic configuration diagram for the ground state and the first two singlet excited states of the ammonia molecule.
- (8.8)* Explain why spin-orbit coupling can allow radiative transitions between singlet and triplet states.
- (8.9) Under certain circumstances it is possible to observe weak emission at 760 nm from the pyrromethene dye studied in Fig. 8.10. This emission is found to have a lifetime of 0.3 ms. Suggest a possible explanation for this result.
- (8.10) Use the data shown in Fig. 8.13 to estimate the energy of the dominant vibrational modes of anthracene crystals and molecules in solution.
- (8.11) Apply the mirror symmetry rule to the absorption spectrum of anthracene crystals shown in Fig. 8.13 to deduce the shape of the emission spectrum.
- (8.12) Repeat Exercise 8.11 for the PDA absorption spectrum shown in Fig. 8.16.
- (8.13) The absorption edge of crystalline anthracene ($C_{14}H_{10}$) occurs at 400 nm, but photoconductivity experiments show a different threshold at 295 nm. Account for the difference, and deduce the value of the binding energy of the ground state Frenkel exciton of anthracene.
- (8.14) In Raman scattering, photons are shifted to lower energy by emitting vibrational quanta as they pass through the sample. The angular frequency of the down-shifted photon is equal to $(\omega - \Omega)$, where ω and Ω are the frequencies of the incoming photon and the vibrational mode. By referring to the data given in Fig. 8.17, estimate the wavelength of the Raman-shifted photons generated when a helium neon laser operating at 633 nm is incident on a sample of MeLPPP.
- (8.15)* Equal numbers of electrons and holes are injected into two identical samples of a molecular material. In one case, the electrons and holes are injected optically; in the other, they are injected electrically. Explain why the luminescence from the electrically excited sample is expected to be four times weaker than that from the optically excited one. (Hint: this is related to the formation of triplet excitons.)
- (8.16) A polymer light-emitting diode emits at 550 nm at an operating current of 10 mA.
- (i) Explain why the maximum quantum efficiency we might expect from the device is only 25 %.
 - (ii) Calculate the total optical power emitted on the assumption that the internal quantum efficiency is 25 %.
 - (iii) If the operating voltage is 5 V, what is the power conversion efficiency? Would you expect to obtain this efficiency in a practical device ?

* Exercises marked with an asterisk are more challenging.

9

Luminescence centres

9.1	Vibronic absorption and emission	186
9.2	Colour centres	189
9.3	Paramagnetic impurities in ionic crystals	192
9.4	Solid state lasers and optical amplifiers	196
9.5	Phosphors	199

In this chapter we will be considering the physics of defects and impurities that act as luminescence centres in crystalline host materials. We briefly considered a good example of optically active impurities when we discussed the transmission spectrum of ruby in Section 1.4.5. Rubies consists of chromium ions doped into sapphire crystals. Pure sapphire itself is colourless, and the characteristic red colour arises from the optical transitions of the chromium impurities. Similarly, colourless alkali halide crystals acquire strong colours when certain types of vacancies are present.

The electronic states of the luminescence centres are strongly coupled to the phonons of the host crystal by the electron–phonon interaction. We are thus dealing with a vibronic system, with optical properties analogous to the vibrational–electronic spectra of molecular materials studied in Chapter 8. We therefore begin by reviewing the physics of vibronic transitions, and then focus on two general categories of luminescence centres, namely colour centres and paramagnetic ion impurities. These materials are widely used in solid state lasers and phosphors.

9.1 Vibronic absorption and emission

The electronic states of the impurity atoms doped into a crystal couple strongly to the vibrational modes of the host material through the electron–phonon interaction. This gives rise to continuous vibronic bands that are conceptually different from the electronic bands studied in the band theory of solids. The electronic states are localized near specific lattice sites in the crystal, and the continuous spectral bands arise by coupling the discrete electronic states to a continuous spectrum of vibrational (phonon) modes. This contrasts strongly with interband transitions which involve continuous bands of delocalized electronic states.

The basic processes involved in the vibrational–electronic transitions in molecular materials were described in Sections 8.3.1–8.3.3. The principles developed there form a good starting point for the more general vibronic systems that we will be studying here. There are, however, two additional aspects of the physics that need to be discussed.

- (1) We will usually be considering the optical transitions involving a low density of luminescent dopant ions or defects within an optically inert crystal. The interaction with the crystal host therefore has a strong effect on the spectra.
- (2) We must consider the coupling of the electronic states to a continuous spectrum of vibrational modes, rather than the discrete modes of a

In principle, molecular crystals also have continuous phonon bands. In practice, however, the vibronic transitions in many molecular materials involve localized vibrational modes associated with the internal vibrations of the molecule itself rather than delocalized phonon modes of the whole crystal.

molecule. The density of states for the vibrational modes is determined by the phonon dispersion curves.

The formation of vibronic bands is depicted schematically in Fig. 9.1. Figure 9.1(a) shows the optical transitions between the ground state of an isolated atom (e.g. a dopant ion) at energy E_1 and one of its excited states at energy E_2 . If this atom is inserted into a crystalline host material, the electronic levels can couple to the vibrations of the lattice through the electron–phonon interaction. At this stage, we do not wish to enter into the microscopic details of how such an interaction might occur, but merely consider the possibility that the coupling might be present. The presence of the coupling associates a continuous band of phonon modes with each electronic state, as shown in Fig. 9.1(b).

Optical transitions can occur between the vibronic bands if the selection rules permit them. We first consider an absorption transition. Before the photon is incident, the electron will be at the bottom of the ground state band. The absorption of a photon simultaneously puts the electron in an excited electronic state and creates a phonon, as shown in Fig. 9.1(b). Conservation of energy requires that the angular frequency Ω_2 of the phonon involved must satisfy

$$\hbar\omega_a = (E_2 + \hbar\Omega_2) - E_1 = (E_2 - E_1) + \hbar\Omega_2, \quad (9.1)$$

where $\hbar\omega_a$ is the energy of the photon. Equation 9.1 shows that absorption is possible for a band of energies from $(E_2 - E_1)$ up to the maximum energy of the phonon modes.

After the photon has been absorbed, the electron relaxes non-radiatively to the bottom of the upper band. The system then returns to the ground state band by a vibronic transition of energy:

$$\hbar\omega_e = E_2 - (E_1 + \hbar\Omega_1) = (E_2 - E_1) - \hbar\Omega_1, \quad (9.2)$$

where Ω_1 is the frequency of the phonon created in the ground state band. Once the electron is in the ground state band, it relaxes to the bottom of the band by non-radiative transitions, dissipating the excess vibrational energy as heat in the lattice.

In the relaxation process, the vibrational energy of the localized phonon excited during the absorption transition rapidly spreads throughout the whole crystal and ultimately becomes heat.

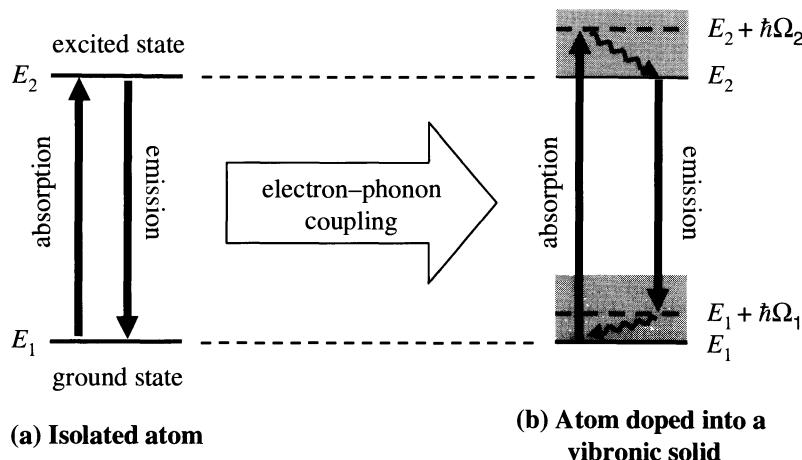


Fig. 9.1 (a) Optical transitions between the ground state and an excited state of an isolated atom. (b) Absorption and emission transitions in a vibronic solid, in which the electron–phonon interaction couples each electronic state to a continuous band of phonons.

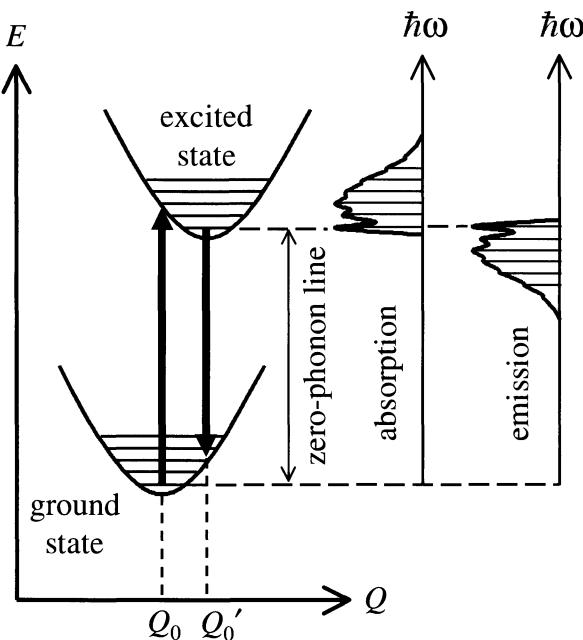


Fig. 9.2 Configuration diagram for the ground state and one of the excited electronic states of a vibronic solid. The optical transitions are indicated by the vertical arrows. The right hand side of the figure shows the general shape of the absorption and emission spectra that would be expected.

On first encountering configuration diagrams, it is quite confusing to understand exactly what the configuration coordinate represents physically. In the case of molecules discussed in Section 8.3.2, it is easy to see that Q corresponds to the amplitude of one of the normal modes of the vibrating molecule. In a vibronic solid, Q might, for example, represent the average separation of the dopant ion from the cage of neighbouring ions in the host lattice. In this case, the vibrations would correspond to a breathing mode in which the environment pulsates radially about the optically active ion. This is equivalent to a localized phonon mode of the whole crystal. In general there will a large number of vibrational modes in a solid, and the configuration coordinate can represent the amplitude of any one of these modes or perhaps a linear combination of several of them.

On comparing eqns 9.1 and 9.2, we see that in a vibronic system the emission generally occurs at a lower energy than the absorption. This red shift is called the **Stokes shift**. It is apparent from Fig. 9.1(b) that the Stokes shift arises from the vibrational relaxation that takes place within the vibronic bands. This contrasts with isolated atoms in which the absorption and emission lines occur at the same frequency.

The Stokes shift between absorption and emission can be understood in more detail by using configuration diagrams. The concept of configuration diagrams was introduced in Section 8.3.2 in the context of the vibrational-electronic spectra of molecules. This model carries over directly to the discussion of the optical transitions in a vibronic solid. The electronic energy of the optically active species is a function of the vibrational configuration of the system as shown schematically in Fig. 9.2. This diagram shows the energy of two electronic states of a vibronic system as a function of Q , the configuration coordinate. We have assumed that the electronic states are bound, and they therefore have a minimum energy for some value of Q . In general, the equilibrium positions for the two states will occur at different values of the configuration coordinate. Therefore we label the position of the minima for the ground state and excited states as Q_0 , and Q'_0 respectively.

The basic physical processes involved in the optical transitions of a vibronic solid are similar to those in a molecule, and we only give a brief summary here. More details can be found in Section 8.3.2. The energy of the electronic ground state can be expanded as a Taylor series about the minimum at Q_0 as follows:

$$E(Q) = E(Q_0) + \frac{dE}{dQ}(Q - Q_0) + \frac{1}{2} \frac{d^2E}{dQ^2}(Q - Q_0)^2 + \dots \quad (9.3)$$

Since we are at a minimum, we know that dE/dQ must be zero. Hence the $E(Q)$ curve will be approximately parabolic for small displacements from Q_0 . The same analysis can be applied to the excited state. This means that to first order we have harmonic oscillator potentials with a series of equally spaced energy levels as sketched in Fig. 9.2.

The Franck–Condon principle discussed in Section 8.3.3 tells us that optical transitions are represented by vertical arrows on the configuration diagram. The absorption transition begins in the lowest vibrational level of the ground state, while the emission commences at the lowest vibrational level of the excited state following non-radiative relaxation. This gives rise to vibronic absorption and emission bands as shown in the right hand side of the figure. In principle, the absorption and emission bands for a particular vibrational mode should consist of a series of discrete lines similar to those observed in molecules, each corresponding to the creation of a specific number of phonons. However, in practice the electronic states can couple to many different phonon modes with a whole range of frequencies, and thus the spectra usually fill out to form continuous bands.

The transitions from the lowest vibrational level of the ground state to the lowest level of the excited state are called the **zero-phonon lines**. Since there are no vibrational quanta involved, the absorption and emission lines occur at the same frequency. In the absorption spectrum there will be a band of vibronic transitions to higher energy of the zero-phonon line, while in the emission spectra there will be a corresponding band to lower energy. The shape of the absorption and emission bands depends on the overlap of the vibrational wave functions as determined by the Franck–Condon factor given in eqn 8.11. In general, the peak occurs away from the zero-phonon line due to the difference between Q_0 and Q'_0 . As with molecules, we would expect mirror symmetry between the emission and absorption about the zero-phonon line.

In the sections that follow, we will apply these general principles to the optical spectra of colour centres and luminescent impurities. In many cases, it will be sufficient to use simpler level diagrams of the type shown in Fig. 9.1(b) to explain the absorption and emission, without delving into the complications of the configuration coordinate model in any detail.

9.2 Colour centres

Colour centres are optically active vacancies in ionic crystals such as the alkali halides. The perfect crystals are colourless insulators with band gaps in the ultraviolet spectral region. (See Table 4.3.) However, it is quite common to find alkali halide crystals that are coloured. Extensive work has demonstrated that this colouration is related to the presence of cation vacancies in the crystal. These defects are aptly named **colour centres** or **F-centres**, where the F stands for *Farbe*, the German word for colour.

Figure 9.3 gives a schematic representation of an F-centre in an alkali halide crystal. The F-centre consists of an electron trapped at a cation vacancy. The cation vacancies are typically created by introducing an excess of the metal ion. This might be done, for example, by heating the crystal in alkali vapour and then cooling it quickly. Alternatively, the vacancies can be produced by

Cations are negative ions.
Anions are positive ions.

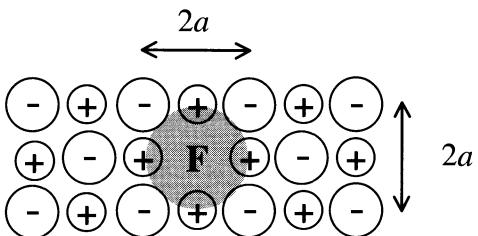


Fig. 9.3 An F-centre in an alkali halide crystal. The centre consists of an electron trapped at a cation vacancy. The shaded region represents the orbit of the electron.

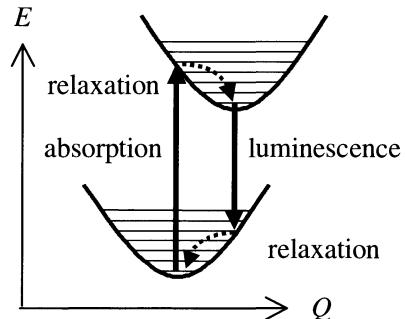


Fig. 9.4 Configuration diagram corresponding to the vibronic transitions of the trapped electron in an F-centre.

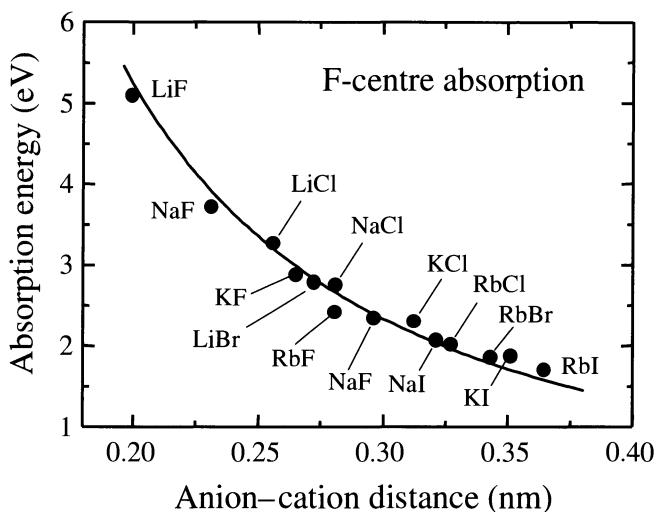


Fig. 9.5 Energy (E) of the peak absorption in the F-band for several face-centred cubic alkali halide crystals. The energies are plotted against the anion–cation distance a . The solid line is a fit with $E \propto 1/a^2$. After [1], reprinted with permission from Plenum Publishers.

irradiation with X-rays or by electrolysis. The absence of the negative ion acts like a positive hole that can attract an electron. The trapped electron is in a bound state with characteristic energy levels.

Optical transitions between the bound states of the trapped electron cause the colouration of the crystals. The trapped electrons couple to the vibrations of the host crystal and this gives rise to vibronic absorption and emission. The processes that take place are illustrated in the generic configuration diagram shown in Fig. 9.4. These transitions are known as F-bands.

Experimental measurements on the F-centres in alkali halides indicate that the frequency of the F-band absorption is proportional to a^{-2} where a is the cation–anion distance in the host crystal. This is clearly evident in the data shown in Fig. 9.5, which plots the energy of the peak in the F-band absorption

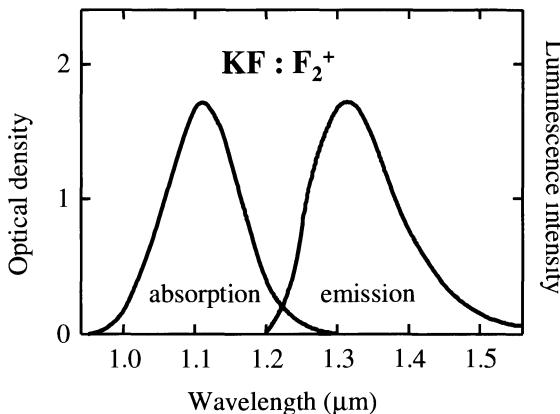


Fig. 9.6 Absorption and emission bands of the F_2^+ centre in KF. After [2], copyright 1985 Excerpta Medica Inc., reprinted with permission.

as a function of a . The solid line is a fit to the data with the energy proportional to $1/a^2$.

This inverse square dependence on a can be explained by a simple model which gives an intuitive understanding of the basic physics. We assume that the trapped electron is confined inside a rigid cubic box of dimension $2a$, as shown in Fig. 9.3. The energy levels of an electron of mass m_0 trapped in such a box are given by:

$$E = \frac{\hbar^2 \pi^2}{2m_0(2a)^2} (n_x^2 + n_y^2 + n_z^2), \quad (9.4)$$

where n_x , n_y and n_z are quantum numbers that specify the bound electronic states. (See Exercise 9.1.) The ground state has $n_x = n_y = n_z = 1$, while the first excited state has one of the quantum numbers equal to 2. The lowest energy transition thus occurs at a photon energy given by:

$$\hbar\nu = \frac{3\hbar^2}{8m_0} \frac{1}{(2a)^2}. \quad (9.5)$$

The model therefore predicts the correct a^{-2} dependence of the F-band absorption energy, but overestimates the transitions energies somewhat. (See Exercise 9.2.)

The simple electron in a box model can also explain the microscopic origin of the coupling between the trapped electrons and the vibrations of the host crystal. A displacement of the neighbouring ions from their equilibrium positions would alter the size of the box in which the electron is trapped. This in turn would alter the electronic energy through eqn 9.4. Such a displacement of the ions could be caused by a vibration of the crystal. Hence the vibrations are coupled to the electronic energy levels, and we have a vibronic system.

Figure 9.6 shows the absorption and emission bands of a slightly more complicated type of colour centre, namely the F_2^+ centre in KF. The emission bands of this F-centre are in the near-infrared spectral region, and the crystals can be used to make tunable lasers, as will be discussed further in Section 9.4. The F_2^+ centre consists of a single electron trapped at two adjacent cation vacancies. Since the centre consists of one electron and two holes, it has a net positive charge of one unit. The Stokes shift and the mirror symmetry between the absorption and emission is clearly evident in the data.

Most of the alkali halide crystals have the face-centred cubic sodium chloride structure, in which the cube edge dimension is equal to $2a$. The exceptions are CsCl, CsBr and CsI, which have simple cubic structures.

This is only a very simple model, and it is not surprising that eqn 9.5 overestimates the energy. A more realistic approach would have to calculate how the electron wave function tries to maximize its overlap with the positive ions while minimizing the overlap with the negative ones.

The transition energies in the F_2^+ centres are lower than those of F-centres because the electron can move over two lattice sites, and hence the box in which the electron is confined is larger. (See Exercise 9.4.)

9.3 Paramagnetic impurities in ionic crystals

Table 9.1 Atomic number Z and electronic configuration of the atoms from the transition metal and rare earth series of the periodic table.

Series	Z	Configuration
Transition metal	21–30	$3d^n 4s^2$
Rare earth	58–70	$4f^n 5s^2 5p^6 6s^2$

In this section we will discuss the optical transitions of paramagnetic metal ions doped into ionic crystals. We will focus on ions from the transition metal and rare earth series of the periodic table. These have optically active unfilled $3d$ or $4f$ shells respectively, as listed in Table 9.1. They are naturally present in certain minerals, but are deliberately doped into synthetic crystals for technological applications. The optical transitions of these doped crystals are the basis for many solid state lasers, and are also widely used in phosphors for fluorescent lighting and cathode ray tubes.

9.3.1 The crystal field effect and vibronic coupling

Metal ions doped as impurities in an ionic crystal substitute at the anion lattice sites. For example, when Cr_2O_3 is doped into an Al_2O_3 crystal to form ruby, the Cr^{3+} ions directly substitute for the Al^{3+} ions. The impurities will normally be present at a low density, so that the interactions between neighbouring dopants are negligible due to their large separation. Hence, the main effect that we need to consider is the perturbation of the electronic levels of the dopant ions due to the crystalline environment in which they are placed.

The optical properties of free ions in the gas phase are characterized by sharp emission and absorption lines with wavelengths determined by their discrete electronic levels. When the same ions are doped into a crystalline host, the optical properties will be modified by the interactions with the crystal. If the interaction is weak, the emission and absorption spectra will remain as discrete lines but perhaps with their frequency slightly shifted and certain degeneracies lifted. On the other hand, if the interaction is strong, the frequencies of the transitions will be quite different from those of the isolated ions, and the spectra may be broadened into a continuum. We will see below that the $4f$ series dopants are generally weakly coupled to the crystal, while the $3d$ series tend to be strongly coupled.

A positive ion doped into a crystal finds itself surrounded by a regular matrix of cations. For example, the Cr^{3+} ions in ruby are surrounded by six O^{2-} ions arranged in an octahedral arrangement, as depicted in Fig. 9.7. These negative ions produce an electric field at the anion site, which perturbs the atomic levels of the ion. This interaction is known as the **crystal field effect**.

The shift of the energy levels of the dopant ion caused by the crystal field can be calculated by perturbation theory. The calculation starts with the gross structure of the free ion with the electrons arranged in the principal atomic shells. It then proceeds by adding on perturbations in order of diminishing size. The details of these perturbation calculations are beyond the scope of this volume, and at this level we are just able to make four qualitative remarks.

(1) The crystal field coupling can be considered to consist of two different contributions. The first arises from the **static** crystal field. This is the perturbation to the energy levels caused by the electric field of the crystal when all of the ions are at their time-averaged equilibrium positions. The second is the **dynamic** effect. This refers to the additional perturbation caused by displacing the neighbouring cations from their equilibrium position, which alters the electric field experienced by the dopant ion and hence alters the perturbation to the energy levels of the atom.

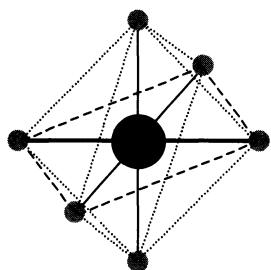


Fig. 9.7 The octahedral crystal environment. The anion dopant is surrounded by six equidistant cations which are located at the corners of an octahedron.

(2) The lifting of the degeneracies of the atomic levels of the free ion due to the *static* field is determined by the symmetry of the crystalline environment. (See, for example, Exercise 9.6.) A useful analogy can be made here with the case of a free atom in a magnetic field. The free atom is spherically symmetric, which implies that the magnetic levels are degenerate. The application of an external field defines a preferred axis, and the levels split by the Zeeman effect. The same is true for the ions doped in the crystal. The magnetic levels of the free ion are degenerate, but are split in the crystal because the host defines axes so that not all directions are equivalent. This point is illustrated in Fig. 1.8.

(3) The *dynamic* crystal field effect is the origin of the vibronic coupling in these systems. Vibrations of the crystal cause the ions to be displaced from their equilibrium positions and therefore alter the electric field experienced by the dopant ion. This in turn alters the perturbation of the electronic levels, and thus couples the vibrations to the electronic levels of the system. This is equivalent to an electron–phonon interaction. One way to look at this is to consider the phonon as acting like an amplitude modulation on the crystal field. This induces side bands at the phonon frequency on the electronic levels through the crystal field effect. In some cases it is possible to resolve distinct side bands in the optical spectra that correspond to specific phonon frequencies, but more often than not, the side bands form a continuum due to the continuous distribution of frequencies of the phonon modes.

(4) The magnitude of the crystal field effects for the transition metal and rare earth ions are very different. This is a consequence of the electronic configurations of the optically active electrons. (See Table 9.1.) Transition metal ions are formed when the outermost $4s$ electrons of the neutral atoms are removed. The $3d$ orbitals therefore lie on the outside of the ion and have a large radius. Rare earth ions, by contrast, are formed when the outermost $6s$ electrons are removed, leaving the optically active $4f$ orbitals inside the filled $5s$ and $5p$ shells. This means that they have a smaller radius (see Exercise 9.5) and are also partly shielded from external fields. These two factors mean that the transition metal ions are much more sensitive to the crystal field than the rare earths.

These points apply to a wide range of paramagnetic ions in crystalline hosts. In the following subsections, we will discuss the properties of the $3d$ and $4f$ series ions separately, starting with the rare earths.

9.3.2 Rare earth ions

The magnitude of the crystal field effect in rare earth ions is relatively small, due to the screening of the optically active levels (see point 4 above). Furthermore, the spin–orbit coupling is quite large because it varies as Z^4 , and Z is in the range 58–70. This means that the crystal field effects are smaller than the spin–orbit coupling. Therefore, in treating the crystal field effects by perturbation theory, we must apply the spin–orbit interaction first.

The spin–orbit interaction splits the gross structure of the free ions into fine structure terms defined by the quantum numbers $|LSJ\rangle$ (Russell–Saunders coupling), denoted in spectroscopic notation as $^{2S+1}L_J$. The crystal field then perturbs these states, shifting their energies slightly and causing new splittings. However, the size of these shifts is much smaller than the spin–orbit splittings,

Group theory provides an extremely powerful tool for working out the lifting of degeneracies by the crystal field. For example, it tells us that the five-fold degenerate $3d$ orbitals of a free transition metal ion are split into a doublet and triplet by an octahedral crystal field. However, group theory cannot tell us whether it is the doublet or triplet that is at the lower energy, and nor can it tell us the size of the splitting. This requires detailed numerical modelling of the interaction between the dopant anion and the electric field of the neighbouring cations.

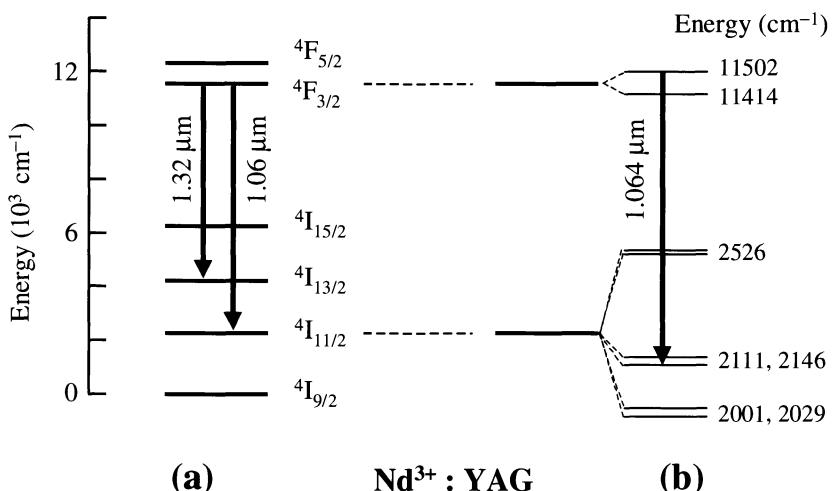


Fig. 9.8 (a) Energy level diagram for Nd³⁺ ions in a YAG crystal. (b) Crystal field fine structure of the 1.06 μm transition. The energies are given in wave number units. ($1 \text{ cm}^{-1} \equiv 1.240 \times 10^{-4} \text{ eV}$.)

The $4F_{3/2} \rightarrow 4I_{13/2}$ and $4F_{3/2} \rightarrow 4I_{11/2}$ laser transitions in Nd³⁺ have $\Delta J = 5$ and 4 respectively. They are therefore electric-dipole forbidden for the free ion. In the crystal, the crystal field perturbation mixes states of different J and provides a finite probability that the transitions can occur. The upper state still has a long lifetime of 230 μs at 300 K, which allows it to store energy efficiently. This explains why Nd³⁺ lasers give such high pulse energies.

and so the optical spectra of the dopant ions are generally fairly similar to those of the free ions.

As an example of these effects, we can consider the optical spectra of Nd³⁺ ions doped into an yttrium aluminium garnet (Y₃Al₅O₁₂ or ‘YAG’) crystal. We choose this example because Nd:YAG crystals form the gain medium in one of the most important solid state lasers. The electronic configuration of Nd³⁺ is $4f^3$. Hund’s rules tell us that the ground state has $S = 3/2$, $L = 6$ and $J = 9/2$, that is, a $4I_{9/2}$ term. Above this ground state there is a progression of excited states. Figure 9.8(a) shows the first five excited states without the crystal field fine structure. Two important transitions are identified, namely the $4F_{3/2} \rightarrow 4I_{13/2}$ line at 1.32 μm and the $4F_{3/2} \rightarrow 4I_{11/2}$ line at 1.06 μm . Lasing has been demonstrated for both transitions, although the 1.06 μm line is the more important.

Figure 9.8(b) shows the crystal field fine structure for the $4F_{3/2} \rightarrow 4I_{11/2}$ transition at 1.06 μm . The octahedral symmetry of the YAG crystal field lifts the degeneracy of the m_J states of the free ion, with states of the same $|m_J|$ having the same energy. Thus the upper $4F_{3/2}$ term, which has four degenerate m_J states in the free ion corresponding to $m_J = -3/2, -1/2, +1/2$ and $+3/2$, is split by the crystal field into two levels identified by $|m_J| = 3/2$ and $|m_J| = 1/2$. Similarly, the lower $4I_{11/2}$ term splits into six sublevels. The size of the crystal field splittings is of order 100 cm^{-1} , which is approximately an order of magnitude smaller than the spin-orbit splitting.

Figure 9.9 shows experimental data for the emission spectrum of the $4F_{3/2} \rightarrow 4I_{11/2}$ transition at 77 K and 300 K. The spectrum consists of sharp lines rather than a continuum, which demonstrates the weak nature of the crystal field interaction. Transitions involving most of the sublevels of the upper and lower terms are clearly identifiable in the spectra. The laser transition at 1.064 μm is identified in the emission spectra, and the states involved are indicated in Fig. 9.8(b).

The emission lines in Fig. 9.9 are broader at 300 K than at 77 K. This is a consequence of the stronger electron–phonon coupling at the higher temperature. The linewidth of the 1.064 μm emission line is 120 GHz at 300 K. As

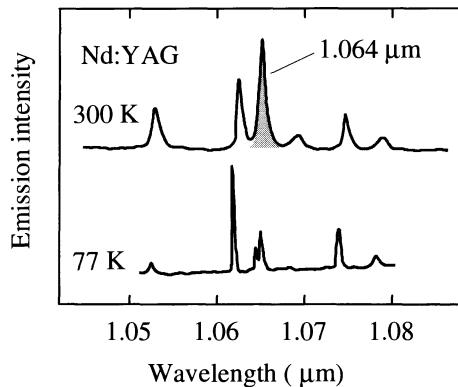


Fig. 9.9 Emission spectrum for the ${}^4F_{3/2} \rightarrow {}^4I_{11/2}$ transition in a Nd:YAG crystal at 77 K and 300 K. The laser transition at 1.064 μm has been highlighted by shading. After [3], copyright © 1966 by Academic Press, reproduced by permission of the publisher.

we will see in Section 9.4, this broadening is very beneficial for making short pulse lasers.

9.3.3 Transition metal ions

Transition metal ion dopants have a strong interaction with the crystal field. As mentioned in point 4 of Section 9.3.1, this is due to the relatively large radius of the $3d$ orbitals and the fact that they are unshielded by outer filled shells. This makes their electronic states very sensitive to the crystalline environment. A striking example of this is that Cr^{3+} ions are responsible for both the red colour of ruby and the green colour of emerald. The change in colour arises from the shift in the energy levels in changing the host crystal from sapphire (Al_2O_3) in ruby to beryl ($\text{Be}_3\text{Al}_2\text{Si}_6\text{O}_{18}$) in emerald. This can be contrasted with the behaviour of rare earth dopants when the crystal host is changed. For example, the 1.064 μm transition of Nd:YAG only shifts to 1.053 μm when the host crystal is changed to YLF (YLiF_4).

On comparing the $3d$ group dopants with the rare earths, the crystal field effect is stronger and the spin-orbit interaction is smaller. The latter point is a consequence of the Z^4 dependence of the spin-orbit interaction. This means that in treating the crystal field effect by perturbation theory, we should consider the crystal field interactions first, and then apply the spin-orbit coupling afterwards. This means that the character of the states will be very different from those of the free ion.

As an example we can consider a transition metal ion dopant in the octahedral crystalline environment shown in Fig. 9.7. We take the simplest case in which the metal has only a single $3d$ electron, namely Ti^{3+} , with an electronic configuration of $3d^1$. The metal anion lies at the origin with six cation nearest neighbours at $(\pm a, 0, 0)$, $(0, \pm a, 0)$ and $(0, 0, \pm a)$. The octahedral crystal field interacts with the degenerate $3d$ levels of the free atom and splits them into a doublet and a triplet, as shown in Fig. 9.10.

The nomenclature used for the crystal-field-split levels in Fig. 9.10 is taken from group theory. The doublet is labelled as an E state, and the triplet as a T_2 state. These states are sometimes further specified by their spin multiplicity and their parity. Thus the doublet is a ${}^2\text{E}_g$ state, with ${}^2\text{T}_{2g}$ for the triplet. The superscript prefix of 2 tells us that there are two spin states for each electron

The transmission spectrum of ruby is given in Fig. 1.7. The red colour is caused by the two strong absorption bands in the green/yellow and blue spectral regions respectively.

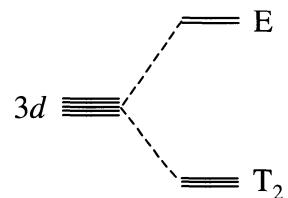


Fig. 9.10 Splitting of the degenerate $3d$ levels of a transition metal ion in an octahedral crystal environment.

The splitting of the $3d$ levels into a doublet and triplet can be deduced by group theory, and can also be worked out by explicitly calculating the perturbation due to the crystal field.

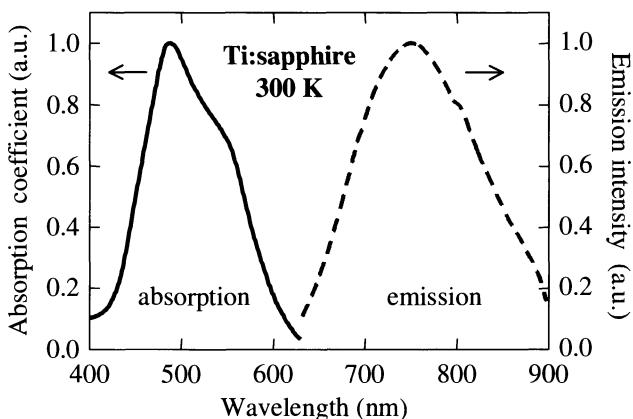


Fig. 9.11 Absorption and emission spectra for Ti^{3+} ions doped into sapphire (Al_2O_3) at 300 K. After [4], copyright 1983 Pennwell Corp., reprinted with permission.

state, while the subscript g refers to the parity and is short for *gerade*, the German word for ‘even’.

Figure 9.11 shows the absorption and emission spectrum of Ti^{3+} ions doped into the octahedral sapphire (Al_2O_3) host at 300 K. The spectra correspond to transitions between the T_{2g} ground state level and the E_g excited state. Since the upper and lower levels both have even parity, conventional electric dipole transitions are forbidden. The transitions become possible because the introduction of the Ti^{3+} impurities slightly distorts the octahedral environment of the host, and mixes in states of odd parity. This gives an upper state radiative lifetime of $3.9 \mu\text{s}$, which is shorter than the non-radiative lifetime at 300 K. The luminescent efficiency at 300 K is therefore high (see eqn 5.5 and Exercise 5.4), which explains why Ti:sapphire makes good laser crystals.

It is evident from the experimental data that the absorption and emission spectra consist of continuous bands rather than sharp lines. This is a consequence of the strong vibronic broadening of the ground state and the excited states. The Stokes shift of the emission is also apparent in the data, together with the approximate mirror symmetry of the emission and absorption about the zero phonon wavelength of around 630 nm.

The general shape of the spectra shown in Fig. 9.11 is typical of many other transition-metal-ion-doped crystals. The details of the level schemes become more complicated if there is more than one $3d$ electron present and/or the crystal environment has lower than octahedral symmetry. However, the basic point is that the crystal field splits the atomic levels derived from the $3d$ states, and then the strong coupling to phonons broadens these states into continuous vibronic bands. This gives rise to continuous vibronic absorption and emission bands, which are particularly useful for making tunable lasers, as we will see in the next section.

9.4 Solid state lasers and optical amplifiers

A brief explanation of how population inversion produces optical gain is given in Section 5.4.3. The reader is referred to the bibliography listed under Further Reading for a more detailed treatment of the physics of solid state lasers.

Many important solid state lasers use transition metal ion or rare earth ions as the gain medium. For example, the first laser ever demonstrated used ruby (Cr^{3+} doped into Al_2O_3) as the active material. The lasers can be generally classified as having either a fixed or tunable wavelength. The emission spectra

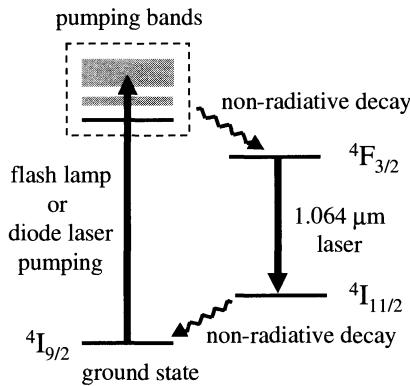


Fig. 9.12 Population inversion scheme for the $1.064\text{ }\mu\text{m}$ transition in a Nd:YAG laser.

of rare earth ions usually consist of very specific wavelengths, and generally fall into the first category. Transition metal ions, on the other hand, usually show broad emission bands, which give rise to the possibility for tunable laser operation over a very wide range of wavelengths.

The optical gain in a laser crystal is produced by **population inversion** between the upper and lower laser levels. The population inversion is achieved by ‘pumping’ atoms into the upper laser level by a variety of mechanisms. Figure 9.12 indicates how this is done for the $1.064\text{ }\mu\text{m}$ line of the Nd:YAG laser. The upper laser level is the ${}^4\text{F}_{3/2}$ state. This level is populated by first pumping electrons from the ground state to excited states such as the ${}^4\text{F}_{5/2}$ level identified in Fig. 9.8(a). Alternatively, the upper laser level can be populated by pumping to other excited states not shown in Fig. 9.8(a). Some of these are broadened into bands by vibronic coupling, and can thus absorb a wide range of frequencies, which makes it easier to pump them. The electrons in the higher excited states then relax to the upper laser level by rapid non-radiative decay. This gives rise to population inversion with respect to the ${}^4\text{I}_{11/2}$ state, and if a suitable cavity is provided, lasing can occur. Rapid non-radiative decay to the ${}^4\text{I}_{9/2}$ state ensures that the electrons do not accumulate in the lower laser level and reduce the population inversion.

Nd:YAG lasers have been traditionally pumped by bright flash lamps. However, the transition from the ground state to the ${}^4\text{F}_{5/2}$ state conveniently matches the optimum emission wavelength of GaAs quantum well diode lasers around 800 nm (see Section 6.6). This has given rise to a new generation of Nd:YAG lasers pumped by semiconductor lasers, which are much more efficient and stable than their counterparts using flash lamps.

The population inversion mechanism in Ti:sapphire lasers follows the general procedure shown in Fig. 9.13. Electrons are pumped from the ground state of the ${}^2\text{T}_2$ band to an excited level within the ${}^2\text{E}$ band. These electrons relax to the bottom of the ${}^2\text{E}$ band by phonon emission, and this creates population inversion with respect to the vibronic levels of the ${}^2\text{T}_2$ band. Laser emission can then occur over a broad range of wavelengths within the emission band shown in Fig. 9.11. Phonon emission finally depletes the lower laser level, allowing a large population inversion to be achieved.

Ti:sapphire lasers are frequently pumped by argon ion lasers, whose emission lines at 488 nm and 514 nm match very well to the absorption bands of the

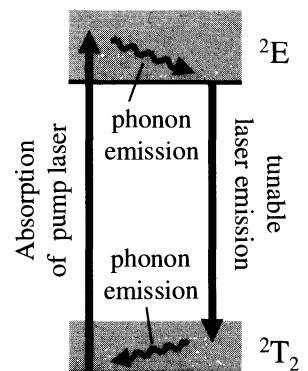


Fig. 9.13 Level diagram for the tunable vibronic emission in a Ti:sapphire laser.

Table 9.2 Common solid state lasers based on rare earth or transition metal ions. All results are at room temperature except for the cobalt laser which operates at 77 K. YAG stands for yttrium aluminium garnet ($\text{Y}_3\text{Al}_5\text{O}_{12}$).

Laser	Ion	Configuration	Host	Wavelength (nm)
Titanium	Ti^{3+}	$3d^1$	Al_2O_3	700–1100
Ruby	Cr^{3+}	$3d^3$	Al_2O_3	694
Alexandrite	Cr^{3+}	$3d^3$	BeAl_2O_4	700–820
Cobalt	Co^{2+}	$3d^7$	MgF_2	1500–2500
Neodymium	Nd^{3+}	$4f^3$	YAG	1064
Neodymium	Nd^{3+}	$4f^3$	glass	1054
Erbium	Er^{3+}	$4f^{11}$	optical fibre	1530–1560

It might seem strange to use one laser to pump another, but it actually makes sense because it is an efficient way to convert the discrete frequencies of a fixed wavelength high power laser to continuously tunable radiation.

Ti:sapphire crystal. (See Fig. 9.11.) Another possibility is to use a frequency doubled Nd:YAG laser operating at 532 nm. The 532 nm radiation is obtained by doubling the frequency of the 1064 nm laser line using the techniques of nonlinear optics discussed in Chapter 11.

Table 9.2 lists a number of important solid state lasers based on transition metal or rare earth ions. As is apparent from the table, it is possible to cover a wide range of frequencies in the visible and near-infrared spectral regions using these sources. Of the lasers listed, the Nd^{3+} lasers have found the most widespread applications in industrial and medical environments, due to their high power output and rugged structure.

The last gain medium listed in Table 9.2, namely the erbium-doped optical fibre, has become increasingly important for use in telecommunications systems. The level scheme for the Er^{3+} ion is shown in Fig. 9.14(a). The ${}^4\text{I}_{11/2}$ band of the Er^{3+} ions is 1.27 eV above the ground state, which makes it suitable for pumping with 980 nm diode lasers. Rapid non-radiative relaxation occurs to the bottom of the ${}^4\text{I}_{13/2}$ band, where the electrons accumulate due to the long lifetime of the state (11 ms). This creates population inversion with respect to the vibronic band of the ${}^4\text{I}_{15/2}$ ground state, in a similar fashion to the Ti:sapphire laser shown in Fig. 9.13. We thus obtain optical gain for the ${}^4\text{I}_{13/2} \rightarrow {}^4\text{I}_{15/2}$ vibronic band between 1.53 μm and 1.56 μm .

The erbium ions are doped into a section of fibre, and the output of the 980 nm pump laser is introduced using a fibre coupler, as shown schematically in Fig. 9.14(b). Lasing can occur if mirrors are placed around the gain medium, but usually there is no cavity and the gain of the erbium ions is used to amplify signals. The gain peaks around 1.55 μm , which is one of the preferred wavelengths for silica fibre systems. Amplification factors of about 10^3 can be obtained with a few metres of erbium fibre.

We mentioned in Section 9.2 that colour centres can also be used as laser crystals. The pumping mechanism follows the same general scheme as for the Ti:sapphire laser shown in Fig. 9.13. The electrons are first excited to the upper band by absorption of photons from the pump laser. They then relax to the bottom of the band by emitting phonons, before emitting the laser light and finally returning to the bottom of the ground state band by a second phonon emission process. The colour centre systems are mainly used for spectroscopic studies in the infrared spectral region around the telecommunications wavelengths of 1.3 μm and 1.55 μm . For example, laser operation has been demonstrated for

The fibre losses are very small at 1.55 μm . Nevertheless, in long distance systems (e.g. transatlantic) the signals in the fibre must still be amplified at regular intervals to compensate for the small (but non-zero) losses.

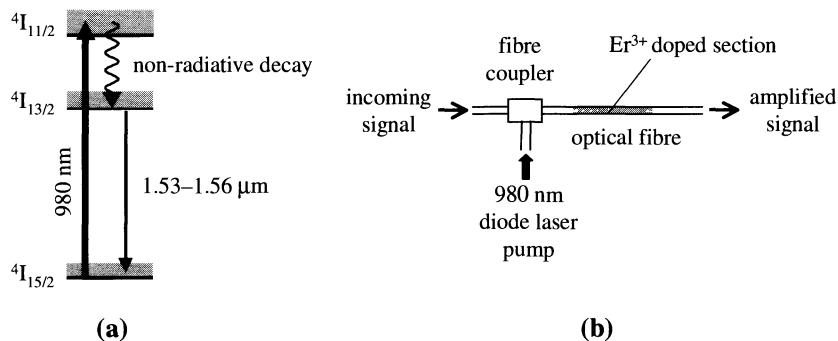


Fig. 9.14 The erbium-doped fibre amplifier. (a) Level scheme. The ${}^4I_{11/2}$ band is 1.27 eV above the ground state, and is suitable for pumping with 980 nm diode lasers. Rapid non-radiative relaxation occurs to the bottom of the ${}^4I_{13/2}$ band. This creates population inversion for the ${}^4I_{13/2} \rightarrow {}^4I_{15/2}$ vibronic transition, and hence gain between 1.53 μm and 1.56 μm . (b) Schematic diagram of the fibre amplifier. The 980 nm pump laser is coupled into the erbium-doped section by means of a fibre coupler.

the F_2^+ centre in KF between 1.22 μm and 1.50 μm , which covers most of the emission bands of this crystal (see Fig. 9.6). The laser is conveniently pumped by the 1.064 μm line of a Nd:YAG laser, which matches well to the absorption band between 1.0 μm and 1.2 μm . With other combinations of host crystals and types of colour centre, it is possible to cover a wide range of wavelengths in the infrared spectral region between 1 μm and 2 μm .

In many modern laser applications it is desirable to be able to produce very short light pulses. The duration Δt of the shortest pulses that can be produced by a laser is set by the spectral width $\Delta\nu$ of the emission line according to:

$$\Delta\nu\Delta t \sim 1. \quad (9.6)$$

This **time-bandwidth product** is a type of uncertainty principle. It means that laser crystals with broad emission lines are good candidates for producing very short pulses. The precise value of the time-bandwidth product depends on the shape of the pulse. For example, if the pulses are Gaussian, $\Delta\nu\Delta t = 0.441$. (See Exercise 9.10.)

We mentioned in connection with Fig. 9.9 that the linewidth of the 1064 nm emission line in Nd:YAG is about 120 GHz at 300 K. This allows for the possibility of generating pulses as short as a few picoseconds, which has indeed been demonstrated. However, the present world record is set by a mode-locked Ti:sapphire laser. The extremely broad spectral width of the emission band ($\Delta\nu \sim 10^{14}$ Hz) makes it possible to generate pulses shorter than 10 fs with this laser.

The ultra-short pulse Ti:sapphire laser mentioned above represents a *tour de force* of present-day solid state optical technology discussed throughout this book. The pump source – a 532 nm diode-pumped Nd:YAG laser – combines quantum well technology (Chapter 6) with nonlinear optics (Chapter 11) and solid state laser technology (the present chapter). The ultra-short pulse Ti:sapphire laser pushes the technology to its theoretical limit, and provides an extremely useful tool for studying dynamical effects in many interesting solid state systems.

9.5 Phosphors

The term **phosphor** covers a wide range of solids that emit visible light when excited either by a beam of electrons or by ultraviolet light. In this section we

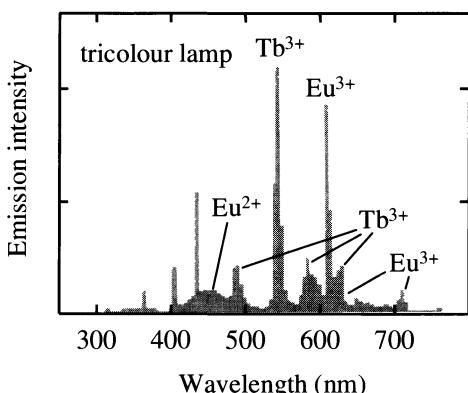


Fig. 9.15 Emission spectrum of a tricolour fluorescent lamp with a colour balance equivalent to a black body source at 4000 K. The main emission lines from the blend of Eu²⁺, Eu³⁺ and Tb³⁺ phosphors in the lamp are identified. The sharp lines at 405 nm and 436 nm originate from the mercury discharge. There is also a mercury line at 545 nm which is very close to the main Tb³⁺ emission line. After [5], reprinted with permission from Plenum Publishers.

The description of electric lights as ‘fluorescent’ is somewhat misleading. The light emission process should more accurately be described as phosphorescence rather than fluorescence. This is because we are dealing with dipole-forbidden transitions with long lifetimes.

will briefly discuss some phosphors based on rare earth ions. These phosphors find widespread application in cathode ray tubes (oscilloscope screens, televisions and computer monitors) and also in fluorescent lighting. In the former case, an electron beam strikes a screen coated with a suitable phosphor which then releases the energy in the form of light. In the latter case, the phosphors coated on the inside of the lamp tube are excited by the ultraviolet photons from a mercury discharge and re-emit in the visible.

Fluorescent lighting was developed around the time of the second world war and quickly established itself for general lighting applications due to its greater efficiency compared to incandescent lamps. The tubes contain mercury vapour at a low pressure, and the inside of the glass is coated with the phosphor. An electrical discharge excites the mercury atoms, which emit ultraviolet radiation at 254 nm and 185 nm. This ultraviolet light is then absorbed by the phosphor, and re-emitted in the visible spectral region.

For many years fluorescent lighting was dominated by halophosphate phosphors incorporating Sb³⁺ and Mn²⁺ dopants. However, the advent of rare-earth based phosphors in 1975 revolutionized the market. As discussed in Section 9.3.2, rare earth ion dopants emit with characteristic colours determined mainly by the spin-orbit splittings of the atoms. By using a blend of three rare earth dopants, one emitting in the blue, one in the green, and one in the red, it is possible to make highly efficient lights with a very good white light colour balance.

Figure 9.15 shows the emission spectrum of a tricolour lamp blended to give a colour balance equivalent to a black body source at 4000 K. The lamp incorporates a carefully selected mixture of BaMgAl₁₀O₁₇:Eu²⁺, CeMgAl₁₁O₁₉:Tb³⁺, and Y₂O₃:Eu³⁺. The Eu²⁺ ($4f^7$) ions emit in the blue at 450 nm, the Tb³⁺ ($4f^8$) ions in the green at 550 nm, and the Eu³⁺ ($4f^6$) ions in the red at 610 nm. These emission lines are clearly visible in the spectrum of the tricolour lamp, together with other weaker emission lines from the phosphors and the mercury lines at 405 nm, 436 nm and 545 nm. These tricolour lamps are much more efficient than the older halophosphates, and also offer a much better colour balance. Similar phosphors are used for the red, green and blue pixels in colour televisions and computer monitors.

A recent advance in phosphor science has been the development of white light semiconductor LEDs (light emitting diodes). This approach is much sim-

pler than trying to blend separate blue, green and red LEDs to make white light. The white light diodes incorporate GaN-based materials. As discussed in Section 5.4, GaN itself emits in the UV, while GaInN alloys can be grown to emit in the blue. This suggests two possibilities for making white LEDs. The first is to use the UV emission of GaN to excite a phosphor blend as in the tricolour lamp. The second is to use some of the blue emission of a GaInN alloy directly, and to use the rest to excite a blend of green and red phosphors. These options offer the possibility of developing highly efficient low voltage white light emitters that could completely revolutionize the lighting market.

Chapter summary

- Luminescence centres are optically active defects and impurities within crystalline hosts. The electronic states of the centres are localized at the defect or impurity from which they arise.
- The electronic states couple to the lattice vibrations of the host crystal through the electron–phonon interaction. Optical transitions between the states are vibronic and involve the simultaneous excitation of phonons.
- The vibronic coupling leads to broad absorption and emission bands in many materials. The emission occurs at a lower energy than the absorption. This red-shift of the emission bands is called the Stokes shift.
- Colour centres (F-centres) consist of an electron trapped at a cation vacancy in an ionic crystal. Vibronic transitions between the bound states of the electron give rise to broad absorption and emission bands. The energy of the absorption peak is inversely proportional to the square of the lattice constant of the host crystal.
- The energy levels of paramagnetic ions doped into ionic crystals are perturbed by the crystal field of their local environment. In the rare earth ions, the crystal field effects are quite small, but in the transition metal ions, the crystal field effects are very large.
- The optical spectra of rare earth ions tend to consist of discrete lines. The crystal field effect causes small splittings of transitions that are degenerate in the free ions.
- The optical spectra of transition metal ions consist of broad vibronic bands. The emission is Stokes-shifted with respect to the absorption.
- Paramagnetic ions and colour centres can be used as the gain medium in solid state lasers. Rare earth ion lasers tend to operate on discrete wavelengths, while the transition metal ions and colour centres give rise to tunable laser wavelengths. Erbium ions doped into optical fibres can be used as optical amplifiers at $1.55\text{ }\mu\text{m}$.
- Rare earth ions are frequently used as the light-emitting material in phosphors for fluorescent lighting and cathode ray tubes.

Further reading

The basic physics of colour centres is covered in Ashcroft and Mermin (1976), Burns (1985) or Kittel (1996). A good introduction to luminescent impurities is given by Elliott and Gibson (1974). The crystal field effect is explained in more detail by Blundell (2001).

Henderson and Imbusch (1989) give an authoritative treatment of vibronic systems. Detailed information about solid state lasers is given in Svelto (1998) or Silfvast (1996). A collection of review papers on colour centres, transition metal ions and phosphors may be found in Di Bartolo (1992).

References

- [1] Baldacchini, G. (1992). In *Optical properties of excited states in solids* (ed. B. di Bartolo), NATO ASI Series B, Vol. 301. Plenum Press, New York, p. 255.
- [2] Mollenauer, L.F. (1985). In *Laser handbook*, Vol. 4 (ed. M.L. Stitch and M. Bass). Elsevier Science Publishers, North Holland, p. 144.
- [3] Van Uitert, L.G. (1966). In *Luminescence of inorganic solids* (ed. P. Goldberg). Academic Press, New York, p. 465.
- [4] Moulton, P.F. (1983). *Laser Focus* **14** (May), 83.
- [5] Smets, B. (1992). In *Optical properties of excited states in solids* (ed. B. di Bartolo), NATO ASI Series B, Vol. 301. Plenum Press, New York, p. 349.

Exercises

- (9.1) A colour centre may be modelled as an electron of mass m_0 confined to move in a cubic box with a cube edge length of $2a$. On the assumption that the potential barriers at the edge of the box are infinite, solve the Schrödinger equation for the electron and hence derive eqn 9.4.
- (9.2) The solid line in Fig. 9.5 is a fit to the data with $E = 0.21/a^2$, where E is measured in eV and a in nm. How does this fit compare to the prediction of eqn 9.4?
- (9.3) The anion–cation distance in KBr is 0.33 nm. Estimate the energy of the F-band absorption peak in this crystal.
- (9.4) An electron is trapped in a hard rectangular box with square ends orientated along the z axis. Calculate the energy of the electron if the length of the box is $2b$ and its cross-sectional area is b^2 . Hence explain why we might expect the transitions of an F_2^+ centre to be at about half the energy of the equivalent F centre. Does this model fit the experimental data for KF given in Figs 9.5 and 9.6?
- (9.5) According to the Bohr model of the atom, the radius of the n th quantum level in an atom with atomic number Z is proportional to n^2/Z . Use this model to argue that:
 - (a) The radius of the $3d$ orbitals in the transition metal series is larger than that of the $4f$ orbitals in the rare earth series.
 - (b) The $3d$ orbitals of a transition metal ion are the outermost orbitals of the ion, whereas the $4f$ orbitals of a rare earth ion are not.
- (9.6)* Consider the interaction between an electron in an outer p orbital with the electric field of a crystalline host environment.
 - (i) Explain why the p_x , p_y and p_z orbitals are degenerate if the ion is placed in an octahedral crystal, as sketched in Fig. 9.7.
 - (ii) Explain why the p states split into a singlet and a doublet if the crystal has uniaxial symmetry, that

- is, if the ions of the crystal host are closer along the z axis than in the x and y directions.
- (iii) State whether the energy of the singlet is higher or lower than that of the doublet if the nearest neighbour ions are negative.
- (9.7) Explain why the intensity of the $1.064\text{ }\mu\text{m}$ line of the Nd:YAG crystal is greater at 300 K than at 77 K . (See Fig. 9.9.)
- (9.8) Explain why population inversion between two levels gives rise to optical gain at the energy difference between the two levels.
- (9.9)* The lower level of the 694 nm laser line in a ruby crystal is the ground state. A bright flash lamp pulse pumps 60% of the atoms from the ground state to the upper laser level of a ruby rod, which then emits a short laser pulse. Calculate the maximum energy of this pulse if the laser rod has a volume of 10^{-6} m^3 and the doping density of the Cr^{3+} ions in the crystal is $1 \times 10^{25}\text{ m}^{-3}$.
- (9.10) A laser emits pulses with a Gaussian time dependence of the form $I(t) = I_0 \exp(-t^2/\tau^2)$. The centre frequency of the laser light is ω_0 .
- (i) By considering the Fourier transform of the electric field, show that the pulses have a spectrum of the form $I(\omega) = I(\omega_0) \exp[-\tau^2(\omega - \omega_0)^2]$.
- (ii) Hence show that the time-bandwidth product of the pulses, namely $\Delta\nu\Delta t$, where $\Delta\nu$ and Δt are the full width at half maximum of the pulse in the frequency and time domains respectively, is equal to $2\ln 2/\pi$.
- (9.11) The linewidth of the $1.054\text{ }\mu\text{m}$ transition of Nd^{3+} in a phosphate glass host is $7.5 \times 10^{12}\text{ Hz}$. Suggest a possible explanation for why this is about 60 times larger than that of the $1.064\text{ }\mu\text{m}$ line in Nd:YAG crystal. Estimate the duration of the shortest pulses that can be obtained from a Nd:glass laser.
- (9.12) Explain why the radiative lifetime for the $E_g \rightarrow T_{2g}$ transition in titanium doped sapphire is in the microsecond range. Would you classify this emission as fluorescence or phosphorescence?
- (9.13) The radiative lifetime of the upper laser level of Co:MgF_2 is 1.8 ms . The measured excited state lifetime decreases from 1.4 ms at 77 K to 0.06 ms at 300 K . Account for the temperature dependence of the excited state lifetime, and explain why the operating temperature for the Co:MgF_2 laser is 77 K and not 300 K .
- (9.14) A titanium-doped sapphire laser operating at 800 nm is pumped by an argon ion laser at 514 nm . Calculate the maximum possible power output if the pump power is 5 W , stating the assumptions you make. What happens to the energy that is not emitted as laser light?

* Exercises marked with an asterisk are more challenging.

10

Phonons

10.1	Infrared active phonons	204
10.2	Infrared reflectivity and absorption in polar solids	206
10.3	Polaritons	214
10.4	Polarons	215
10.5	Inelastic light scattering	218
10.6	Phonon lifetimes	222

In this chapter we will turn our attention to the interaction between light and the phonons in a solid. Phonons are vibrations of the atoms in a crystal lattice, and have resonant frequencies in the infrared spectral region. This contrasts with the optical properties of bound electrons, which occur at visible and ultraviolet frequencies.

The main optical properties of phonons can be explained to a large extent by classical models. We will therefore make extensive use of the classical dipole oscillator model developed in Chapter 2. This will allow us to understand why polar solids reflect and absorb light strongly within a band of infrared frequencies. We will then introduce the concepts of polaritons and polarons, before moving on to discuss the physics of inelastic light scattering. We will see how Raman and Brillouin scattering techniques give us complementary information to infrared reflectivity data, which is why they are so extensively used in phonon physics. Finally we will briefly discuss why phonons have a finite lifetime, and how this affects the reflectivity and inelastic scattering spectra.

We will assume that the reader is familiar with the basic physics of phonons, which is covered in all introductory solid state physics texts. A partial list of suitable preparatory reading is given under Further Reading at the end of the chapter.

10.1 Infrared active phonons

The atoms in a solid are bound to their equilibrium positions by the forces that hold the crystal together. When the atoms are displaced from their equilibrium positions, they experience restoring forces, and vibrate at characteristic frequencies. These vibrational frequencies are determined by the phonon modes of the crystal.

The resonant frequencies of the phonons occur in the infrared spectral region, and the modes that interact directly with light are called **infrared active** (IR active). Detailed selection rules for deciding which phonon modes are IR active can be derived by using group theory. At this level we just discuss the general rules based on the dispersion of the modes, their polarization, and the nature of the bonding in the crystal.

The phonon modes of a crystal are subdivided into two general categories:

- acoustic or optical;
- transverse or longitudinal.

It will come as no surprise to realize that it is the ‘optical’ rather than the acoustic modes that are directly IR active. These optically active phonons are

The group theory approach is beyond the scope of this book, although we will give some simple arguments based on symmetry when we consider inelastic light scattering in Section 10.5.

able to absorb light at their resonant frequency. The basic process by which a photon is absorbed by the lattice and a phonon is created is represented in Fig. 10.1. Conservation laws require that the photon and the phonon must have the same energy and momentum. We will see below that this condition can only be satisfied for the optical modes.

Figure 10.2 shows the generic dispersion curves for the acoustic and optical phonons in a simple crystal. The angular frequency Ω of the acoustic and optical phonons is plotted against the wave vector q in the positive half of the first Brillouin zone. At small wave vectors the slope of the acoustic branch is equal to v_s , the velocity of sound in the medium, while the optical modes are essentially dispersionless near $q = 0$.

The figure also shows the dispersion of the light waves in the crystal, which have a constant slope of $v = c/n$, where n is the refractive index. The refractive index has been highly exaggerated here in order to make the dispersion of the photon noticeable on the same scale as the phonon dispersion. The requirement that the photon and phonon should have the same frequency and wave vector is satisfied when the dispersion curves intersect. Since $c/n \gg v_s$, the only intersection point for the acoustic branch occurs at the origin, which corresponds to the response of the crystal to a static electric field. The situation is different for the optical branch: there is an intersection at finite ω , which is identified with the circle in Fig. 10.2. Since the optical branch is essentially flat for small q , the frequency of this resonance is equal to the frequency of the optical mode at $q = 0$.

Electromagnetic waves are transverse, and can only apply driving forces to the transverse vibrations of the crystal. Therefore they can only couple to the transverse optic (TO) phonon modes. This does not mean that we can now completely forget about the longitudinal optic (LO) phonons. As we will see in Section 10.2.2, the LO modes do in fact play an important role in the infrared properties of crystals.

Photons couple to phonons through the driving force exerted on the atoms by the AC electric field of the light wave. This can only happen if the atoms are charged. Therefore, if the atoms are neutral, there will be no coupling to

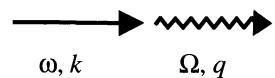


Fig. 10.1 Lattice absorption process by an infrared active phonon. The straight arrow represents the photon that is absorbed, while the wiggly arrow represent the phonon that is created.

The phonon dispersion curves for real crystals are more complicated than those shown in Fig. 10.2 because the longitudinal and transverse polarizations tend to have different frequencies.

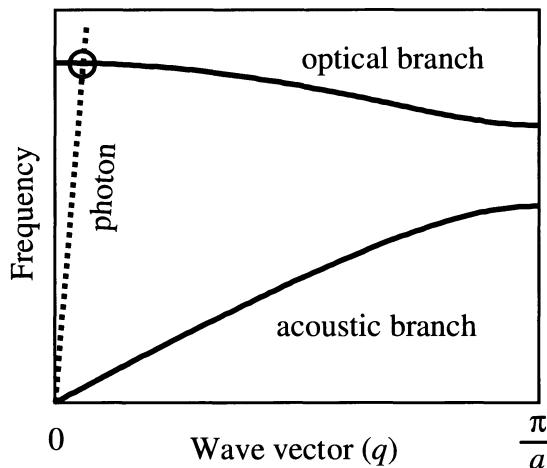


Fig. 10.2 Dispersion curves for the acoustic and optical phonon branches in a typical crystal with a lattice constant of a . The dispersion of the photon modes in the crystal is shown by the dotted line.

the light. This means that the crystal must have some ionic character in order for its TO phonons to be optically active.

The ionicity of a solid arises from the way the crystal binding occurs. An ionic crystal consists of an alternating sequence of positive and negative ions held together by their mutual Coulomb attraction. Covalent crystals, by contrast, consist of neutral atoms with the electrons shared equally between the neighbouring nuclei. This means that none of the optical phonons of purely covalent solids like silicon are IR active. Most other materials fall somewhere between these two limits. For example, the bond in a III–V semiconductor is only partly covalent, and the shared electrons lie slightly closer to the group V atoms than to the group III atoms, which gives the bond a partly ionic character. The bonds with an ionic character are called **polar** bonds to stress the point that the asymmetric electron cloud between the atoms creates a dipole that can interact with electric fields. Provided the bond has some polar character, its phonons can be IR active.

The conclusions of this section are summarized in Table 10.1.

Table 10.1 Infrared activity of the phonon modes in polar and non-polar crystals. LA: longitudinal acoustic, TA: transverse acoustic, LO: longitudinal optic, TO: transverse optic.

Mode	Polar crystal	Non-polar crystal
LA	no	no
TA	no	no
LO	no	no
TO	yes	no

10.2 Infrared reflectivity and absorption in polar solids

Experimental data show that polar solids absorb and reflect light very strongly in the infrared spectral region when the frequency is close to resonance with the TO phonon modes. We have come across several examples of this already. For example, the transmission spectra of sapphire and CdSe given in Fig. 1.4 show that there are spectral regions in the infrared where no light is transmitted. This is a consequence of lattice absorption.

The aim of this section is to account for this result by modelling the interaction of photons with TO phonons. To do this we will make extensive use of the classical oscillator model developed in Chapter 2, especially Section 2.2. This will allow us to calculate the frequency dependence of the complex dielectric constant $\tilde{\epsilon}_r(\omega)$, from which we will be able to determine the important optical properties such as the reflectivity and absorption.

10.2.1 The classical oscillator model

The interaction between electromagnetic waves and a TO phonon in an ionic crystal is most easily treated by considering a linear chain, as illustrated in Fig. 10.3. The chain consists of a series of unit cells, each containing a positive ion (black circle) and a negative ion (grey circle). The waves are taken to be propagating along the chain in the z direction. We are dealing with a transverse mode, and so the displacement of the atoms is in the x or y directions. Furthermore, in an optic mode the different atoms within each unit cell move in opposite directions, with a fixed ratio between their displacements which is not necessarily equal to unity.

We are interested in the interaction between a TO phonon mode with $q \approx 0$ and an infrared light wave of the same frequency and wave vector. This means that we are considering phonons with a very long wavelength of $\sim 10 \mu\text{m}$ matched to that of an infrared photon. This phonon wavelength is huge

LO modes do not interact with light because the displacement of the atoms generates a longitudinal electric field, which is perpendicular to that of the light wave.

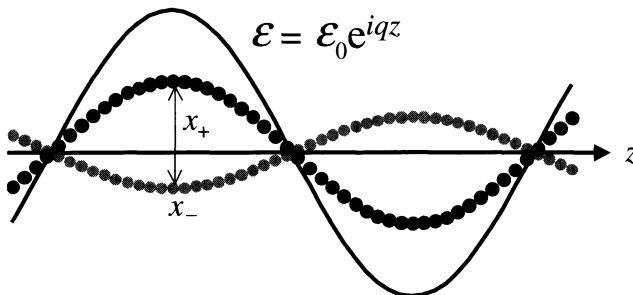


Fig. 10.3 Interaction of a TO phonon mode propagating in the z direction with an electromagnetic wave of the same wave vector. The black circles represent positive ions, while the grey circles represent the negative ions. The solid line represents the spatial dependence of the electric field of the electromagnetic wave.

compared to the size of a unit cell in a crystal, which is usually less than 10^{-9}m . The size of the atoms has been highly exaggerated in Fig. 10.3 to make the physics of the interaction clearer. In fact, the real size of the atoms is tiny compared to the wavelength, and there will be thousands of unit cells within one period of the wave.

The solid line in the figure represents the spatial dependence of the AC electric field of the infrared light wave. At resonance, the wave vector of the photon and the phonon are the same. This means that the driving force exerted by the light on the positive and negative ions is in phase with the lattice vibration. At the same time, the antiparallel displacements of the oppositely charged atoms generate an AC electric field in phase with the external light. This implies that there is a strong interaction between the TO phonon mode and the light wave when the wave vectors and frequencies match.

For long wavelength TO modes with $q \approx 0$, the motion of the atoms in different unit cells is almost identical, and we therefore need to concentrate on what is happening within the unit cell itself. This enables us to see that there is a close connection between the TO phonons at $q = 0$ and the vibrational modes of the molecules from which the crystal is formed. We can therefore make use of some of the principles developed in molecular physics, for example: the selection rules for deciding whether a particular phonon mode is IR or Raman active. (cf. Section 10.5.2.)

The interaction between the TO phonon and the light wave can be modelled by writing down the equations of motion for the displaced ions. The displacements of the positive and negative ions in a TO mode are in opposite directions and are given the symbols x_+ and x_- respectively, as indicated in Fig. 10.3. The appropriate equations of motion are:

$$m_+ \frac{d^2x_+}{dt^2} = -K(x_+ - x_-) + q\mathcal{E}(t) \quad (10.1)$$

$$m_- \frac{d^2x_-}{dt^2} = -K(x_- - x_+) - q\mathcal{E}(t) \quad (10.2)$$

where m_+ and m_- are the masses of the two ions, K is the restoring constant of the medium, and $\mathcal{E}(t)$ is the external electric field due to the light wave. The effective charge per ion is taken to be $\pm q$.

By dividing eqn 10.1 by m_+ and eqn 10.2 by m_- , and then subtracting, we obtain:

$$\frac{d^2}{dt^2}(x_+ - x_-) = -\frac{K}{\mu}(x_+ - x_-) + \frac{q}{\mu}\mathcal{E}(t), \quad (10.3)$$

The data for SiO_2 glass shown in Fig. 2.7 illustrates the connection between the infrared absorption in solids and that of the constituent molecules quite well. The glass is amorphous, and therefore does not have long range order with delocalized phonon modes. The absorption in the range 10^{13} – 10^{14}Hz is basically caused by the vibrational absorption of the SiO_2 molecules themselves, although the frequencies are not necessarily exactly the same in the solid as in the free molecule.

For a strongly ionic crystal such as NaCl , q would just be equal to $\pm e$. However, for crystals with polar covalent bonds such as the III–V compounds, q will represent an effective charge which is determined by the asymmetry of the electron cloud within the bond.

where μ is the reduced mass given by

$$\frac{1}{\mu} = \frac{1}{m_+} + \frac{1}{m_-}. \quad (10.4)$$

By putting $x = x_+ - x_-$ for the relative displacement of the positive and negative ions within their unit cell, we can recast eqn 10.3 in the simpler form:

$$\frac{d^2x}{dt^2} + \Omega_{\text{TO}}^2 x = \frac{q}{\mu} \mathcal{E}(t), \quad (10.5)$$

where we have written Ω_{TO}^2 for K/μ . Ω_{TO} represents the natural vibrational frequency of the TO mode at $q = 0$ in the absence of the external light field.

Equation 10.5 represents the equation of motion for undamped oscillations of the lattice driven by the forces exerted by the AC electric field of the light wave. In reality, we should have incorporated a damping term to account for the finite lifetime of the phonon modes. The physical significance of the phonon lifetime will be discussed further in Section 10.6. At this stage, we simply introduce a phenomenological damping rate γ , and rewrite eqn 10.5 as

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \Omega_{\text{TO}}^2 x = \frac{q}{\mu} \mathcal{E}(t). \quad (10.6)$$

This now represents the response of a damped TO phonon mode to a resonant light wave.

Equation 10.6 is identical in form to eqn 2.5 in Chapter 2, with m_0 replaced by μ , ω_0 by Ω_{TO} and $-e$ by q . Therefore, we can use all the results derived in Section 2.2 to model the response of the medium to a light field of angular frequency ω with $\mathcal{E}(t) = \mathcal{E}_0 e^{i\omega t}$. In particular, we can go directly to the formula for the frequency dependence of the dielectric constant without repeating all the steps in the derivation. By adapting the symbols appropriately in eqn 2.14, we immediately write down:

$$\epsilon_r(\omega) = 1 + \chi + \frac{Nq^2}{\epsilon_0 \mu} \frac{1}{(\Omega_{\text{TO}}^2 - \omega^2 - i\gamma\omega)}, \quad (10.7)$$

where $\epsilon_r(\omega)$ is the complex dielectric constant at angular frequency ω . χ represents the non-resonant susceptibility of the medium, and N is the number of unit cells per unit volume.

Equation 10.7 can be tidied up by introducing the static and high frequency dielectric constants ϵ_{st} and ϵ_∞ respectively. In the limits of low and high frequency, we obtain from eqn 10.7:

$$\epsilon_{st} \equiv \epsilon_r(0) = 1 + \chi + \frac{Nq^2}{\epsilon_0 \mu \Omega_{\text{TO}}^2}, \quad (10.8)$$

and

$$\epsilon_\infty \equiv \epsilon_r(\infty) = 1 + \chi. \quad (10.9)$$

Thus we can write:

$$\epsilon_r(\omega) = \epsilon_\infty + (\epsilon_{st} - \epsilon_\infty) \frac{\Omega_{\text{TO}}^2}{(\Omega_{\text{TO}}^2 - \omega^2 - i\gamma\omega)}. \quad (10.10)$$

In principle, we should consider the local field corrections discussed in Section 2.2.4 here. This is an unnecessary complication at this level which does not add much to the main conclusions. We will therefore neglect local field effects, and base our discussion on eqn 10.10.

This is our main result, which will be used in the next subsections to derive the infrared optical coefficients. As discussed in Section 2.2.2, and in particular in connection with Fig. 2.6, we should understand ‘ $\omega = \infty$ ’ in a relative sense here. ϵ_∞ represents the dielectric constant at frequencies well above the phonon resonance, but below the next natural frequency of the crystal due, for example, to the bound electronic transitions in the visible/ultraviolet spectral region.

10.2.2 The Lyddane–Sachs–Teller relationship

Before working out the frequency dependence of the infrared reflectivity, it is useful to investigate one rather striking implication of eqn 10.10. Suppose we have a lightly damped system so that we can set $\gamma = 0$. Then at a certain frequency which we label ω' , eqn 10.10 tells us that the dielectric constant can fall to zero. The condition for this to happen is:

$$\epsilon_r(\omega') = 0 = \epsilon_\infty + (\epsilon_{st} - \epsilon_\infty) \frac{\Omega_{TO}^2}{(\Omega_{TO}^2 - \omega'^2)}. \quad (10.11)$$

This can be solved to obtain:

$$\omega' = \left(\frac{\epsilon_{st}}{\epsilon_\infty} \right)^{\frac{1}{2}} \Omega_{TO}. \quad (10.12)$$

What does $\epsilon_r = 0$ mean physically? In a medium with no free charges, the total charge density will be zero. Hence Gauss’s law (eqn A.10) tells us that

$$\nabla \cdot \mathbf{D} = \nabla \cdot (\epsilon_r \epsilon_0 \mathbf{E}) = 0, \quad (10.13)$$

where we have made use of eqn A.3 to relate the electric displacement \mathbf{D} to the electric field \mathbf{E} in a dielectric medium. When we consider the propagation of electromagnetic waves through the dielectric, we look for wave solutions of the form:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}. \quad (10.14)$$

On substituting eqn 10.14 into eqn 10.13, we usually assume that $\epsilon_r \neq 0$ and therefore conclude that $\mathbf{k} \cdot \mathbf{E} = 0$. This tells us that the electric field must be perpendicular to the direction of the wave and therefore that the waves are transverse. However, if $\epsilon_r = 0$, we can satisfy eqn 10.13 with waves in which $\mathbf{k} \cdot \mathbf{E} \neq 0$, that is, with longitudinal waves. Thus we conclude that the dielectric can support longitudinal electric field waves at frequencies which satisfy $\epsilon_r(\omega) = 0$.

In the same way that TO phonon modes generate a transverse electric field wave, the LO phonon modes generate a longitudinal electric field wave. Thus the waves at $\omega = \omega'$ correspond to LO phonon waves, and we identify ω' with the frequency of the LO mode at $q = 0$, namely Ω_{LO} . This allows us to rewrite eqn 10.12 in the following form:

$$\frac{\Omega_{LO}^2}{\Omega_{TO}^2} = \frac{\epsilon_{st}}{\epsilon_\infty}. \quad (10.15)$$

We came across another situation in which ϵ_r is zero when we discussed plasmons in Section 7.5. We saw there that the plasma oscillations that cause ϵ_r to be zero also correspond to longitudinal electric field waves.

Table 10.2 Comparison of the measured ratio $\Omega_{\text{LO}}/\Omega_{\text{TO}}$ for several materials to the value predicted by the Lyddane–Sachs–Teller relationship. After [1].

Crystal	$\Omega_{\text{LO}}/\Omega_{\text{TO}}$	$(\epsilon_{\text{st}}/\epsilon_{\infty})^{\frac{1}{2}}$
Si	1	1
GaAs	1.07	1.08
AlAs	1.12	1.11
BN	1.24	1.26
ZnSe	1.19	1.19
MgO	1.81	1.83
AgF	1.88	1.88

This result is known as the **Lyddane–Sachs–Teller (LST) relationship**. The validity of the relationship can be checked by comparing the values of $\Omega_{\text{LO}}/\Omega_{\text{TO}}$ deduced from neutron or Raman scattering experiments with those calculated from eqn 10.15 using known values of the dielectric constants. Some results are given in Table 10.2. It is apparent that the agreement is generally very good.

An interesting corollary of the LST relationship is that it implies that the LO phonon and TO phonon modes of non-polar crystals are degenerate. This follows because there is no infrared resonance, and therefore $\epsilon_{\text{st}} = \epsilon_{\infty}$. This is indeed the case for the purely covalent crystals of the group IV elements, namely diamond (C), silicon and germanium.

10.2.3 Restrahlen

Having discussed the properties of the system at the special frequency of $\omega = \Omega_{\text{LO}}$, we can now calculate the infrared optical constants. It is easier to understand the general behaviour if we assume that the damping term is small. We thus set $\gamma = 0$ in eqn 10.10, and discuss the properties of a material with a dielectric constant that has the following frequency dependence:

$$\epsilon_r(\nu) = \epsilon_{\infty} + (\epsilon_{\text{st}} - \epsilon_{\infty}) \frac{\nu_{\text{TO}}^2}{(\nu_{\text{TO}}^2 - \nu^2)}. \quad (10.16)$$

We have divided all the angular frequencies by 2π here, so that we can compare the predictions to experimental data, which are usually presented against frequency (ν) rather than angular frequency (ω). We will discuss the effect of including the damping term when we compare our model to the experimental data in connection with Fig. 10.5.

Figure 10.4(a) plots the frequency dependence of the dielectric constant $\epsilon_r(\nu)$ calculated from eqn 10.16 for a polar crystal with the following parameters: $\nu_{\text{TO}} = 10 \text{ THz}$, $\nu_{\text{LO}} = 11 \text{ THz}$, $\epsilon_{\text{st}} = 12.1$ and $\epsilon_{\infty} = 10$. These figures are quite close to those that would be found in a typical III–V semiconductor. Note that the phonon frequencies have been chosen to satisfy the LST relationship given in eqn 10.15.

1 THz = 10^{12} Hz

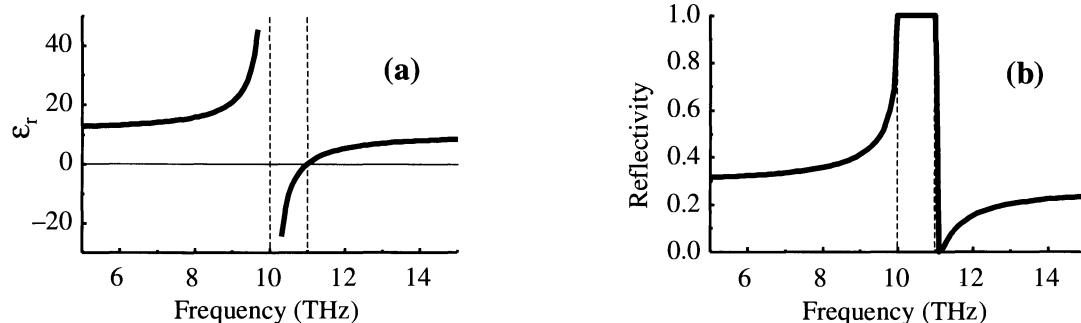


Fig. 10.4 Frequency dependence of the dielectric constant and reflectivity for a crystal with $\nu_{\text{TO}} = 10 \text{ THz}$, $\nu_{\text{LO}} = 11 \text{ THz}$, $\epsilon_{\text{st}} = 12.1$ and $\epsilon_{\infty} = 10$. The curves have been calculated from eqns 10.16 and 10.17. Phonon damping is ignored in this calculation.

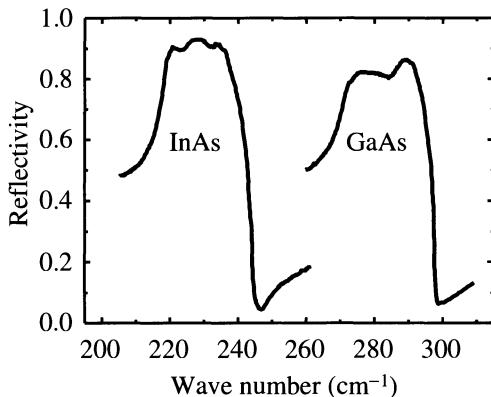


Fig. 10.5 Infrared reflectivity of InAs and GaAs at 4.2 K. A wave number of 1 cm^{-1} is equivalent to a frequency of $2.998 \times 10^{10}\text{ Hz}$. After [2], copyright © 1967 by Academic Press, reproduced by permission of the publisher.

At low frequencies the dielectric constant is just equal to ϵ_{st} . As ν increases from 0, $\epsilon_r(\nu)$ gradually increases until it diverges when the resonance at ν_{TO} is reached. Between ν_{TO} and ν_{LO} , ϵ_r is negative. Precisely at $\nu = \nu_{LO}$, $\epsilon_r = 0$. Thereafter, ϵ_r is positive, and gradually increases asymptotically towards the value of ϵ_∞ .

The most important optical property of a polar solid in the infrared spectral region is the reflectivity. This can be calculated from the dielectric constant using eqn 1.26:

$$R = \left| \frac{\tilde{n} - 1}{\tilde{n} + 1} \right|^2 = \left| \frac{\sqrt{\epsilon_r} - 1}{\sqrt{\epsilon_r} + 1} \right|^2. \quad (10.17)$$

Figure 10.4(b) plots the reflectivity calculated using eqn 10.17 for the dielectric constant shown in Fig. 10.4(a). At low frequencies the reflectivity is $(\sqrt{\epsilon_{st}} - 1)^2 / (\sqrt{\epsilon_{st}} + 1)^2$. As ν approaches ν_{TO} , R increases towards unity. In the frequency region between ν_{TO} and ν_{LO} , $\sqrt{\epsilon_r}$ is imaginary, so that R remains equal to unity. R drops rapidly to zero as ν increases above ν_{LO} (see Exercise 10.2), and then increases gradually towards the high frequency asymptote of $(\sqrt{\epsilon_\infty} - 1)^2 / (\sqrt{\epsilon_\infty} + 1)^2$.

We see from this analysis that the reflectivity is equal to 100 % in the frequency region between ν_{TO} and ν_{LO} . This frequency region is called the **restrahlen** band. **Restrahlen** is the German word for ‘residual rays’. Light cannot propagate into the medium in the restrahlen band.

Figure 10.5 shows experimental data for the reflectivity of InAs and GaAs in the infrared spectral region. InAs has TO and LO phonon frequencies at 218.9 cm^{-1} and 243.3 cm^{-1} respectively, while for GaAs we have $\nu_{TO} = 273.3\text{ cm}^{-1}$ and $\nu_{LO} = 297.3\text{ cm}^{-1}$. We see that the reflectivity is very high for frequencies between the TO and LO phonon frequencies in both materials, and there is a sharp dip in the reflectivity just above the LO phonon resonance.

On comparing these results with the prediction shown in Fig. 10.4(b), we see that the general agreement between the model and the experimental data is very good. The main difference is that in both materials the maximum reflectivity in the restrahlen band is less than 100 %. This reduction in the reflectivity is caused by ignoring the damping term. (See Example 10.1 and Exercise 10.4.) The damping also broadens the edge so that there is only a minimum in R just above ν_{LO} rather than a zero.

We can see from eqn 1.17 that $\sqrt{\epsilon_\infty}$ corresponds to the refractive index of the medium at frequencies well above the optical phonon resonances. This will be the refractive index measured at near-infrared and visible frequencies the below band gap of the material.

Experimental infrared spectra are frequently plotted against the wave number $\bar{\nu} \equiv 1/\lambda$. The wave number is effectively a frequency unit, with 1 cm^{-1} equivalent to $2.998 \times 10^{10}\text{ Hz}$.

The magnitude of γ can be found by fitting the experimental data to the full dependence given in eqn 10.10. The values of γ obtained in this way are around 10^{11} – 10^{12} s $^{-1}$, which implies that the optical phonons have a lifetime of about 1–10 ps. The physical significance of this short lifetime will be discussed in Section 10.6.

$1\text{ ps} = 10^{-12}\text{ s}$.

10.2.4 Lattice absorption

When we introduced the classical oscillator model in Section 2.2 of Chapter 2, we made the point that we expect high absorption coefficients whenever the frequency matches the natural resonances of the medium. The reader might therefore be wondering why we have been concentrating on calculating the reflectivity rather than the absorption due to the TO phonon resonances.

This question is further prompted by recalling the analogy between the infrared absorption of polar solids and that of isolated molecules. In both cases we are basically treating the interaction of photons with quantized vibrational modes. In molecular physics we usually discuss this in terms of the infrared absorption spectrum. The absorption spectra show strong peaks whenever the frequency coincides with the infrared active vibrational modes and the molecule can absorb a photon by creating one vibrational quantum. This is directly analogous to the process for solids shown in Fig. 10.1 in which a photon is absorbed and a phonon is created.

The answer to these questions is that the lattice does indeed absorb very strongly whenever the photon is close to resonance with the TO phonon. As stressed in Chapter 2, the fundamental optical properties of a dielectric – the absorption, refraction and reflectivity – are all related to each other because they are all determined by the complex dielectric constant. The distinction between absorption and reflection is merely a practical one. Polar solids have such high absorption coefficients in the infrared that unless the crystal is less than $\sim 1\text{ }\mu\text{m}$ thick, no light at all will be transmitted. This is clearly seen in the transmission spectra of Al_2O_3 and CdSe shown in Fig. 1.4. For this reason, it is only sensible to consider lattice absorption in thin film samples. In thick crystals, we must use reflectivity measurements to determine the vibrational frequencies. This contrasts with molecular physics, where we are usually dealing with low density gases, which give rise to much smaller absorption coefficients.

The absorption coefficients expected at the resonance with the TO phonon can be calculated from the imaginary part of the dielectric constant. At $\omega = \Omega_{\text{TO}}$ we have from eqn 10.10:

$$\epsilon_r(\Omega_{\text{TO}}) = \epsilon_\infty + i(\epsilon_{\text{st}} - \epsilon_\infty) \frac{\Omega_{\text{TO}}}{\gamma}. \quad (10.18)$$

The extinction coefficient κ can be worked out from ϵ_r using eqn 1.23, and then the absorption coefficient α can be determined from κ using eqn 1.16. Typical values for α are in the range 10^6 – 10^7 m^{-1} . (See Example 10.1 and Exercise 10.6.) This is why the sample must be thinner than $\sim 1\text{ }\mu\text{m}$ in order to perform practical absorption measurements. Infrared absorption measurements on thin film samples do indeed confirm that the absorption is very high at the TO phonon resonance frequency.

Example 10.1

The static and high frequency dielectric constants of NaCl are $\epsilon_{st} = 5.9$ and $\epsilon_\infty = 2.25$ respectively, and the TO phonon frequency ν_{TO} is 4.9 THz.

- Calculate the upper and lower wavelengths of the restrahlen band.
- Estimate the reflectivity at 50 μm , if the damping constant γ of the phonons is 10^{12} s^{-1} .
- Calculate the absorption coefficient at 50 μm .

Solution

- The restrahlen band runs from ν_{TO} to ν_{LO} . We are given ν_{TO} , and we can calculate ν_{LO} from the LST relationship (eqn 10.15). This gives

$$\nu_{LO} = \left(\frac{\epsilon_{st}}{\epsilon_\infty} \right)^{\frac{1}{2}} \times \nu_{TO} = \left(\frac{5.9}{2.25} \right)^{\frac{1}{2}} \times 4.9 \text{ THz} = 7.9 \text{ THz}.$$

Therefore the restrahlen band runs from 4.9 THz to 7.9 THz, or 38 μm to 61 μm .

- At 50 μm we are in middle of the restrahlen band. We therefore expect the reflectivity to be high. We insert the values for ϵ_{st} , ϵ_∞ , γ and $\Omega_{TO} = 2\pi\nu_{TO}$ into eqn 10.10 with $\omega = 2\pi\nu$ ($\nu = 6 \text{ THz}$) to find:

$$\epsilon_r = 2.25 + 3.65 \frac{(4.9)^2}{(4.9)^2 - 6^2 - i(1)(6)/2\pi} = -5.0 + 0.57i.$$

We then obtain the real and imaginary parts of the refractive index from eqns 1.22 and 1.23:

$$n = \frac{1}{\sqrt{2}} \left(-5.0 + [(-5.0)^2 + (0.57)^2]^{\frac{1}{2}} \right)^{\frac{1}{2}} = 0.13,$$

and

$$\kappa = \frac{1}{\sqrt{2}} \left(+5.0 + [(-5.0)^2 + (0.57)^2]^{\frac{1}{2}} \right)^{\frac{1}{2}} = 2.2.$$

We finally substitute these values of n and κ into eqn 1.26 to find the reflectivity:

$$R = \frac{(n-1)^2 + \kappa^2}{(n+1)^2 + \kappa^2} = \frac{(-0.87)^2 + (2.2)^2}{(1.13)^2 + (2.2)^2} = 0.91.$$

This value is close to the measured reflectivity of NaCl in the restrahlen band at room temperature.

- We can calculate the absorption coefficient α from the extinction coefficient using eqn 1.16. We have already worked out that $\kappa = 2.2$ in part (ii). Hence we find:

$$\alpha = \frac{4\pi\kappa}{\lambda} = \frac{4\pi \times 2.2}{50 \times 10^{-6}} = 5.5 \times 10^5 \text{ m}^{-1}.$$

This shows that the light would be absorbed in a thickness of about 2 μm .

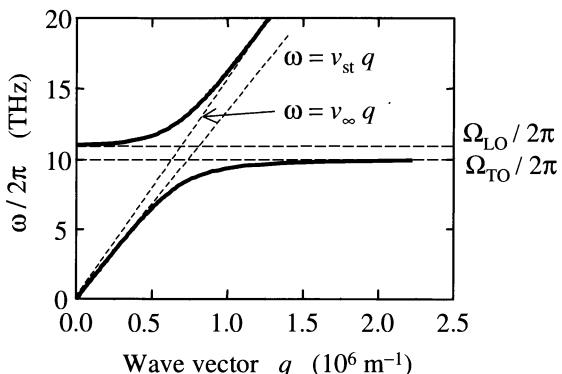


Fig. 10.6 Polariton dispersion predicted from eqn 10.19 with ϵ_r given by eqn 10.16. The curves are calculated for a crystal with $v_{TO} = 10 \text{ THz}$, $\epsilon_{st} = 12.1$ and $\epsilon_\infty = 10$. The asymptotic velocities v_{st} and v_∞ are equal to $c/\sqrt{\epsilon_{st}}$ and $c/\sqrt{\epsilon_\infty}$ respectively.

10.3 Polaritons

The dispersion curves of the photons and TO phonons were discussed in broad terms in connection with Fig. 10.2. We now wish to consider the circled intersection point in Fig. 10.2 in more detail. As we will see, the two dispersion curves do not actually cross each other. This is a consequence of the strong coupling between the TO phonons and the photons when their frequencies and wave vectors match. This leads to the characteristic anticrossing behaviour which is observed in many coupled systems.

The coupled phonon–photon waves are called **polaritons**. As the name suggests, these classical waves are mixed modes which have characteristics of both polarization waves (the TO phonons) and the photons. The dispersion of the polaritons can be deduced from the relationship:

$$\omega = vq = \frac{c}{\sqrt{\epsilon_r}} q, \quad (10.19)$$

where the second part of the equation comes from eqn A.29, with $\mu_r = 1$. The resonant response of the polar solid is contained implicitly in the frequency dependence of ϵ_r .

Figure 10.6 shows the polariton dispersion calculated for a lightly damped medium. The dielectric constant is given by eqn 10.16, and is plotted for the same parameters as in Fig. 10.4(a). At low frequencies the dielectric constant is equal to ϵ_{st} , and the dispersion of the modes is given by $\omega = cq/\sqrt{\epsilon_{st}}$. As ω approaches Ω_{TO} , the dielectric constant increases, and the velocity of the waves decreases, approaching zero at Ω_{TO} itself. For frequencies in the restrahlen band between Ω_{TO} and Ω_{LO} , the dielectric constant is negative. No modes can propagate, and all the photons that are incident on the medium are reflected. For frequencies above Ω_{LO} , ϵ_r is positive again and propagating modes are possible once more. The velocity of the waves gradually increases with increasing frequency, approaching a value of $c/\sqrt{\epsilon_\infty}$ at high frequencies.

The dispersion of the polariton modes has been measured for a number of materials. Figure 10.7 shows the measured dispersion of the TO phonons and LO phonons in GaP at small wave vectors. The results were obtained by Raman scattering techniques. (See Section 10.5.2.) The experimental data reproduce very well the polariton dispersion model indicated in Fig. 10.6. The solid line

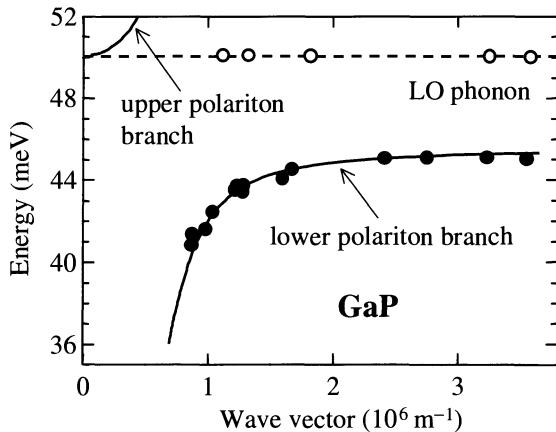


Fig. 10.7 Dispersion of the TO and LO phonons in GaP measured by Raman scattering. The solid lines are the predictions of the polariton model with $\hbar\nu_{\text{TO}} = 45.5 \text{ meV}$, $\epsilon_\infty = 9.1$ and $\epsilon_{\text{st}} = 11.0$. After [3], copyright 1965 American Institute of Physics, reprinted with permission.

is the calculated polariton dispersion, which gives a very accurate fit to the experimental points. Note that the LO phonons do not show any dispersion here because they do not couple to the light waves.

10.4 Polarons

So far in this chapter we have been considering the direct interaction between a light wave and the phonons in a crystal. As we have seen, this gives rise to strong absorption and reflection in the infrared spectral region. The optical phonons can, however, contribute indirectly to a whole host of other optical properties that depend primarily on the electrons through the **electron–phonon coupling**. In this section we will consider the **polaron** effect, which is one of the most important examples of this.

Consider the motion of a free electron through a polar solid, as shown in Fig. 10.4. The electron will attract the positive ions that are close to it, and repel the negative ones. This produces a local displacement of the lattice in the immediate vicinity of the electron. The lattice distortion accompanies the electron as it moves through the crystal. The electron with its local lattice

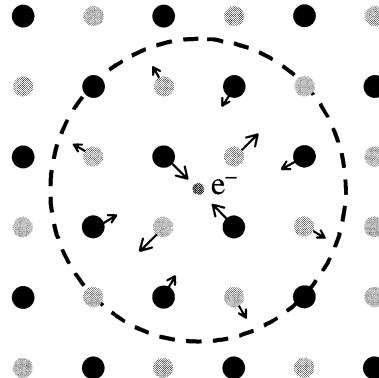


Fig. 10.8 Schematic representation of a polaron. A free electron moving through an ionic lattice attracts the positive (black) ions, and repels the negative (grey) ones. This produces a local distortion of the lattice within the polaron radius shown by the dashed circle

It can be shown that the average number of virtual LO phonons that move with the electron is equal to $\alpha_{\text{ep}}/2$. We do not consider the longitudinal acoustic modes here because they do not produce a polarization in the medium: the positive and negative ions move in the same direction, and this produces no electric dipole moment.

The polaron theory can be applied equally to electrons or holes by taking the appropriate effective masses in the formulae. In a non-polar crystal such as silicon, $\epsilon_{\infty} = \epsilon_{\text{st}}$, and $\alpha_{\text{ep}} = 0$. There is therefore no polaron effect.

distortion is equivalent to a new elementary excitation of the crystal, and is called a polaron.

The polaron effect can be conceived in terms of an electron surrounded by a cloud of virtual phonons. We think of the electron absorbing and emitting phonons as it moves through the crystal. These phonons produce the local lattice distortion. The displacement of the ions is in the same direction as the electric field of the electron, we are therefore dealing with longitudinal optic phonons.

The strength of the electron–phonon interaction in a polar solid can be quantified by the dimensionless coupling constant α_{ep} , which is given by:

$$\alpha_{\text{ep}} = \frac{1}{137} \left(\frac{m^* c^2}{2\hbar\Omega_{\text{LO}}} \right)^{\frac{1}{2}} \left[\frac{1}{\epsilon_{\infty}} - \frac{1}{\epsilon_{\text{st}}} \right], \quad (10.20)$$

where $1/137$ is the fine structure constant from atomic physics. The mass m^* that appears here is the usual effective mass deduced from the curvature of the band structure (c.f. eqn C.6):

$$m^* = \hbar^2 \left(\frac{d^2 E}{dk^2} \right)^{-1}. \quad (10.21)$$

Values for α_{ep} for three binary compound semiconductors, namely GaAs, ZnSe and AgCl, are given in Table 10.3. We see that the coupling constant increases from GaAs (0.06) through ZnSe (0.40) to AgCl (2.2). This is because the ionicity increases as we go from the III–V semiconductor, in which the bonding is predominantly covalent, to the I–VII compound, which is highly ionic.

This effective mass given by eqn 10.21 is calculated by assuming that the lattice is rigid. However, the concept of a rigid lattice is only a theoretical one, and any experiment we perform to measure m^* will actually measure the polaron mass m^{**} instead. This is because it is not possible to hold the lattice rigid as the electron moves. The polaron mass is larger than the rigid lattice mass because the electron has to drag the local lattice distortion with it as it moves.

An example of an experiment to measure the effective mass is **cyclotron resonance**. In this technique, we measure the infrared absorption in the presence of a magnetic field B . As discussed in Section 3.3.6, the electron energy is quantized in terms of the cyclotron energy:

$$E_n = (n + \frac{1}{2})\hbar\omega_c, \quad (10.22)$$

where n is an integer, and

$$\omega_c = \frac{eB}{m^*}. \quad (10.23)$$

Optical transitions with $\Delta n = \pm 1$ can take place between the ladder of levels defined by eqn 10.22. We therefore observe absorption at a wavelength λ given by:

$$\frac{hc}{\lambda} = \frac{e\hbar B}{m^*}. \quad (10.24)$$

This absorption usually occurs in the far-infrared spectral region, and the effective mass can be deduced from the values of λ and B at resonance. In a

Table 10.3 Electron-phonon coupling constant α_{ep} calculated from eqn 10.20 for GaAs, ZnSe, and AgCl. The figures for ZnSe are for the cubic crystal structure. After [1].

	GaAs	ZnSe	AgCl
m_e^*/m_0	0.067	0.13	0.30
ϵ_{∞}	10.9	5.4	3.9
ϵ_{st}	12.4	7.6	11.1
Ω_{LO} (THz)	53.7	47.7	36.9
α_{ep}	0.06	0.40	2.2

typical experiment, we use a fixed wavelength source from an infrared laser and find the value of B which gives the maximum absorption. For example, the cyclotron resonance occurs at about 6.1 T in GaAs ($m^* = 0.067m_0$) for the $118\ \mu\text{m}$ line from a methanol laser. The effective mass we find this way is the polaron mass m^{**} , not the value determined by the curvature of the bands given by eqn 10.21.

If the electron–phonon coupling constant α_{ep} is small, we can give an explicit relationship between the rigid lattice effective mass m^* and the polaron mass m^{**} :

$$\frac{m^{**}}{m^*} = \frac{1}{1 - \alpha_{\text{ep}}/6} \approx 1 + \frac{1}{6}\alpha_{\text{ep}}. \quad (10.25)$$

Values of m^* are actually worked from the measured values of m^{**} by applying eqn 10.25. For III–V semiconductors like GaAs with $\alpha_{\text{ep}} < 0.1$, m^{**} only differs from m^* by about 1 %. The polaron effect is thus only a small correction. This correction becomes more significant for II–VI compounds (e.g. $\sim 7\%$ for ZnSe). With highly ionic crystals like AgCl, the small α_{ep} approximation is not valid. The actual polaron mass of AgCl is 0.43, which is about 50 % larger than the rigid lattice value.

It can be shown that, in addition to the change of the mass, the polaron effect causes a reduction in the band gap by an amount:

$$\Delta E_g = -\alpha_{\text{ep}} \hbar \Omega_{\text{LO}}. \quad (10.26)$$

With a III–V material like GaAs, this again produces only a relatively small effect: $\Delta E_g \sim -0.1\%$. In practice, when we measure E_g by optical spectroscopy we always measure the polaron value.

Another important parameter of the polaron is its radius, r_p , which specifies how far the lattice distortion extends. This is depicted schematically in Fig. 10.4 by the dashed circle drawn around the electron that causes the lattice distortion. If α_{ep} is small, we can give an explicit formula for r_p :

$$r_p = \left(\frac{\hbar}{2m^* \Omega_{\text{LO}}} \right)^{\frac{1}{2}}. \quad (10.27)$$

This gives $r_p = 4.0\ \text{nm}$ for GaAs and $3.1\ \text{nm}$ for ZnSe. Both values are significantly larger than the unit cell size ($\sim 0.5\ \text{nm}$), which is important because the theory used to derive eqns 10.25–10.27 assumes that we can treat the medium as a polarizable continuum. This approximation is only valid if the radius of the polaron is very much greater than the unit cell size. A polaron which satisfies this criterion is called a **large polaron**. In highly ionic solids such as AgCl and the alkali halides, α_{ep} is not small and the polaron radius is comparable to the unit cell size. In this case we have a **small polaron**. The mass and radius have to be calculated from first principles.

The small polaron effect in highly ionic crystals leads to **self-trapping** of the charge carriers. The local lattice distortion is very strong, and the charge carrier can get completely trapped in its own lattice distortion. The carrier effectively digs itself into a pit and cannot get out of it. This is particularly the case for the holes in alkali halide crystals. The only way they can move is by **hopping** to a new site. The electrical conductivity of most alkali halide crystals is limited by this thermally activated hopping process at room temperature.

A particularly clear manifestation of the electron–phonon coupling can be observed in cyclotron resonance experiments when $B = m^* \Omega_{\text{LO}}/e$, so that $\omega_c = \Omega_{\text{LO}}$. The degenerate electron and phonon modes anticross with each other as the field is swept through this condition, and the cyclotron resonance line splits into a doublet. The magnitude of the splitting is directly proportional to the electron–phonon coupling constant α_{ep} . This effect was first observed in n-type InSb.

Polaronic hopping effects are also important in the conduction processes in organic semiconductors like polydiacetylene.

Self-trapping effects are important in determining the energies of Frenkel excitons. As discussed in Section 4.5, these are bound electron–hole pairs localized at individual atom or molecule sites within the lattice. The self-trapping of either the electron or hole can exacerbate the tendency for the exciton to localize, thereby instigating the transition from Wannier (free) to Frenkel exciton behaviour. The ground state excitons observed in many alkali halide, rare gas and organic crystals are of the self-trapped Frenkel type.

10.5 Inelastic light scattering

Inelastic light scattering describes the phenomenon by which a light beam is scattered by an optical medium and changes its frequency in the process. It contrasts with elastic light scattering, in which the frequency of the light is unchanged. The interaction process is illustrated in Fig. 10.9. Light incident with angular frequency ω_1 and wave vector \mathbf{k}_1 is scattered by an excitation of the medium of frequency Ω and wave vector \mathbf{q} . The scattered photon has frequency ω_2 and wave vector \mathbf{k}_2 . Inelastic light scattering can be mediated by many different types of elementary excitations in a crystal, such as phonons, magnons or plasmons. In this chapter we will be concerned exclusively with phonon processes.

Inelastic light scattering from phonons is generally subdivided as to whether it is the optical or acoustic phonons that are involved:

- **Raman scattering.** This is inelastic light scattering from optical phonons.
- **Brillouin scattering.** This is inelastic light scattering from acoustic phonons.

The physics of the two processes is essentially the same, but the experimental techniques differ. We will thus consider the general principles first, and then consider the details of each technique separately.

10.5.1 General principles of inelastic light scattering

Inelastic light scattering can be subdivided into two generic types:

- **Stokes scattering;**
- **Anti-Stokes scattering.**

Stokes scattering corresponds to the emission of a phonon (or some other type of material excitation), while anti-Stokes scattering corresponds to phonon absorption. The interaction shown in Fig. 10.9 is thus a Stokes process. Conservation of energy during the interaction requires that:

$$\omega_1 = \omega_2 \pm \Omega, \quad (10.28)$$

while conservation of momentum gives:

$$\mathbf{k}_1 = \mathbf{k}_2 \pm \mathbf{q}. \quad (10.29)$$

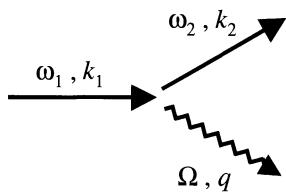


Fig. 10.9 An inelastic light scattering process. The straight arrows represent photons, while the wiggly arrow represents the phonon. The process shown corresponds to Stokes scattering in which the photon is shifted to lower frequency.

The + signs in eqns 10.28 and 10.29 correspond to phonon emission (Stokes scattering), while the - signs correspond to phonon absorption (anti-Stokes scattering). Thus the light is shifted down in frequency during a Stokes process, and up in frequency in an anti-Stokes event.

Anti-Stokes scattering will only be possible if there are phonons present in the material before the light is incident. The probability for anti-Stokes scattering therefore decreases on lowering the temperature as the phonon populations decrease. This means that the probability for anti-Stokes scattering from optical phonons is very low at cryogenic temperatures. On the other hand, Stokes scattering does not require a phonon to be present and can therefore occur at any temperature. The full quantum mechanical treatment shows that the ratio of anti-Stokes to Stokes scattering events is given by:

$$\frac{I_{\text{anti-Stokes}}}{I_{\text{Stokes}}} = \exp(-\hbar\Omega/k_B T). \quad (10.30)$$

This will be the ratio of the intensities of the anti-Stokes and Stokes lines observed in the Raman or Brillouin spectra.

The frequencies of the phonons involved can be deduced from the frequency shift of the scattered light using eqn 10.28. Thus the main use of inelastic light scattering is to measure phonon frequencies. This means that inelastic light scattering can give complementary information to that obtained from the infrared spectra. For example, infrared reflectivity measurements tell us nothing about the acoustic phonons, but we can measure the frequencies of some of the acoustic modes using Brillouin scattering experiments. We will consider this complementarity in more detail when we discuss the selection rules for Raman scattering in subsection 10.5.2 below.

The maximum phonon frequency in a typical crystal is about $10^{12} - 10^{13}$ Hz. This is almost two orders of magnitude smaller than the frequency of a photon in the visible spectral region. Equation 10.28 therefore tells us that the maximum frequency shift for the photon will be around 1 %. The wave vector of the photon is directly proportional to its frequency, and we can therefore make the approximation:

$$|\mathbf{k}_2| \approx |\mathbf{k}_1| = \frac{n\omega}{c}, \quad (10.31)$$

where n is the refractive index of the crystal and ω is the angular frequency of the incoming light.

We know from eqn 10.29 that $|\mathbf{q}| = |\mathbf{k}_1 - \mathbf{k}_2|$. The maximum possible value of $|\mathbf{q}|$ thus occurs for the **back-scattering geometry** in which the outgoing photon is emitted in the direction back towards the source. In this case, we have:

$$q \approx |\mathbf{k} - (-\mathbf{k})| \approx 2 \frac{n\omega}{c}. \quad (10.32)$$

By inserting typical values into eqn 10.32, we conclude that the maximum value of q that can be accessed in an inelastic light scattering experiment is of order 10^7 m^{-1} . This is very small compared to the size of the Brillouin zone in a typical crystal ($\sim 10^{10} \text{ m}^{-1}$). Inelastic light scattering is thus only able to probe small wave vector phonons.

Raman and Brillouin scattering are generally weak processes, and we therefore expect that the scattering rate will be small. This is because we are dealing with a higher order interaction than for linear interactions such as absorption.

Figure 10.9 shows us that three particles are present in the Feynman diagram for inelastic light scattering rather than the two for absorption (see Fig. 10.1). Therefore, a higher order perturbation term must be involved. This means that we usually have to employ very sensitive detectors to observe the signals even when using a powerful laser beam as the excitation source.

10.5.2 Raman scattering

C.V. Raman was awarded the Nobel prize in 1930 for his discovery of inelastic light scattering from molecules. The process which now carries his name refers to scattering from high frequency excitations such as the vibrational modes of molecules. In the present context of phonon physics, it refers specifically to inelastic light scattering from optical phonons.

Optical phonons are essentially dispersionless near $q = 0$. We argued above that inelastic light scattering can only probe the phonon modes with $q \approx 0$. Therefore, Raman scattering gives little information about the dispersion of optical phonons, and its main use is to determine the frequencies of the LO and TO modes near the Brillouin zone centre. For example, when Raman techniques are used to measure polariton dispersion curves (see Section 10.3, and especially Fig. 10.7), we are only probing a very small portion of the Brillouin zone near $q = 0$.

The complementarity of infrared reflectivity and inelastic light scattering measurements become more apparent when we consider the selection rules for deciding whether a particular optical phonon is Raman active or not. These rules are not the same as those for determining whether the mode is IR active. The full treatment requires the use of group theory. However, a simple rule can be given for crystals that possess inversion symmetry. In these centrosymmetric crystals, the vibrational modes must either have even or odd parity under inversion. The odd parity modes are IR active, while the even parity modes are Raman active. Thus the Raman active modes are not IR active, and *vice versa*. This is called the **rule of mutual exclusion**, and is a well-known result in molecular physics. In non-centrosymmetric crystals, some modes may be simultaneously IR and Raman active.

As an example of these rules, we can compare silicon and GaAs. Silicon has the diamond structure with inversion symmetry, while GaAs has the non-centrosymmetric zinc blende structure. The TO modes of silicon are not IR active, but they are Raman active, while the TO modes of GaAs are both Raman and IR active.

The observation of a Raman spectrum requires specialized apparatus to overcome the difficulties that are inherent to the technique. We pointed out above that the signal is relatively weak, which means that we have to use an intense source such as a laser to produce a sizeable scattering rate. However, the frequency shift of the scattered photons is quite small. We thus need to resolve a weak Raman signal which is very close in wavelength to the elastically scattered light from the laser.

Figure 10.10 shows a basic experimental arrangement that can be used to measure Raman spectra. The sample is excited with a suitable laser, and the scattered light is collected and focussed onto the entrance slit of a scanning spectrometer. The number of photons emitted at a particular wavelength is

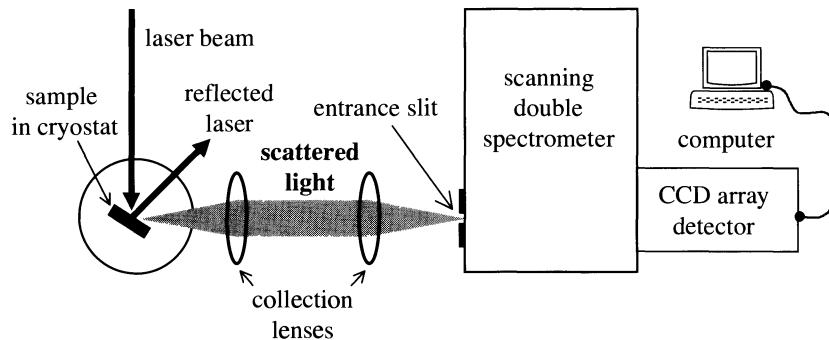


Fig. 10.10 Experimental apparatus used to record Raman spectra. The sample is excited with a laser, and the scattered photons are collected and focussed into a spectrometer. The signals are recorded using a sensitive photon-counting detector such as a photomultiplier tube or a charge coupled device (CCD).

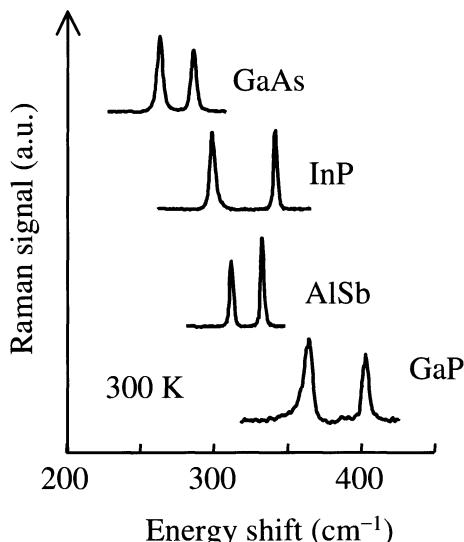


Fig. 10.11 Raman spectra for the TO and LO phonons of GaAs, InP, AlSb and GaP at 300 K using a Nd:YAG laser at $1.06\text{ }\mu\text{m}$. The spectra are plotted against the wave number shift: 1 cm^{-1} is equivalent to an energy shift of 0.124 meV . The LO mode is the one at higher frequency. After [4], copyright 1972 Excerpta Medica Inc., reprinted with permission.

registered using a photon-counting detector and then the results are stored on a computer for analysis. Photomultiplier tubes have traditionally been employed as the detector in this application, but modern arrangements now tend to use array detectors made with charge coupled devices (CCD arrays). By orientating the sample appropriately, the reflected laser light can be arranged to miss the collection optics. However, this still does not prevent a large number of elastically scattered laser photons entering the spectrometer, and this could potentially saturate the detector. To get around this problem, a high resolution spectrometer with good stray light rejection characteristics is used.

Figure 10.11 shows the Raman spectrum obtained from four III–V crystals at 300 K. The laser source was a Nd:YAG laser operating at $1.06\text{ }\mu\text{m}$, and a double monochromator with a photomultiplier tube were used to detect the signal. Two strong lines are observed for each crystal. These correspond to the Stokes-shifted signals from the TO phonons and LO phonons, with the LO phonons at the higher frequency. The values obtained from this data agree very well with those deduced from infrared reflectivity measurements. (See Exercise 10.13.)

One way to achieve good stray light rejection is to use a double spectrometer, which is essentially two spectrometers in tandem. This both increases the spectral resolution and enhances the rejection of unwanted photons.

10.5.3 Brillouin scattering

L. Brillouin gave a theoretical discussion of the scattering of light by acoustic waves in 1922. The technique named after him now refers to inelastic light scattering from acoustic phonons. Its main purpose is to determine the dispersion of these acoustic modes.

The frequency shift of the photons in a Brillouin scattering experiment is given by (see Exercise 10.14):

$$\delta\omega = v_s \frac{2n\omega}{c} \sin \frac{\theta}{2}, \quad (10.33)$$

where ω is the angular frequency of the incident light, n is the refractive index of the crystal, v_s is the velocity of the acoustic waves, and θ is the angle through which the light is scattered. Measurements of $\delta\omega$ therefore allow the velocity of the sound waves to be determined if the refractive index is known.

The experimental techniques used for Brillouin scattering are more sophisticated than those for Raman scattering due to the need to be able to detect much smaller frequency shifts. Single-mode lasers must be used to ensure that the laser linewidth is sufficiently small, and a scanning Fabry–Perot interferometer is used instead of a grating spectrometer to obtain the required frequency resolution.

Example 10.2

When light from an argon ion laser operating at 514.5 nm is scattered by optical phonons in a sample of AlAs, two peaks are observed at 524.2 nm and 525.4 nm. What are the values of the TO phonon and LO phonon energies?

Solution

We can work out the energies of the phonons by using eqn 10.28. The photons have been red-shifted, and thus we are dealing with a Stokes process. For the 524.2 nm line we therefore have:

$$\Omega = \omega_1 - \omega_2 = 2\pi c(1/\lambda_1 - 1/\lambda_2) = 6.8 \times 10^{13} \text{ Hz}.$$

For the 525.4 nm line we find $\Omega = 7.6 \times 10^{13}$ Hz. The higher frequency phonon is the LO mode. Hence we find $\hbar\Omega_{\text{TO}} = 45 \text{ meV}$ and $\hbar\Omega_{\text{LO}} = 50 \text{ meV}$.

10.6 Phonon lifetimes

The discussion of the phonon modes as classical oscillators in Section 10.2 led us to introduce a phenomenological damping constant γ . This damping term is needed to explain why the reflectivity in the restrahlen band is less than unity. Analysis of the experimental data led us to conclude that γ is typically in the range $10^{11}\text{--}10^{12} \text{ s}^{-1}$. This very rapid damping is a consequence of the finite

lifetime τ of the optical phonons. Since γ is equal to τ^{-1} , the data implies that τ is in the range 1–10 ps.

The very short lifetime of the optical phonons is caused by anharmonicity in the crystal. Phonon modes are solutions of the equations of motion with the assumption that the vibrating atoms are bound in a harmonic potential well. In reality, this is only an approximation that is valid for small displacements. In general, the atoms sit in a potential well of the form:

$$U(x) = C_2 x^2 + C_3 x^3 + C_4 x^4 + \dots \quad (10.34)$$

An example of how interatomic interactions lead to a potential of this form is considered in Exercise 10.15.

The term in x^2 in eqn 10.34 is the harmonic term. This leads to simple harmonic oscillator equations of motion with a restoring force $-dU/dx$ proportional to $-x$. The terms in x^3 and higher are the anharmonic terms. These anharmonic terms allow phonon–phonon scattering processes. For example, the term in x^3 allows interactions involving three phonons. Figure 10.12 illustrates two possible permutations for a three-phonon process.

Figure 10.12(a) shows a three-phonon interaction in which one phonon is annihilated and two new phonons are created. This type of anharmonic interaction is responsible for the fast decay of the optical phonons. We can see why this is so by referring to the generic phonon dispersion curve for the first Brillouin zone shown in Fig. 10.13. Lattice absorption or Raman scattering creates optical phonons with $q \approx 0$. Three-phonon processes allow these phonons to decay into two acoustic phonons as indicated in Fig. 10.13. Momentum and energy can be conserved if the two acoustic phonons have opposite wave vectors, and their frequency is half that of the optical phonon. With more complex dispersion relationships, and also the possibility for higher order processes, many other types of decay can contribute to the short lifetime of the optical phonons.

The lifetime of the optical phonons can be deduced from Raman data in two different ways. Firstly, the spectral width of the Raman line is affected by lifetime broadening. Provided that other sources of broadening are smaller, the linewidth in frequency units is expected to be $(2\pi\tau)^{-1}$. Thus measurements of

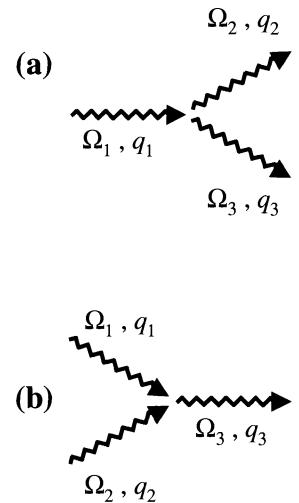


Fig. 10.12 Three phonon interaction processes. Each wiggly arrow represents a phonon. These processes are caused by anharmonicity in the crystal.

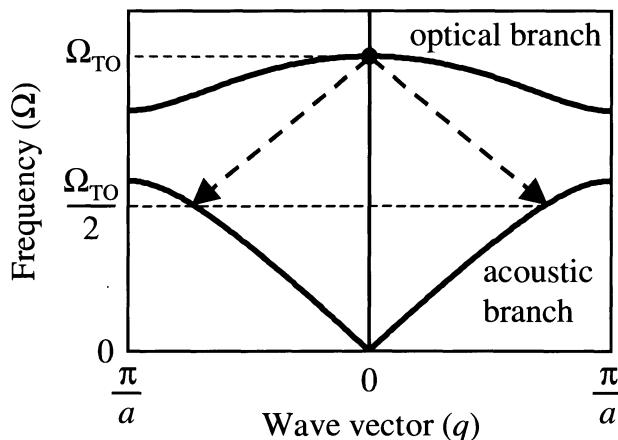


Fig. 10.13 Decay of an optical phonon into two acoustic phonons by a three-phonon interaction of the type shown in Fig. 10.12(a).

the linewidth give a value for τ independently of the reflectivity data. Secondly, τ can be measured directly by time-resolved Raman spectroscopy using short pulse lasers. The lifetime of the LO phonons in GaAs has been determined in this way to be 7 ps at 77 K. This value agrees with the linewidth measured in the conventional Raman spectrum. It is also similar to the lifetime of the TO phonons deduced from reflectivity measurements.

Chapter summary

- The TO phonon modes of polar solids couple strongly to photons when their frequencies and wave vectors match. Acoustic phonons and LO phonons do not couple directly to light waves.
- The interaction between the light and the TO phonon can be modelled by using the classical oscillator model. This model explains why the reflectivity of a polar solid is very high for frequencies in the restrahlen band between ν_{TO} and ν_{LO} .
- The reflectivity in the restrahlen band is 100 % for an undamped system, but damping due to the finite phonon lifetime reduces the reflectivity in real crystals.
- The frequencies of the TO and LO phonon modes are related to each other by the Lyddane–Sachs–Teller relationship given in eqn 10.15.
- The lattice absorbs strongly at the TO phonon frequency. The absorption can be measured directly in thin film samples.
- The strongly coupled phonon–photon waves at frequencies near the restrahlen band are described as polariton modes.
- The electron–phonon coupling in polar crystals leads to polaron effects. Polarons are charge carriers surrounded by a local lattice distortion. The phonon cloud around the electron or hole increases its mass. Polaronic effects are strong in ionic crystals like the alkali halides.
- Raman and Brillouin scattering are inelastic light scattering processes from optical and acoustic phonons respectively. Energy and momentum must be conserved in the scattering process.
- Stokes and anti-Stokes inelastic light scattering processes correspond to phonon emission and absorption respectively. Anti-Stokes scattering from optical phonons is very improbable at low temperatures.
- Optical phonons have short lifetimes due to the possibility of decay into two acoustic phonons by anharmonic interactions.

Further reading

Introductory reading on phonons may be found in practically any solid state physics text, for example: Ashcroft and Mermin (1976), Burns (1985), Ibach and Luth (1995) or Kittel (1996).

The theory of polaritons and polarons is described in more detail in Madelung (1978). Pidgeon (1980) and Seeger (1997) discuss cyclotron resonance experiments in detail. The properties of self-trapped excitons are covered by Song and Williams (1993), while Pope and Swenberg (1999) discuss polaronic hopping transport, especially in organic semiconductors.

A classic text on the infrared physics of molecules and solids is Houghton and Smith (1966). The techniques of inelastic light scattering are described in detail by Mooradian (1972) or Yu and Cardona (1996). The study of phonon dynamics by ultra-fast laser techniques is described by Shah (1999).

References

- [1] Madelung, O. (1996). *Semiconductors, basic data*, (2nd edn). Springer-Verlag, Berlin.
- [2] Hass, M. (1967). In *Semiconductors and Semimetals, Vol. 3: Optical properties of III-V compounds* (ed. R.K. Willardson and A.C. Beer). Academic Press, New York, p. 3.
- [3] Henry, C.H. and Hopfield, J.J. (1965). *Phys. Rev. Lett.*, **15**, 964.
- [4] Mooradian, A. (1972). In *Laser Handbook*, Vol. II (ed. F.T. Arecchi and E.O. Schulz-duBois). North Holland, Amsterdam, p. 1409.
- [5] Turner, W.J. and Reese, W.E. (1962). *Phys. Rev.*, **127**, 126.

Exercises

- (10.1) State, with reasons, which of the following solids would be expected to show strong infrared absorption: (a) ice, (b) germanium, (c) solid argon at 4 K, (d) ZnSe, (e) SiC.
- (10.2) Show that the reflectivity of an undamped polar solid falls to zero at a frequency given by

$$\nu = \left(\frac{\epsilon_{st} - 1}{\epsilon_{\infty} - 1} \right)^{\frac{1}{2}} \nu_{TO},$$

where ϵ_{st} and ϵ_{∞} are the low and high frequency dielectric constants, and ν_{TO} is the frequency of the TO phonon mode at the Brillouin zone centre.

- (10.3) The static and high frequency dielectric constants of LiF are $\epsilon_{st} = 8.9$ and $\epsilon_{\infty} = 1.9$ respectively, and the TO phonon frequency ν_{TO} is 9.2 THz. Calculate the upper and lower wavelengths of the restrahlen band.
- (10.4) Estimate the reflectivity in the middle of the restrahlen band for a crystal with $\nu_{TO} = 10$ THz, $\epsilon_{st} = 12.1$, and $\epsilon_{\infty} = 10$, when the damping constant γ is (a) 10^{11} s^{-1} and (b) 10^{12} s^{-1} .
- (10.5) Figure 10.14 shows the measured infrared reflectivity of AlSb crystals. Use this data to estimate:

- (i) the frequencies of the TO and LO phonons of AlSb near the Brillouin zone centre;
- (ii) the static and high frequency dielectric constants, ϵ_{st} and ϵ_{∞} ;
- (iii) the lifetime of the TO phonons.

Are the experimental values found in parts (i) and (ii) consistent with the Lyddane–Sachs–Teller relationship?

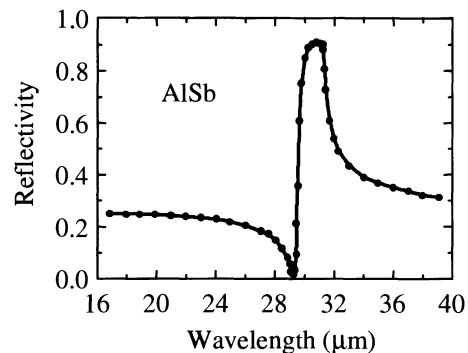


Fig. 10.14 Infrared reflectivity of AlSb. After [5], copyright 1962 American Institute of Physics, reprinted with permission.

- (10.6) Estimate the absorption coefficient at the TO phonon frequency in a typical polar solid with a damping constant γ of (a) 10^{11} s^{-1} and (b) 10^{12} s^{-1} . Take $\nu_{\text{TO}} = 10 \text{ THz}$, $\epsilon_{\text{st}} = 12.1$, and $\epsilon_{\infty} = 10$.
- (10.7) Explain qualitatively why the reflectivity of NaCl in the middle of the restrahlen band is observed to decrease from 98 % at 100 K to 90 % at 300 K.
- (10.8) The static and high frequency dielectric constants of InP are $\epsilon_{\text{st}} = 12.5$ and $\epsilon_{\infty} = 9.6$ respectively, and the TO phonon frequency ν_{TO} is 9.2 THz. Calculate the wave vector of a polariton mode with a frequency of 8 THz. (Ignore phonon damping.)
- (10.9) In an infrared absorption experiment on n-type CdTe, the cyclotron resonance condition is satisfied at 3.4 T for the $306 \mu\text{m}$ line from a deuterated methanol laser. Calculate (i) the polaron mass, and (ii) the rigid lattice electron effective mass, given that $\epsilon_{\infty} = 7.1$, $\epsilon_{\text{st}} = 10.2$, and $\Omega_{\text{LO}} = 31.9 \text{ THz}$.
- (10.10) Discuss the qualitative differences you would expect between the Raman spectrum observed from diamond to that shown for the III–V crystals in Fig. 10.11.
- (10.11) In an inelastic light scattering experiment on silicon using an argon ion laser at 514.5 nm, Raman peaks are observed at 501.2 nm and 528.6 nm. Account for the origin of the two peaks, and estimate their intensity ratios if the sample temperature is 300 K.
- (10.12) NaCl is a centrosymmetric crystal. Would you expect the TO phonon modes to be IR active, or Raman active, or both?
- (10.13) Use the data in Fig. 10.11 to deduce the energies in meV of the TO and LO phonons of GaAs, InP, AlSb, and GaP at 300 K. How do the values for GaAs obtained from this data relate to the infrared reflectivity data given in Fig. 10.5 ?
- (10.14) A photon of angular frequency ω is scattered inelastically through an angle θ by an acoustic phonon of angular frequency Ω . By considering the conservation of momentum in the process, show that Ω is given by:
- $$\Omega = v_s \frac{2n\omega}{c} \sin \frac{\theta}{2},$$
- where v_s and n are the velocity of sound and the refractive index in the medium respectively. (You may assume that $\omega \gg \Omega$.) Hence justify eqn 10.33.
- (10.15)* The potential energy per molecule of an ionic crystal with a nearest neighbour separation of r may be approximated by the following form:
- $$U(r) = \frac{\beta}{r^{12}} - \frac{\alpha e^2}{4\pi\epsilon_0 r},$$
- where α is the Madelung constant of the crystal, and β is a fitting parameter.
- (i) Account for the functional form of $U(r)$.
 - (ii) Show that $U(r)$ has a minimum value when $r = r_0$, where $r_0^{11} = 48\beta\pi\epsilon_0/\alpha e^2$.
 - (iii) Expand $U(r)$ as a Taylor series about r_0 , and hence show that the potential takes the form given by eqn 10.34 for small displacements about r_0 , stating the value of the constant C_3 in terms of α and r_0 .
- (10.16) High resolution Raman experiments on a GaAs crystal indicate that the LO phonon line has a spectral width of 0.85 cm^{-1} . Use this value to estimate the lifetime of the LO phonons, on the assumption that the spectrum is lifetime-broadened.

* Exercises marked with an asterisk are more challenging.

Nonlinear optics

11

Practically everything we have been describing so far in this book falls into the realm of linear optics, where it is assumed that properties such as the refractive index, absorption coefficient and reflectivity are independent of the optical power. This approximation is only valid at low power levels. With a high power laser, it is possible to enter a different realm of behaviour called **nonlinear optics**. In this subject we consider the consequences of allowing the electric susceptibility, and all the properties that follow from it, to vary with the strength of the electric field of the light beam.

Nonlinear optics is a subject in its own right that has grown in importance as applications of lasers have become more common. A solid state optics text such as this would be incomplete without some mention of the types of phenomena that can occur. The objective here is to give a brief introduction to the subject, with a particular emphasis on the relevant aspects of the solid state physics. The treatment we give is predominantly classical, and is mainly based on the dipole oscillator model developed in Chapter 2. We will also make some use of the properties of excitons discussed in Chapter 4 and Section 6.4.4. It is hoped that this may form a basis for further reading in more comprehensive treatments. A partial list of introductory nonlinear optics texts is given under Further Reading.

11.1 The nonlinear susceptibility tensor

The optical properties of materials are described through the real and imaginary parts of the dielectric constant ϵ_r . The dielectric constant is derived from the polarization \mathbf{P} of the medium according to:

$$\begin{aligned}\mathbf{D} &= \epsilon_0 \mathbf{E} + \mathbf{P} \\ &= \epsilon_0 \epsilon_r \mathbf{E}.\end{aligned}\tag{11.1}$$

In linear optics, we assume that \mathbf{P} depends linearly on the electric field \mathbf{E} of the light wave, so that we can write:

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E},\tag{11.2}$$

where χ is the electric susceptibility. By combining eqns 11.1 and 11.2 we derive the usual relationship between ϵ_r and χ , namely:

$$\epsilon_r = 1 + \chi.\tag{11.3}$$

In nonlinear optics we consider the possibility that the relationship between \mathbf{P} and \mathbf{E} is more general than that given by eqn 11.2. We start by considering a

11.1	The nonlinear susceptibility tensor	227
11.2	The physical origin of optical nonlinearities	230
11.3	Second-order nonlinearities	236
11.4	Third-order nonlinear effects	243

nonlinear medium in which the polarization is parallel to the electric field, so that we do not need to consider the vector nature of \mathbf{P} and \mathbf{E} at this stage. We split the polarization P into the first-order linear response $P^{(1)}$, plus a whole series of nonlinear terms of increasing order according to:

$$P^{\text{nonlinear}} = P^{(1)} + P^{(2)} + P^{(3)} + \dots, \quad (11.4)$$

where $P^{(n)}$ is the n th-order nonlinear polarization.

In analogy with eqn 11.2, we now introduce the **nonlinear susceptibility** $\chi^{\text{nonlinear}}$, and the n th-order nonlinear susceptibility, $\chi^{(n)}$. These are defined by the following equations:

$$\begin{aligned} P^{\text{nonlinear}} &= \epsilon_0 \chi^{\text{nonlinear}} \mathcal{E} \\ &= \epsilon_0 \left(\chi^{(1)} \mathcal{E} + \chi^{(2)} \mathcal{E}^2 + \chi^{(3)} \mathcal{E}^3 + \dots \right), \end{aligned} \quad (11.5)$$

where \mathcal{E} is the magnitude of the applied field.

The various terms in eqns 11.4 and 11.5 correspond directly with each other so that

$$P^{(1)} = \epsilon_0 \chi^{(1)} \mathcal{E} \quad (11.6)$$

$$P^{(2)} = \epsilon_0 \chi^{(2)} \mathcal{E}^2 \quad (11.7)$$

$$P^{(3)} = \epsilon_0 \chi^{(3)} \mathcal{E}^3 \quad (11.8)$$

⋮

By comparing eqns 11.3 and 11.5, we see that:

$$\epsilon_r^{\text{nonlinear}} = 1 + \chi^{\text{nonlinear}} \quad (11.9)$$

$$= 1 + \chi^{(1)} + \chi^{(2)} \mathcal{E} + \chi^{(3)} \mathcal{E}^2 + \dots \quad (11.10)$$

where $\chi^{(1)}$ is just the normal linear susceptibility. Equation 11.10 implies that the dielectric constant depends on the electric field through the nonlinear susceptibilities. Since the optical power is proportional to \mathcal{E}^2 , this means that ϵ_r also depends on the optical power. Hence properties like the refractive index and absorption coefficient become power-dependent in nonlinear materials.

The different order nonlinear susceptibilities give rise to a whole host of nonlinear effects. The majority of these phenomena can be attributed to either the $\chi^{(2)}$ or $\chi^{(3)}$ terms in the polarization. These are either called **second-order** or **third-order nonlinear** effects as appropriate. Some of these will be discussed in Sections 11.3 and 11.4.

The well-defined axes of crystalline materials make it necessary to consider that the nonlinear response of the medium may depend on the directions in which the fields are applied. For example, we could apply two optical fields in different directions and then generate a nonlinear polarization along a third direction. This type of behaviour can be described by generalizing eqns 11.7 and 11.8 to allow for the anisotropic response of the medium. For example, the components of the second-order nonlinear polarization $\mathbf{P}^{(2)}$ can be written in the following form:

$$P_i^{(2)} = \epsilon_0 \sum_{j,k=x,y,z} \chi_{ijk}^{(2)} \mathcal{E}_j \mathcal{E}_k. \quad (11.11)$$

It is not necessary that \mathcal{E}_k and \mathcal{E}_l in eqn 11.11 should be derived from different light beams. In many cases there will only be a single light beam incident on the crystal, and \mathcal{E}_k and \mathcal{E}_l will just be the components of the polarization vector of the light resolved along the appropriate axes.

The quantity $\chi_{ijk}^{(2)}$ that appears here is the second-order nonlinear susceptibility tensor, and the subscripts i , j and k correspond to the cartesian coordinate axes x , y and z . It will usually be convenient to define these axes so that they coincide with the axes of the crystal whenever this is possible.

Equation 11.11 shows that there are nine different contributions for each component of $\mathbf{P}^{(2)}$. For example, the term with $\chi_{xyz}^{(2)}$ gives the nonlinear polarization generated along the x axis when one optical field is applied along the y axis and another along the z axis. We can also generate a nonlinear polarization along the x axis by applying two fields along x using the $\chi_{xxx}^{(2)}$ term, and so on for all nine possible permutations of j and k . At first sight it might therefore appear that we have to measure 27 different quantities in order to fully quantify the second-order nonlinear response of an anisotropic medium. Fortunately, this is not usually the case, because the high degree of symmetry found in crystals requires that many of the terms are zero, and many of the others are the same. This point is developed further in Section 11.3.2.

The third-order nonlinear response of an anisotropic medium is also described by a tensor relationship. We can generalize eqn 11.8 by writing the components of the third-order nonlinear polarization as follows:

$$P_i^{(3)} = \epsilon_0 \sum_{j,k,l=x,y,z} \chi_{ijkl}^{(3)} \mathcal{E}_j \mathcal{E}_k \mathcal{E}_l, \quad (11.12)$$

where $\chi_{ijkl}^{(3)}$ is the third-order nonlinear susceptibility tensor. $\chi_{ijkl}^{(3)}$ is a fourth-rank tensor with 81 components. As with the second-order nonlinear susceptibility, symmetry may require that many of these terms are the same or zero.

Example 11.1

Potassium dihydrogen phosphate (KDP) is a uniaxial crystal with a four-fold axis of rotation about the z axis. The tetragonal 42m symmetry class of the crystal demands that the only non-zero components of the second-order nonlinear susceptibility tensor are the ones with i , j and k all different, namely $\chi_{xyz}^{(2)}$, $\chi_{yxz}^{(2)}$, $\chi_{xzy}^{(2)}$, $\chi_{zxy}^{(2)}$, $\chi_{yzx}^{(2)}$, and $\chi_{zyx}^{(2)}$. Furthermore, symmetry also requires that

$$\chi_{xyz}^{(2)} = \chi_{yxz}^{(2)} = \chi_{xzy}^{(2)} = \chi_{zxy}^{(2)}, \quad (11.13)$$

and

$$\chi_{zxy}^{(2)} = \chi_{zyx}^{(2)}. \quad (11.14)$$

Determine the direction of the nonlinear polarization when a powerful laser beam is propagating along the optic axis.

Solution

If the laser is propagating in the z direction, then the electric field of the light will be polarized along the x or y directions. The nonlinear polarization is therefore given by eqn 11.11 with $\mathcal{E}_z = 0$. This gives:

$$P_i^{(2)} = \epsilon_0 \left(\chi_{ixx}^{(2)} \mathcal{E}_x \mathcal{E}_x + \chi_{ixy}^{(2)} \mathcal{E}_x \mathcal{E}_y + \chi_{iyx}^{(2)} \mathcal{E}_y \mathcal{E}_x + \chi_{iyy}^{(2)} \mathcal{E}_y \mathcal{E}_y \right). \quad (11.15)$$

If $i = x$ or $i = y$, then all the terms on the right hand side are zero because $\chi_{ijk}^{(2)}$ is zero unless i , j and k are all different. This means that the nonlinear polarization vector is given by:

$$\begin{aligned} P_x^{(2)} &= 0 \\ P_y^{(2)} &= 0 \\ P_z^{(2)} &= \epsilon_0 \left(\chi_{zxy}^{(2)} \mathcal{E}_x \mathcal{E}_y + \chi_{zyx}^{(2)} \mathcal{E}_y \mathcal{E}_x \right). \end{aligned} \quad (11.16)$$

We therefore conclude that the nonlinear polarization is pointing along the optic axis, irrespective of the direction of the polarization of the input laser beam.

11.2 The physical origin of optical nonlinearities

The discussion in the previous section gave no indication as to why a particular material should be nonlinear or not. The magnitude of the electric field that binds an electron to an atom is typically around 10^{10} – 10^{11} V m $^{-1}$. (See Exercise 11.1.) It might therefore be expected that nonlinear effects will become important when the electric field of the light is comparable to this value. From the relationship between the intensity of a light beam and its electric field given by eqn A.40 in Appendix A, namely:

$$I = \frac{1}{2} c \epsilon_0 n \mathcal{E}^2, \quad (11.17)$$

we see that we need optical intensities around 10^{19} W m $^{-2}$ to produce fields of this magnitude. Intensities as high as this can just about be achieved with very powerful lasers, but in fact the nonlinear effects set in at much lower intensity levels. This is because we can produce a sizeable macroscopic result by adding together the very small nonlinear effects in a very large number of atoms. This only works if the nonlinear phenomena in all the individual atoms are in phase with each other. This effect is called ‘phase-matching’, and is discussed in Section 11.3.3.

The approach taken to explaining the microscopic origin of optical nonlinearities depends on whether the frequency is close to one of the natural transition frequencies of the atoms or not. If it is, then we are dealing with a **resonant** nonlinear effect, while if it is not, we are considering a **non-resonant** nonlinearity. These two situations are discussed separately below, starting with the non-resonant nonlinearities. It turns out that the non-resonant effects can be explained in terms of the classical oscillator model by introducing anharmonic terms. On the other hand, we need to use a quantum model to account properly for resonant effects.

11.2.1 Non-resonant nonlinearities

In Chapter 2 we explained how we can calculate the response of a medium to electromagnetic waves by assuming that it consists of a series of oscillators with characteristic resonant frequencies. In the near-infrared, visible or

ultraviolet spectral ranges we will normally be considering the response due to the electrons. We have been assuming that these are bound to the atoms by harmonic restoring forces such that the displacement induced by the driving field of the light wave is linear. As with most oscillatory systems, this will only be true for small displacements. If the system is driven hard by the strong field of an intense laser beam, the displacements will be large, and it may no longer be valid to assume that the displacement varies linearly with the driving field.

We can account for the non-resonant nonlinear effects by assuming that the electron is bound in an anharmonic potential well of the form:

$$U(x) = \frac{1}{2}m_0\omega_0^2x^2 + \frac{1}{3}m_0C_3x^3 + \frac{1}{4}m_0C_4x^4 + \dots, \quad (11.18)$$

where ω_0 is the natural resonant frequency and $x = 0$ corresponds to the equilibrium position of the electron. It is assumed that $\omega_0^2 \gg C_3x \gg C_4x^2 \dots$, so that it makes sense to carry out the power series expansion, and that the harmonic term dominates for small displacements. The power series expansion is a simplification of the more complicated functional forms that would appear in a real atom. (See Exercise 10.15 for a worked example for the vibrational potential energy.)

We concentrate here on the second-order effects, and consider only the x^3 term in eqn 11.18. The restoring force for displacements from the equilibrium position is given by:

$$F(x) = -\frac{dU}{dx} = -\left(m_0\omega_0^2x + m_0C_3x^2\right). \quad (11.19)$$

This shows that the strength of the restoring force now depends on the direction of the displacement: the electron experiences a stronger force for positive displacements than for negative displacements. If we drive the electron with the AC electric field of a light wave, the displacements will be smaller during the positive part of the cycle than for the negative part. Since the dipole moment per unit volume of the medium is equal to $-Nex$, the polarization will likewise be asymmetric in the field direction. Hence the relationship between P and E will not be linear and will involve powers of E greater than one.

The nonlinear relationship between P and E is sketched in Fig. 11.1. For small fields the departure from the linear response shown by the dashed line is negligible. Hence the polarization closely follows the applied field, as shown in Fig. 11.1(a). However, if the magnitude of the applied field is increased, the response becomes asymmetric, with larger displacements for negative fields. This point is illustrated in Fig. 11.1(b), which shows how the application of a sinusoidal electric field gives a distorted output if the material is nonlinear. It is well known from electrical circuit theory that the distorted output can be described by including higher harmonics. In the case shown in Fig. 11.1(b), the output contains a second harmonic wave with 20 % of the amplitude of the fundamental.

This simple discussion shows that the inclusion of the anharmonic term generates a signal at twice the frequency as the applied wave. We can see from eqn 11.7 that this is equivalent to a second-order nonlinearity, because

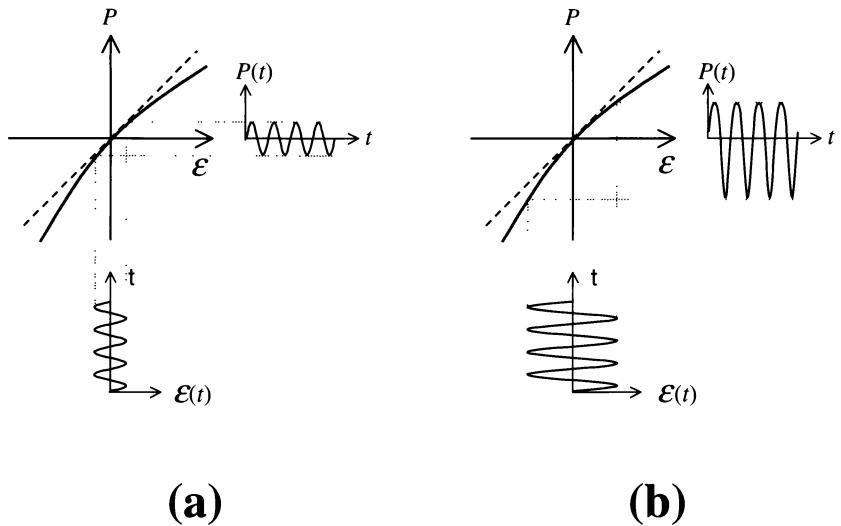


Fig. 11.1 Response of an anharmonic medium to a sinusoidal driving field. The linear dependence between P and \mathcal{E} is shown by the dashed line, while the solid line corresponds to a nonlinear dependence. (a) Small electric field: the departure from linear response is small. (b) Large applied field: the polarization is asymmetric, with larger displacements for negative \mathcal{E} .

if $\mathcal{E}(t) = \mathcal{E}_0 \sin \omega t$, then

$$\begin{aligned} P^{(2)}(t) &= \epsilon_0 \chi^{(2)} \mathcal{E}_0^2 \sin^2 \omega t \\ &= \frac{1}{2} \epsilon_0 \chi^{(2)} \mathcal{E}_0^2 (1 - \cos 2\omega t). \end{aligned} \quad (11.20)$$

Thus if $\chi^{(2)}$ is non-zero, the medium generates a wave at 2ω when driven at frequency ω . This is the same conclusion as that derived from the consideration of the anharmonic term in the potential, and shows that the two treatments are equivalent.

The relationship between C_3 and $\chi^{(2)}$ can be made more precise by finding an approximate solution to the equation of motion of the electron when driven by an AC electric field at frequency ω . To do this, we proceed as in Section 2.2, but now include the anharmonic term in the restoring force. The equation of motion is thus:

$$m_0 \frac{d^2x}{dt^2} + m_0 \gamma \frac{dx}{dt} + m_0 \omega_0^2 x + m_0 C_3 x^2 = -e\mathcal{E}, \quad (11.21)$$

where γ is the usual damping term. \mathcal{E} is the driving field of the electromagnetic wave, which is assumed to have the following time dependence:

$$\mathcal{E}(t) = \mathcal{E}_0 \cos \omega t = \frac{1}{2} \mathcal{E}_0 (e^{i\omega t} + e^{-i\omega t}). \quad (11.22)$$

We have seen above that the inclusion of the C_3 term leads to a response at frequency 2ω in addition to the one at ω . Therefore we write the time-dependence of the electron displacement as:

$$x(t) = \frac{1}{2} (X_1 e^{i\omega t} + X_2 e^{2i\omega t} + \text{c.c.}) \quad (11.23)$$

where ‘c.c.’ stands for complex conjugate. We assume that the nonlinear term is small, so that $X_1 \gg X_2$.

It is important to keep track of all the conjugate terms in this model, for otherwise we can lose some of the important cross-terms.

On substituting eqn 11.23 into eqn 11.21, we obtain:

$$\begin{aligned} & (-\omega^2 + i\omega\gamma + \omega_0^2) \frac{1}{2}(X_1 e^{i\omega t} + \text{c.c.}) \\ & + (-4\omega^2 + 2i\omega\gamma + \omega_0^2) \frac{1}{2}(X_2 e^{2i\omega t} + \text{c.c.}) \\ & + \frac{C_3}{4}(X_1^2 e^{2i\omega t} + 2X_1^* X_2 e^{i\omega t} + \dots + \text{c.c.}) \\ & = \frac{-e\mathcal{E}_0}{m_0} \frac{1}{2}(e^{i\omega t} + \text{c.c.}) \end{aligned} \quad (11.24)$$

where it is assumed that the anharmonic term is small and the ellipsis represents the higher order cross-terms at frequencies other than ω and 2ω . For eqn 11.24 to hold at all times, the coefficients of $e^{\pm i\omega t}$ and $e^{\pm 2i\omega t}$ must be the same on both sides of the equation. We are assuming that the nonlinear response is small, and so we can neglect the term at frequency ω generated from the anharmonic part of the potential. Hence we obtain:

$$X_1 = \frac{-e\mathcal{E}_0}{m_0} \frac{1}{(\omega_0^2 - \omega^2) + i\gamma\omega}. \quad (11.25)$$

This is exactly the same result as eqn 2.9 in Section 2.2, which is hardly surprising, since it represents the linear response of the system. The polarization at frequency ω follows directly:

$$\begin{aligned} P(\omega, t) &= -Nex(\omega, t) \\ &= -Ne \frac{1}{2}(X_1 e^{i\omega t} + \text{c.c.}) \\ &= \epsilon_0 \chi(\omega) \mathcal{E}(t), \end{aligned} \quad (11.26)$$

where the third line is just the standard definition of the linear susceptibility as in eqn 11.2. By combining eqns 11.22, 11.25 and 11.26, we obtain the usual result for the linear susceptibility:

$$\chi(\omega) = \frac{Ne^2}{m_0 \epsilon_0 [(\omega_0^2 - \omega^2) + i\gamma\omega]}. \quad (11.27)$$

We now solve for X_2 to find the nonlinear response by equating the coefficients of $e^{2i\omega t}$ in eqn 11.24. This gives:

$$(-4\omega^2 + 2i\omega\gamma + \omega_0^2) \frac{X_2}{2} + \frac{C_3}{4} X_1^2 = 0, \quad (11.28)$$

from which we obtain, using eqns 11.25 and 11.27:

$$\begin{aligned} X_2 &= \frac{-C_3 X_1^2}{2(\omega_0^2 - 4\omega^2 + 2i\omega\gamma)} \\ &= \frac{-C_3 e^2 \mathcal{E}_0^2}{2m_0^2 (\omega_0^2 - \omega^2 + i\gamma\omega)^2 (\omega_0^2 - 4\omega^2 + 2i\omega\gamma)} \\ &= \frac{-m_0 C_3 \epsilon_0^3 \chi(\omega)^2 \chi(2\omega)}{2N^3 e^4} \mathcal{E}_0^2. \end{aligned} \quad (11.29)$$

The polarization at frequency 2ω is given by

$$P(2\omega, t) = -Nex(2\omega, t) = -Ne \frac{1}{2}(X_2 e^{2i\omega t} + \text{c.c.}). \quad (11.30)$$

Now in the case we are considering where the polarization at frequency 2ω is generated by nonlinear conversion of the driving field at frequency ω , $P(2\omega)$ will also be given by eqns 11.7 and 11.22 as:

$$P(2\omega, t) = \epsilon_0 \chi^{(2)} \mathcal{E}(t)^2 = \epsilon_0 \chi^{(2)} (\frac{1}{2} \mathcal{E}_0)^2 (e^{2i\omega t} + c.c.) . \quad (11.31)$$

We have restricted our attention to the second-order nonlinearity here, but it is obvious that the derivation can be generalized by including higher order anharmonic terms, which would then explain the origin of higher order nonlinearities.

Thus by combining eqns 11.29–11.31 we obtain the final result:

$$\chi^{(2)} = \frac{m_0 C_3 \chi(\omega)^2 \chi(2\omega) \epsilon_0^2}{N^2 e^3} . \quad (11.32)$$

This shows that the second-order nonlinear susceptibility is directly proportional to C_3 , the anharmonic term in the equation of motion. Equation 11.32 is reasonably successful in predicting the dispersion of $\chi^{(2)}$ in a large number of crystals. This is because it is found empirically that the anharmonic constant C_3 does not vary very much from material to material.

Equation 11.32 tells us that $\chi^{(2)}$ increases as ω approaches ω_0 , through the frequency dependence of $\chi(\omega)$ given by eqn 11.27. This effect is known as resonance enhancement. In a lightly damped system, the classical treatment breaks down as we get closer to the resonant frequency due to the divergence in $\chi(\omega)$. Hence we have to adopt a different approach if we are close to resonance. This is discussed in the next section.

11.2.2 Resonant nonlinearities

The off-resonant nonlinear effects discussed in the previous section are all ‘virtual’ processes. This means that no real transitions take place because the photon energy does not coincide with any of the transition frequencies of the atoms. The situation is obviously completely different if the laser frequency is in resonance with an atomic transition. In this case, the atoms can absorb photons and make transitions to excited states as the beam propagates through the medium.

The absorption rate is normally determined by the matrix element for the transition and the density of states according to Fermi’s golden rule (eqn B.13). This allows us to determine the absorption coefficient for a particular material at a particular frequency. All this presupposes that the intensity of the light beam on the sample is small. If the intensity is high, we will find that the absorption coefficient becomes intensity dependent. Since the absorption coefficient is related to the dielectric constant, this means that the dielectric constant is intensity dependent. In other words, we are dealing with an optical nonlinearity.

The intensity dependence of the absorption rate can be understood through the Einstein B coefficients discussed in Section B.1 of Appendix B. Consider the propagation of an intense laser beam of frequency ν through an absorbing medium. Figure 11.2 illustrates the simplest case to discuss, namely a medium containing atoms with just two levels: level 1 at energy E_1 and level 2 at energy E_2 , where $E_2 > E_1$. We assume that there are N_1 atoms per unit volume in the lower level, and N_2 per unit volume in level 2. The total number of atoms per unit volume is $N_0 = N_1 + N_2$.

We consider the case in which the laser is resonant with the atomic transition frequency of the atoms, such that $h\nu = E_2 - E_1$. The laser beam will be

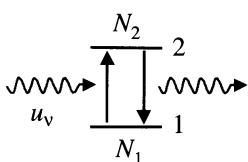


Fig. 11.2 Transitions induced by a resonant laser beam of energy density u_v . Photons are removed from the beam by absorption transitions from level 1 \rightarrow 2 and are added by stimulated emission transitions from level 2 \rightarrow 1.

absorbed as it propagates through the medium by transitions upwards from level 1 to level 2. At the same time, photons will be added to the beam by stimulated emission. The stimulated emission rate is not normally considered when discussing the propagation through an absorbing medium because it is assumed that N_2 is negligible. If the laser intensity is large, we can no longer make this assumption, because the absorption transitions will create a substantial population in the upper level. This gives rise to a significant stimulated emission rate, and effectively reduces the absorption coefficient.

The reduction of the absorption coefficient due to stimulated emission can be modelled by considering an incremental beam slice of thickness dz as illustrated in Fig. 11.3. The number of photons absorbed per unit time in the incremental slice is given by:

$$\delta N_{\text{absorbed}} = B_{12}N_1 u_v \times Adz, \quad (11.33)$$

where u_v is the energy density of the beam at position z , and A is its area. Adz is thus the volume of the slice. In the same way, we can see from eqn B.6 that the number of photons added per unit time to the beam by stimulated emission is given by:

$$\delta N_{\text{stimulated}} = B_{21}N_2 u_v \times Adz. \quad (11.34)$$

Hence the total reduction in the photon number per unit time is given by:

$$\begin{aligned} \delta N_{\text{total}} &= \delta N_{\text{absorbed}} - \delta N_{\text{stimulated}} \\ &= (B_{12}N_1 - B_{21}N_2)u_v Adz. \end{aligned} \quad (11.35)$$

This is effectively the net absorption rate.

The intensity of the beam decreases as it propagates through the medium due to absorption. If we take the intensity at position z to be $I(z)$, and the reduction in the intensity in the incremental slice to be dI , we can write

$$\begin{aligned} AdI &= \delta N_{\text{total}} \times hv \\ &= -(B_{12}N_1 - B_{21}N_2)u_v hv Adz. \end{aligned} \quad (11.36)$$

The left hand side is the energy absorbed per unit time from the beam in the incremental slice at z . The right hand side is the change in the photon number per unit time given by eqn 11.35 multiplied by the energy of each photon. Conservation of energy requires that these two quantities must be the same.

Equation 11.36 can be simplified by noting from eqn A.35 in Appendix A that $I = cu_v/n$, where n is the refractive index of the medium. Hence we obtain:

$$\frac{dI}{dz} = -\frac{B_{12}(N_1 - N_2)hv n}{c} I, \quad (11.37)$$

where we have assumed that the degeneracies of the two levels are the same so that $B_{12} = B_{21}$ (cf. eqn B.10). Equation 11.37 can be compared to the standard definition of the absorption coefficient α given in eqn 1.4, which implies that

$$\frac{dI}{dz} = -\alpha I. \quad (11.38)$$

Hence by comparing eqns 11.37 and 11.38 we see that

$$\alpha = B_{12}(N_1 - N_2)hv n/c. \quad (11.39)$$

Equation 11.33 follows from eqn B.5, with the appropriate redefinition of the Einstein B coefficient to the case of a monochromatic light source with total energy density u_v instead of a continuous spectrum with spectral energy density $u(v)$.

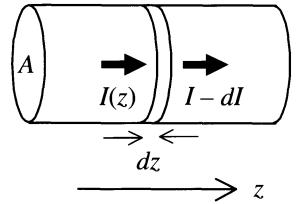


Fig. 11.3 Propagation of a laser beam of area A through an absorbing medium.

The argument can easily be generalized to the case where the degeneracies are different.

This shows that the absorption coefficient is proportional to the population difference between the lower and upper levels.

At low intensities we can assume that $N_1 \approx N_0$ and $N_2 \approx 0$. Equation 11.39 then just reduces to the usual result where the absorption coefficient is proportional to the number of atoms in the system. However, at high intensities, the laser pumps a large number of atoms into the upper level so that N_2 increases and N_1 decreases. Hence the absorption coefficient begins to decrease as the population difference between the upper and lower levels decreases.

The decrease of the absorption with the laser power can be characterized by introducing the **saturation intensity** I_s . The absorption coefficient is found experimentally to depend on the intensity I according to the following relationship:

$$\alpha(I) = \frac{\alpha_0}{1 + I/I_s}, \quad (11.40)$$

where α_0 is the absorption measured in the linear regime when $I \ll I_s$. A medium which shows the behaviour indicated by eqn 11.40 is called a **saturable absorber**. The saturation intensity measured for a particular absorption line will depend on the detailed rate constants for the transitions of the atoms.

At low intensity levels, eqn 11.40 can be expanded to obtain:

$$\alpha(I) = \alpha_0 - (\alpha_0/I_s)I. \quad (11.41)$$

This shows that the absorption decreases linearly with I . Now α is proportional to the imaginary part of ϵ_r (cf. eqns 1.16 and 1.21), and I is proportional to \mathcal{E}^2 . Hence ϵ_r varies in proportion to \mathcal{E}^2 , and from eqn 11.10 we see that this is equivalent to a $\chi^{(3)}$ process. In other words, the resonant nonlinearities due to saturable absorption are third-order nonlinear effects. (See Exercise 11.9).

The analysis of the saturable absorber above applies primarily to the discrete absorption lines found in atomic systems. However, in solid state materials we will often be more interested in the saturation of an absorption band rather than a discrete line. For example, in Section 11.4.3 we will present data for the saturable absorption of interband transitions and also of excitons.

In treating the nonlinear saturation of interband absorption, it is useful to take a slightly different approach which is based on the Pauli exclusion principle. The dependence of α on $(N_1 - N_2)$ in eqn 11.39 can be considered as a consequence of the Fermi–Dirac statistics of the electrons. For absorption to be possible, the lower level must contain an electron, while the upper level must be empty. Hence the absorption coefficient will obey

$$\alpha = \alpha_0(f_1 - f_2), \quad (11.42)$$

where f_1 and f_2 are the Fermi occupancies of the lower and upper levels respectively. α_0 is the low power absorption when the lower level is full and the upper level empty: that is, when $f_1 = 1$ and $f_2 = 0$. The absorption at high powers is calculated by working out the filling of the levels after a large number of electrons and holes have been excited by absorption of a laser pulse.

11.3 Second-order nonlinearities

In this section we will discuss a few of the more important effects that are associated with the second-order nonlinear susceptibility $\chi^{(2)}$. We begin by

The approach based on the Pauli principle is essentially the same as that based on the Einstein B coefficients, but it is easier to apply in practice.

discussing the general principles of nonlinear frequency mixing, and then discuss the effect of the crystal symmetry on the nonlinear coefficients. Finally, we introduce the concept of phase-matching which is crucially important for obtaining large nonlinear signals.

11.3.1 Nonlinear frequency mixing

The second-order nonlinear polarization is given by eqn 11.7. If the medium is excited by sinusoidal waves at frequencies ω_1 and ω_2 with amplitudes \mathcal{E}_1 and \mathcal{E}_2 respectively, then the nonlinear polarization will be equal to:

$$\begin{aligned} P^{(2)}(t) &= \epsilon_0 \chi^{(2)} \times \mathcal{E}_1 \cos \omega_1 t \times \mathcal{E}_2 \cos \omega_2 t \\ &= \epsilon_0 \chi^{(2)} \mathcal{E}_1 \mathcal{E}_2 \frac{1}{2} [\cos(\omega_1 + \omega_2)t + \cos(\omega_1 - \omega_2)t]. \end{aligned} \quad (11.43)$$

This shows that the second-order nonlinear response generates polarization waves at the sum and difference frequencies of the input fields according to:

$$\omega_{\text{sum}} = \omega_1 + \omega_2 \quad (11.44)$$

$$\omega_{\text{diff}} = \omega_1 - \omega_2 \quad (11.45)$$

The medium then re-radiates at ω_{sum} and ω_{diff} , thereby emitting light at frequencies $(\omega_1 + \omega_2)$ and $(\omega_1 - \omega_2)$. This effect is called **nonlinear frequency mixing**. If the frequencies are the same, the sum frequency is at twice the input frequency. This effect is called **frequency doubling** or **second harmonic generation**, and has already been introduced in the discussion of eqn 11.20.

Nonlinear frequency mixing processes can be represented by Feynman diagrams as indicated in Fig. 11.4. Figure 11.4(a) shows the process for sum frequency mixing, while Fig. 11.4(b) represents difference frequency mixing. As usual, conservation of energy applies at each vertex. The negative input frequency at ω_2 for the difference frequency mixing process in Fig. 11.4(b) reflects the fact that $\cos \omega t = \frac{1}{2}(e^{+i\omega t} + e^{-i\omega t})$, so that we can represent real waves either with positive or negative frequencies on a Feynman diagram. In quantum mechanical terms, we would say that the sum frequency mixing process annihilates two input photons at frequencies ω_1 and ω_2 , with the creation of a new photon at frequency ω_{sum} , while difference frequency mixing annihilates one photon at frequency ω_1 and creates two photons, one at frequency ω_2 and the other at ω_{diff} . The creation of the photon at frequency ω_2 in the latter case is stimulated by the presence of a large number of existing photons at frequency ω_2 from the input field.

One of the most important uses for nonlinear optical processes is to generate new frequencies from fixed-wavelength lasers. The most common technique is frequency doubling. In this case, we just have a single input beam, and the sum frequency mixing works by taking two photons from the input beam and generating a new photon at the doubled frequency.

Figure 11.5 shows a schematic experimental arrangement that can be used to generate the second, third and fourth harmonics of a Nd:YAG laser operating at 1064 nm. The second harmonic at $1064/2 = 532$ nm is generated by frequency doubling of the fundamental. The second harmonic beam can then be doubled again using an additional nonlinear crystal to generate the fourth harmonic at 266 nm. Alternatively, we can generate the third harmonic by carrying out

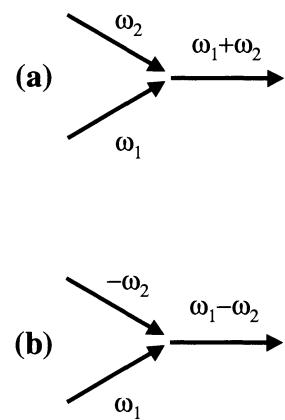


Fig. 11.4 Feynman diagrams for second-order nonlinear frequency mixing processes.

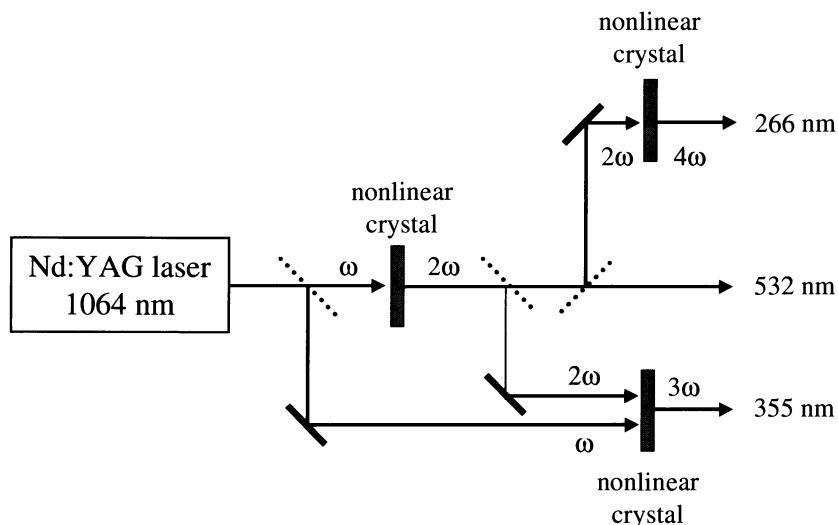


Fig. 11.5 Nonlinear frequency conversion of a Nd:YAG laser operating at 1064 nm. The beam is first doubled to 532 nm. This beam can then either be used as the output, or doubled again to 266 nm. Alternatively, the output at 532 nm can be mixed with the fundamental at 1064 nm to generate the third harmonic at 355 nm. The residual pump beams transmitted through the nonlinear crystals are separated from the harmonics by suitable filters that are not shown in the diagram.

sum frequency mixing in another crystal using the fundamental and the second harmonic beams. These techniques are standard procedures in modern laser physics.

When the two input fields are at the same frequency, we see from eqn 11.45 that the difference frequency is zero. This effect is called **optical rectification**, and refers to the phenomenon by which a static electric field is produced from fields of optical frequencies. The **Pockels effect**, which is also known as the **linear electro-optic effect**, is the reverse of this process. In the Pockels effect we apply a DC electric field \mathcal{E}_{DC} and produce a change in the dielectric constant at optical frequencies which is linear in \mathcal{E}_{DC} . This effect modifies the refractive index of the medium (see Exercise 11.7) and can be used for making optical phase modulators or for inducing a controlled amount of birefringence.

The sum frequency mixing process shown in Fig. 11.4(a) can work the other way round as well. In this case, we bring in a single input field at frequency ω and create two new photons at frequencies of ω_1 and ω_2 , where $\omega_1 + \omega_2 = \omega$. This process is called **down conversion**. It is apparent that the output frequencies generated by down-conversion are not uniquely defined. Any combination of frequencies that satisfies the conservation of energy requirements can in principle be generated. However, the number of photons emitted at any particular frequency will only be large if the phase-matching conditions discussed in Section 11.3.3 below are satisfied.

Down-conversion can be used to amplify a weak beam by a process called **parametric amplification**. If we introduce a weak ‘signal’ field at frequency ω_s in the presence of a strong pump field at frequency ω , it can generate an ‘idler’ field at $\omega_i = \omega - \omega_s$ by difference frequency mixing with the pump field. These new idler photons then generate more signal photons by further mixing with the pump field. This process then repeats itself. If the phase-matching conditions are satisfied, it is possible to transfer power from the pump beam to the signal and idler beams. Furthermore, if the crystal is inside an optical cavity which is resonant with either ω_s or ω_i , then oscillation can occur. This process is called **parametric oscillation**, and can lead to the generation of intense beams

A very interesting aspect of down-conversion processes is that the photons are always created in pairs. This means that the photon statistics at frequency ω_1 are directly correlated with those at ω_2 . This gives rise to a whole host of beautiful quantum optical effects. For example, the pairs of photons are ideal for use in Einstein–Podolsky–Rosen experiments to test for quantum non-locality. Furthermore, the fluctuations in the photon number for the pair can be smaller than the usual Poissonian statistics of laser light.

Table 11.1 Second-order nonlinear effects. The third column lists the frequencies of the light beams incident on the nonlinear crystal, while the fourth gives the frequency of the output beam or the nonlinear polarization. A frequency of zero indicates a DC electric field.

Effect	Alternative name	Input frequencies	Output frequencies
Frequency doubling	Second harmonic generation	ω	2ω
Optical rectification		ω	0
Down conversion		ω	ω_1, ω_2
Sum frequency mixing		ω_1, ω_2	$(\omega_1 + \omega_2)$
Difference frequency mixing		ω_1, ω_2	$ \omega_1 - \omega_2 $
Pockels effect	Linear electro-optic effect	$\omega, 0$	ω

at tunable frequencies even though we started from a fixed-wavelength laser.

The different types of second-order nonlinear effects considered in this section are summarized in Table 11.1.

11.3.2 Crystal symmetry

The second-order nonlinear susceptibility introduced in eqn 11.11 is a third rank tensor with 27 components. Fortunately, it is not necessary to measure all of these components to determine the nonlinear response of the medium. It is immediately obvious that some of the 27 components are the same. For example the term $\chi_{xyz}^{(2)}\mathcal{E}_y\mathcal{E}_z$ must be the same as $\chi_{xzy}^{(2)}\mathcal{E}_z\mathcal{E}_y$ because the response of the medium cannot depend on the mathematical ordering of the fields. Hence there are in fact only 18 physically distinct components of the nonlinear susceptibility. This means that we can write the nonlinear response in a simpler form in terms of a contracted tensor called the **nonlinear optical coefficient tensor** d_{ij} . Written out explicitly, the components of the nonlinear polarizations are given by:

$$\begin{pmatrix} P_x^{(2)} \\ P_y^{(2)} \\ P_z^{(2)} \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{pmatrix} \begin{pmatrix} \mathcal{E}_x\mathcal{E}_x \\ \mathcal{E}_y\mathcal{E}_y \\ \mathcal{E}_z\mathcal{E}_z \\ 2\mathcal{E}_y\mathcal{E}_z \\ 2\mathcal{E}_z\mathcal{E}_x \\ 2\mathcal{E}_x\mathcal{E}_y \end{pmatrix}. \quad (11.46)$$

By comparing this with eqn 11.11 we see that $d_{11} = \epsilon_0\chi_{xxx}^{(2)}$, $d_{14} = \epsilon_0\chi_{xyz}^{(2)}$, etc.

In many crystals the nonlinear optical coefficient tensor can be further simplified because the crystal symmetry requires that many of the terms are zero, and many others are the same. This is a consequence of Neumann's principle which states that the macroscopic physical properties of a crystal must be invariant under the symmetry operations of the crystal. (cf. Section 1.5.1.) This makes the task of characterizing the nonlinear material much easier than it might seem at first.

The simplest case to consider is that of a centrosymmetric crystal. This is a crystal that has inversion symmetry. Suppose we generate a nonlinear polarization in such a crystal by using a single applied field, \mathbf{E} . The components of

$\mathbf{P}^{(2)}$ will be given by eqn 11.11. If we now reverse the direction of the electric field, nothing happens because:

$$\begin{aligned} P_i^{(2)}(-\mathbf{E}) &= \epsilon_0 \sum_{j,k} \chi_{ijk}^{(2)}(-\mathcal{E}_j)(-\mathcal{E}_k) \\ &= \epsilon_0 \sum_{j,k} \chi_{ijk}^{(2)} \mathcal{E}_j \mathcal{E}_k \\ &= P_i^{(2)}(+\mathbf{E}). \end{aligned} \quad (11.47)$$

However, since the crystal has inversion symmetry, we know from Neumann's principle that we must get the same physical result by keeping the field in the original direction and inverting the crystal. In terms of the coordinate axes of the inverted crystal, all the components of \mathbf{E} and $\mathbf{P}^{(2)}$ change sign. Therefore, for the inverted crystal we have:

$$-P_i^{(2)} = \epsilon_0 \sum_{j,k} \chi_{ijk}^{(2)}(-\mathcal{E}_j)(-\mathcal{E}_k). \quad (11.48)$$

The only way that eqn 11.48 can be compatible with eqn 11.47 is if $\chi_{ijk}^{(2)} = 0$ for all permutations of i, j, k . Hence we conclude that the second-order nonlinear susceptibility of centrosymmetric crystals is zero, so that $d_{ij} = 0$ for all i and j .

We can come to the same conclusion by a more obvious but less rigorous route by making reference to eqn 11.18. In a centrosymmetric crystal we must have that $U(x) = U(-x)$ because the physical properties cannot alter on inverting the crystal. This means that $C_3 = 0$. Hence from eqn 11.32 we see that $\chi^{(2)}$ must be zero in a centrosymmetric crystal.

If the crystal does not possess inversion symmetry, then some of the components of d_{ij} will be non-zero. In triclinic crystals with the lowest possible symmetry, it will be necessary to specify all 18 values of d_{ij} to describe the nonlinear response fully. At the other extreme, in materials with the zinc blende structure (class $\bar{4}3m$), it is only necessary to specify one value because the very high degree of crystal symmetry requires that the only non-zero terms are d_{14} , d_{25} , and d_{36} , and that these three are all identical.

In crystals with intermediate symmetry, it will be necessary to specify a varying number of physically different terms in d_{ij} . For example, we already considered the uniaxial nonlinear crystal KDP (potassium dihydrogen phosphate, KH_2PO_4) in Example 11.1. KDP belongs to the tetragonal crystal class $42m$ and has four-fold rotational symmetry about the z axis. The non-zero components of $\chi_{ijk}^{(2)}$ specified in eqns 11.13 and 11.14 imply that the only non-zero components of d_{ij} are d_{14} , d_{25} , and d_{36} , with d_{14} equal to d_{25} . Hence we can characterize the nonlinear response completely by making just two separate measurements to determine d_{14} and d_{36} . In other types of crystal, there will be different relationships between the coefficients of d_{ij} . Tables of these relationships are given in books on crystallography and nonlinear optics. Exercise 11.6 works through a specific example for the case of an orthorhombic nonlinear crystal.

The non-zero terms in d_{ij} for the cubic crystal are derived from the terms of the type $\chi_{xyz}^{(2)}$. It is obvious that $\chi_{xyz}^{(2)} = \chi_{yzx}^{(2)} = \chi_{zxy}^{(2)}$ in a cubic crystal in which the x , y and z axes are equivalent.

11.3.3 Phase matching

Nonlinear effects are generally small, and we therefore need a long length of the nonlinear medium to obtain a useful nonlinear conversion efficiency. For this to work, we need that the phases of the nonlinear waves generated throughout the whole crystal are all the same so that the fields add together coherently. When this is achieved, we are in a regime called **phase matching**. As we will see below, phase matching does not normally occur, and can only be achieved if the nonlinear crystal is orientated in a very precise direction.

We can see why phase matching is an important issue by considering a simple example. Suppose we wish to use a nonlinear crystal to double the frequency of a Nd:YAG laser from 1064 nm to 532 nm, as shown schematically in Fig. 11.5. All materials are dispersive to some extent, and this means that the refractive index at 532 nm will be different to that at 1064 nm. Therefore, the second harmonic waves at 532 nm will propagate with a different phase velocity to the fundamental at 1064 nm. The second harmonic waves generated at the front will arrive at the back of the crystal at a different time to the fundamental, and so the 532 nm waves generated at the back of the crystal will be out of phase with those from the front.

The phase mismatch introduced by the frequency doubling process can be calculated from the wave vectors of the two waves. If the beams are travelling in the z direction, then the nonlinear waves will propagate as $\exp ik^{(2\omega)}z$, where $k^{(2\omega)}$ is the wave vector at frequency 2ω . On the other hand, the fundamental beam propagates as $\exp ik^{(\omega)}z$, where $k^{(\omega)}$ is the wave vector at frequency ω . Now since $P^{(2)} \propto \mathcal{E}^2$, the nonlinear polarization at a given point in the medium will be created with a phase of $(\exp ik^{(\omega)}z)^2 = \exp 2ik^{(\omega)}z$. Hence the phase difference $\Delta\Phi$ between the nonlinear waves created at a distance z into the crystal and those created at the front of the crystal is given by

$$\Delta\Phi = (k^{(2\omega)} - 2k^{(\omega)})z. \quad (11.49)$$

We introduce the coherence length l_c for the nonlinear process as the distance over which the phase mismatch becomes equal to 2π :

$$(k^{(2\omega)} - 2k^{(\omega)}) \times l_c = 2\pi. \quad (11.50)$$

This can be rewritten in terms of the refractive indices $n^{2\omega}$ and n^ω at the two frequencies as:

$$\frac{2\omega}{c} [n^{2\omega} - n^\omega] l_c = 2\pi. \quad (11.51)$$

Hence

$$l_c = \frac{\pi c}{\omega[n^{2\omega} - n^\omega]} = \frac{\lambda}{2[n^{2\omega} - n^\omega]}, \quad (11.52)$$

where λ is the free space wavelength of the fundamental beam. Taking a typical example with $\lambda = 1 \mu\text{m}$ and $n^{2\omega} - n^\omega \sim 10^{-2}$ we find that $l_c \sim 50 \mu\text{m}$.

Equation 11.52 shows us that only the waves emitted within a very short distance of the surface will add together coherently. This clearly greatly restricts the efficiency of the nonlinear conversion process, because only a very short length of the nonlinear crystal is actually useful. The situation would be completely different if we could somehow arrange that $n^{2\omega} = n^\omega$. In this case the nonlinear waves generated throughout the whole crystal would all have the

same phase and would thus add together coherently. This is the phase matching condition.

At first sight it might seem that there is no way to satisfy the condition $n^{2\omega} = n^\omega$ in any material with normal dispersion properties. However, this neglects the fact that the anisotropic crystals that are used for nonlinear mixing are birefringent. This opens new possibilities to balance the dispersion against the birefringence. For example, in a uniaxial crystal with normal dispersion such that $n^{2\omega} > n^\omega$, it is possible to obtain phase matching by propagating the beam at frequency 2ω as an extraordinary ray, and the beam at ω as an ordinary ray. It is shown in Example 11.2 below that phase matching can be achieved for a very specific orientation of the crystal.

The phase matching condition can be given an intuitive physical interpretation if we notice that if $n^{2\omega} = n^\omega$, then $k^{(2\omega)} = 2k^{(\omega)}$. This corresponds to momentum conservation in the nonlinear process. In the more general case when a photon of wave vector \mathbf{k} is generated by mixing two photons with wave vectors \mathbf{k}_1 and \mathbf{k}_2 , the phase matching condition can be written as

$$\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2. \quad (11.53)$$

In down-conversion where one photon is split into two output photons, the phase matching condition of eqn 11.53 applies to each pair of photons created in the process.

Example 11.2

The ordinary and extraordinary refractive indices of a uniaxial crystal are n_e and n_o respectively. A laser beam is propagating at an angle θ to the optic (z) axis as shown in Fig. 2.12. The laser is linearly polarized along the x direction.

- (i) Show that there is an angle θ at which the phase matching condition can be met for second-harmonic waves polarized as extraordinary rays.
- (ii) Evaluate the phase matching angle for potassium dihydrogen phosphate (KDP) at the wavelength of a Nd:YAG laser (1064 nm). The relevant refractive indices for KDP are: $n_o(1064 \text{ nm}) = 1.494$, $n_o(532 \text{ nm}) = 1.512$, and $n_e(532 \text{ nm}) = 1.471$.

Solution

- (i) The general condition for phase matching is that

$$n^{2\omega} = n^\omega. \quad (11.54)$$

The fundamental is polarized along the x axis, and so its refractive index is n_o^ω , irrespective of θ . The refractive index for the second harmonic waves polarized as extraordinary rays is given by the result of Exercise 2.15, namely:

$$\frac{1}{n(\theta)^2} = \frac{\sin^2 \theta}{n_e^2} + \frac{\cos^2 \theta}{n_o^2}, \quad (11.55)$$

where n_0 and n_e are evaluated at 2ω . Hence the phase matching condition given in eqn 11.54 is met when

$$\frac{1}{(n_0^{2\omega})^2} = \frac{\sin^2 \theta}{(n_e^{2\omega})^2} + \frac{\cos^2 \theta}{(n_0^{2\omega})^2}. \quad (11.56)$$

- (ii) We insert the appropriate values for the refractive indices into eqn 11.56 to write:

$$\frac{1}{1.494^2} = \frac{\sin^2 \theta}{1.471^2} + \frac{\cos^2 \theta}{1.512^2}.$$

This is satisfied with $\theta = 41^\circ$. To achieve phase matching we therefore hold the crystal in a gimbal mount and carefully rotate its orientation until the optic axis is at 41° to the direction of the laser beam.

11.4 Third-order nonlinear effects

Third-order nonlinear effects are particularly important in isotropic media, such as gases, liquids and glasses. This is because an isotropic medium possesses inversion symmetry, and all the components of $\chi_{ijk}^{(2)}$ must therefore be zero. (See Section 11.3.2.) Hence the lowest order nonlinear susceptibility with non-vanishing components is $\chi^{(3)}$. In this section we start by giving an overview of third-order nonlinear phenomena, and then focus on isotropic media in more detail, mentioning optical fibres as a specific example. Finally we discuss resonant third-order nonlinear effects in semiconductor crystals.

11.4.1 Overview of third-order phenomena

A third-order nonlinear polarization is generated when three input fields are applied to the nonlinear medium. If the input fields are at frequencies ω_1 , ω_2 and ω_3 , then the nonlinear polarization will be given by eqn 11.8 or more generally by eqn 11.12. In the simplest case without considering the tensor aspect of the susceptibility, this gives:

$$P^{(3)}(t) = \epsilon_0 \chi^{(3)} \times \mathcal{E}_1 \cos \omega_1 t \times \mathcal{E}_2 \cos \omega_2 t \times \mathcal{E}_3 \cos \omega_3 t \quad (11.57)$$

where \mathcal{E}_1 , \mathcal{E}_2 and \mathcal{E}_3 are the amplitudes of the three waves. Hence the frequency ω_4 of the nonlinear polarization must satisfy:

$$\omega_4 = \omega_1 + \omega_2 + \omega_3, \quad (11.58)$$

where the frequencies on the right hand side can be either positive or negative. This reflects the fact that $\cos \omega t = \frac{1}{2}(e^{+i\omega t} + e^{-i\omega t})$ and therefore contains both positive and negative frequency terms.

Figure 11.6 shows the Feynman diagram for several third-order nonlinear processes. Figure 11.6(a) gives the diagram for the general process, with three input photons corresponding to the driving fields and one output photon corresponding to the nonlinear polarization. The output frequency is the sum of

As with the second-order processes, the negative frequencies represent the addition of photons to the driving field in the nonlinear process. This is a legitimate outcome because photons are bosons and hence the nonlinear interaction can stimulate the creation of photons at the input frequency as well as causing annihilation of input photons.

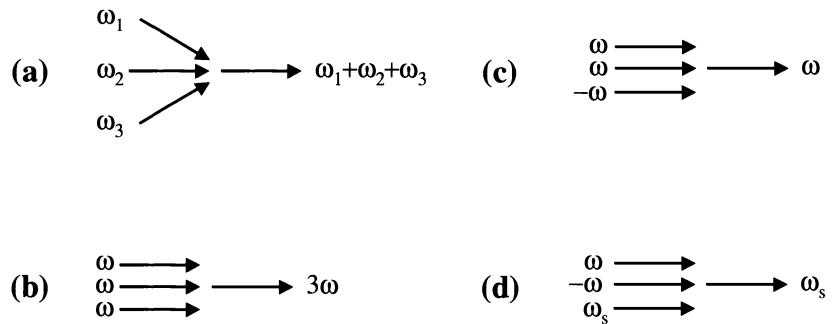


Fig. 11.6 Feynman diagrams for third-order nonlinear processes. (a) Four-wave mixing. (b) Frequency tripling. (c) The optical Kerr effect. (d) The Raman effect.

the input frequencies, as required by conservation of photon energy. Since four photons are involved, the general phenomenon is often called **four-wave mixing**. In the discussion below, we concentrate on the three specific examples of four-wave mixing illustrated in Fig. 11.6(b)–(d), namely: frequency tripling, the optical Kerr effect, and stimulated Raman scattering. There are, of course, many other important third-order nonlinear phenomena, but we do not have space here to discuss them all. The main effects that we will consider here are summarized in Table 11.2.

Frequency tripling

In practice it is usually easier to generate the third harmonic of a laser beam by two second-order processes according to the scheme shown in Fig. 11.5, rather than by using a single third-order conversion using the $\chi^{(3)}$ nonlinearity.

Figure 11.6(b) shows the Feynman diagram for frequency tripling. This is the equivalent of frequency doubling for a $\chi^{(2)}$ process. Three collinear fields at the same frequency are incident on the medium from a single input laser beam. With $\omega_1 = \omega_2 = \omega_3 = +\omega$, we find from eqn 11.58 that the output frequency is equal to 3ω . In other words, the nonlinear process directly generates the third harmonic of the fundamental. As with frequency doubling, the conversion efficiency will only be large if the phase matching condition determined by momentum conservation is satisfied. (See Section 11.3.3.) Frequency tripling experiments are useful for the spectroscopic information they give about the magnitude of $\chi^{(3)}$ and its relation to the atomic transitions of the medium.

Table 11.2 Third-order nonlinear effects. The third column lists the frequencies of the light beams incident on the nonlinear crystal, while the fourth gives the frequency of the output beam or the nonlinear polarization. A frequency of zero indicates a DC electric field. Note that there are several other third-order nonlinear effects such as two-photon absorption which have not been included in this table.

Effect	Alternative names	Input frequencies	Output frequencies
Generic four-wave mixing		$\omega_1, \omega_2, \omega_3$	$ \pm \omega_1 \pm \omega_2 \pm \omega_3 $
Frequency tripling	Third-harmonic generation	ω	3ω
Optical Kerr effect	Degenerate four-wave mixing Nonlinear refractive index Self-phase modulation	ω	ω
DC Kerr effect	Quadratic electro-optic effect	$\omega, 0$	ω
Stimulated four-wave mixing	Stimulated Raman scattering Stimulated Brillouin scattering	ω, ω_s	ω_s

The optical Kerr effect and the nonlinear refractive index

Figure 11.6(c) shows the Feynman diagram for the optical Kerr effect. In this process we again have just a single beam at frequency ω incident on the nonlinear medium, and the nonlinear interaction produces a third-order polarization at the same frequency as the input laser beam. This works by using $\omega_1 = \omega_2 = +\omega$ and $\omega_3 = -\omega$ in eqn 11.58. No phase matching problems occur in this case because the nonlinear polarization is at the same frequency as the driving fields and thus the fields are in-phase throughout the whole medium. Since the frequencies of all four photons are the same, the optical Kerr effect is sometimes called degenerate four-wave mixing.

One of the main consequences of the optical Kerr effect is that the refractive index begins to depend on the intensity of the beam. This can be seen by calculating the change of the dielectric constant produced by the light. From eqn 11.10 we see that the dielectric constant in a nonlinear medium with $\chi^{(2)} = 0$ is given by:

$$\epsilon_r^{\text{nonlinear}} = 1 + \chi^{(1)} + \chi^{(3)} \mathcal{E}^2, \quad (11.59)$$

where $\chi^{(1)}$ is the linear susceptibility and \mathcal{E} is the optical electric field amplitude. We split this into its linear and nonlinear parts by writing:

$$\epsilon_r^{\text{nonlinear}} = \epsilon_r + \Delta\epsilon, \quad (11.60)$$

where

$$\epsilon_r = 1 + \chi^{(1)} \quad (11.61)$$

and

$$\Delta\epsilon = \chi^{(3)} \mathcal{E}^2. \quad (11.62)$$

ϵ_r is the usual dielectric constant for the linear regime and $\Delta\epsilon$ is the change caused by the nonlinear process. In a non-absorbing medium, the refractive index n is equal to the square root of the relative dielectric constant (cf. eqn A.31). Hence we may write:

$$n = (\epsilon_r + \Delta\epsilon)^{\frac{1}{2}} = \sqrt{\epsilon_r} + \frac{\Delta\epsilon}{2\sqrt{\epsilon_r}} = n_0 + \Delta n, \quad (11.63)$$

where we have assumed that $\Delta\epsilon \ll \epsilon_r$ in the second equality, and in the third we have split the refractive index into its linear part $n_0 = \sqrt{\epsilon_r}$ and its nonlinear part Δn . On comparing eqns 11.62 and 11.63 we find that

$$n = n_0 + \frac{\chi^{(3)} \mathcal{E}^2}{2n_0} = n_0 + \frac{\chi^{(3)}}{n_0^2 c \epsilon_0} I, \quad (11.64)$$

where we have used the proportionality between I and \mathcal{E}^2 given by eqn 11.17 in the second identity.

We now introduce the **nonlinear refractive index** n_2 according to:

$$n(I) = n_0 + n_2 I. \quad (11.65)$$

By comparing eqns 11.64 and 11.65 we find that:

$$n_2 = \frac{1}{n_0^2 c \epsilon_0} \chi^{(3)}. \quad (11.66)$$

This shows that n_2 is directly proportional to $\chi^{(3)}$, and hence that third-order nonlinearities cause the refractive index to vary with the intensity.

From eqn 11.64 we see that in the optical Kerr effect, Δn is proportional to \mathcal{E}^2 , where \mathcal{E} is the electric field of the light wave. This is reminiscent of the **DC Kerr effect**, in which the refractive index of a medium at wavelength λ is modulated by a DC electric field according to the relation:

$$\Delta n = \lambda K \mathcal{E}^2 \quad (11.67)$$

where K is the Kerr constant. The optical and DC Kerr effects are therefore both **quadratic electro-optic effects**, since Δn varies as \mathcal{E}^2 . The DC Kerr effect is used for making optical modulators, and can be considered as a $\chi^{(3)}$ process in which $\omega_1 = \omega_2 = 0$ and $\omega_3 = \omega$.

Stimulated Raman scattering

Figure 11.6(d) shows the Feynman diagram for the **stimulated Raman effect**. A weak beam at frequency ω_s is incident on the medium together with a powerful pump beam at frequency ω . With $\omega_1 = +\omega$, $\omega_2 = -\omega$, and $\omega_3 = \omega_s$, we see from eqn 11.58 that the nonlinear wave is at frequency ω_s . Hence the presence of a field at ω_s generates more photons at the same frequency through nonlinear mixing with the pump field. The beam at frequency ω_s can then experience gain by the same sort of parametric mixing that was discussed in Section 11.3.1. The process becomes a Raman effect if we tune the frequency of ω_s such that $\omega - \omega_s = \Omega$, where Ω is the frequency of one of the vibrational modes of the medium. The nonlinear susceptibility is resonantly enhanced in these conditions because the two frequencies are strongly coupled together through the natural vibrations of the medium.

Stimulated Raman scattering was discovered very soon after the invention of the laser. In 1962 E.J. Woodbury and W.K. Ng observed that when passing a strong beam from a ruby laser at 694.3 nm through a nitrobenzene cell, a second intense beam was generated at 766 nm. Analysis of these results showed that the difference in the frequency of the two photons corresponds exactly with one of the vibrational modes of the molecule at 4.0×10^{13} Hz. The same phenomenon has subsequently been observed in many different liquids and gases, and also in solids.

In solids the scattering can be mediated either by the optical or the acoustic vibrational modes. In the former case it is the Raman active LO and TO phonons at $q = 0$ that are involved, where q is the wave vector of the phonon, as discussed in Section 10.5.2. This gives rise to discrete frequency shifts analogous to those observed in molecules. In the latter case, it is the acoustic phonons that are involved, and the process is usually called **stimulated Brillouin scattering**. The frequency shift caused by stimulated Brillouin scattering depends on the angle through which the light is scattered, as determined by the energy and wave vector conservation laws. (See eqn 10.33.)

A further possibility that occurs in solids is that the stimulated light scattering can be mediated by other types of fundamental excitations of the medium. One particularly interesting example is spin-flip Raman scattering in a doped semiconductor in a magnetic field. The Raman scattering is mediated by the

The quadratic Kerr-type effects contrast with the Pockels effect discussed in Section 11.3.1 and Exercise 11.7. The Pockels effect is a linear electro-optic effect in which Δn varies linearly with \mathcal{E} .

The spontaneous Raman scattering processes discussed in Section 10.5 can be related to the stimulated effects we are considering here by making an analogy with spontaneous and stimulated radiative emission. Spontaneous radiative emission can be considered to be a stimulated process that is triggered by a vacuum photon from the zero-point fluctuations of the quantized electromagnetic field. In the same way we can regard spontaneous Raman scattering as a stimulated nonlinear process instigated by vacuum photons. In fact, when we generate stimulated Raman beams by passing a laser through a suitable medium, there is usually no initial field at frequency ω_s to start the process. This field has to come from spontaneous Raman scattering, which itself is instigated by vacuum fluctuations. Hence the generation of a stimulated Raman beam is considered to start from the zero-point fluctuations of the field.

Table 11.3 Non-zero components of the third-order nonlinear susceptibility in an isotropic medium.

$\chi_{xxxx}^{(3)} = \chi_{yyyy}^{(3)} = \chi_{zzzz}^{(3)}$
$\chi_{xxyy}^{(3)} = \chi_{xyxx}^{(3)} = \chi_{xxzz}^{(3)} = \chi_{zzxx}^{(3)} = \chi_{yyzz}^{(3)} = \chi_{zzyy}^{(3)}$
$\chi_{xyxy}^{(3)} = \chi_{yxxy}^{(3)} = \chi_{xzxz}^{(3)} = \chi_{zxzx}^{(3)} = \chi_{yzzy}^{(3)} = \chi_{zyzy}^{(3)}$
$\chi_{xyyx}^{(3)} = \chi_{yxyx}^{(3)} = \chi_{xzxz}^{(3)} = \chi_{zxzx}^{(3)} = \chi_{yzzy}^{(3)} = \chi_{zyzy}^{(3)}$

flip of the electron spin in the magnetic field, and this has been demonstrated to give sufficient gain to allow laser operation.

11.4.2 Isotropic third-order nonlinear media

The third-order nonlinear response is described by the nonlinear susceptibility tensor $\chi_{ijkl}^{(3)}$ defined in eqn 11.12. This has 81 elements, many of which are the same or zero in materials with a high degree of symmetry. In a completely isotropic medium such as a gas, there are 21 non-zero elements which are listed in Table 11.3. The interrelationships listed in the table suggest that there are four independent values, but this is not the case because it can be shown that the susceptibilities must also satisfy:

$$\chi_{xxxx}^{(3)} = \chi_{xxyy}^{(3)} + \chi_{xyxy}^{(3)} + \chi_{xyyx}^{(3)}. \quad (11.68)$$

This means that there are in fact only three independent elements. Furthermore, if we are well below the natural resonance frequencies of the atoms, it will also be true that:

$$\chi_{xxyy}^{(3)} = \chi_{xyxy}^{(3)} = \chi_{xyyx}^{(3)} = \frac{1}{3}\chi_{xxxx}^{(3)}. \quad (11.69)$$

This result is only valid at low frequencies, and is called **Kleinman symmetry**. In this limit, there are thus only two physically different third-order nonlinear susceptibilities.

Glasses are perhaps the most important examples of isotropic optical materials. We usually use glasses at wavelengths where they are transparent. We are therefore well below the band gap in the ultraviolet, and the Kleinman symmetry condition given in eqn 11.69 will normally hold.

When an intense laser propagates through a glass, it can alter the refractive index by the optical Kerr effect in accordance with eqn 11.65. This produces a nonlinear phase shift given by

$$\Delta\Phi^{\text{nonlinear}} = \frac{2\pi}{\lambda} \Delta n l = \frac{2\pi}{\lambda} n_2 I l, \quad (11.70)$$

where λ is the wavelength, l is the length of the medium and I is the intensity. The laser beam therefore alters its own phase, an effect called **self-phase modulation**.

Nonlinear propagation in optical fibres and solitons

Self-phase modulation effects can be observed very clearly in optical fibres, even though the nonlinear refractive index is very small. This is because the

Glasses are isotropic because they do not possess a crystal structure and therefore have no preferred axes. Doped glasses have optical transitions in the visible spectral region (see Section 1.4.5) and the Kleinman symmetry condition does not hold at frequencies close to the band gap of the dopant.

Fig. 11.7 A group of runners on a mattress is an analogy of a soliton pulse in an optical fibre. The indentation of the group slows the faster runners and speeds the slower ones, thereby compensating for the tendency of the group to break up. In the same way, the nonlinear phase shift of an intense short pulse can compensate for the pulse broadening effect due to the dispersion of the optical fibre. From [1], reprinted with permission from Plenum Publishers.



beam is focussed to a very small area in the fibre, so that the intensity is very high even at only moderate power levels. Large phase shifts can then be achieved by using long fibre lengths.

The subject of **solitons** is a very interesting aspect of the propagation of light pulses down an optical fibre in the nonlinear regime. A short laser pulse must necessarily contain a spread of frequencies in order to satisfy the Fourier transform limit given approximately by eqn 9.6. Since the glass that makes up the fibre is dispersive (see Fig. 2.9), the different frequency components of the pulse will experience slightly different refractive indices. This means that their velocities will be different, and the pulse will gradually broaden in time as it propagates down the fibre. (See Exercise 2.14.) This becomes a serious problem if we are trying to transmit a sequence of closely spaced data pulses down the fibre. The soliton effect discovered by John Scott Russell in 1834 can eliminate the problem. Russell noticed that the bow wave of a Scottish canal barge did not disperse when the amplitude of the wave was large enough. He successfully explained his observation by realizing that the dispersion of the water wave was being balanced by the nonlinear effects due to the large amplitude. The same phenomenon can occur in optical fibres.

Solitons can be observed in optical fibres at frequencies where the dispersion is negative. This is because the n_2 of the fibre is positive, and so the dispersion must be negative if the two effects are to cancel. Most glasses have positive dispersion at optical frequencies due to the electronic absorption in the ultraviolet. However, as noted in Section 2.3, the dispersion of the SiO_2 used to make optical fibres is zero at $1.3 \mu\text{m}$ and becomes negative for longer wavelengths. This can be seen in the data shown in Fig. 2.7. This means that we are in the right regime to observe solitons at $1.55 \mu\text{m}$, which is the preferred wavelength for telecommunication systems because the losses of the fibre are smallest.

The basic physics of soliton propagation can be understood in a simple way without recourse to heavy mathematics by reference to the cartoon given in Fig. 11.7. This shows a group of athletes crossing a mattress, with the faster runners at the front. The group creates a valley in the mattress that retards the faster runners but helps the slower ones. On a hard surface the group would break up with the faster runners leaving the slower ones behind, but the valley in the mattress opposes this effect and keeps the group together. In the right conditions the nonlinear phase shift of an intense pulse in an optical fibre can have a similar effect, creating a trapped light pulse similar to the trapped group of runners in the cartoon. The resulting pulse can travel indefinitely long distances along the fibre without significant broadening. This is a very useful property for long-distance optical fibre telecommunication systems.

Example 11.3

A laser pulse at $1.55 \mu\text{m}$ of intensity 10^{12} W m^{-2} is propagating down an optical fibre of length 100 m. Calculate the nonlinear phase shift if the nonlinear refractive index is $3 \times 10^{-20} \text{ m}^2 \text{ W}^{-1}$.

Solution

We calculate the nonlinear phase shift from eqn 11.70 with $\lambda = 1.55 \times 10^{-6} \text{ m}$, $n_2 = 3 \times 10^{-20} \text{ m}^2 \text{ W}^{-1}$, and $I = 10^{12} \text{ W m}^{-2}$. This gives :

$$\Delta\Phi_{\text{nonlinear}} = \frac{2\pi}{1.55 \times 10^{-6}} \times (3 \times 10^{-20}) \times 10^{12} \times 100 = 12.2.$$

The nonlinear phase shift is thus 3.1π . This example shows that the nonlinear phase shift can be very large, even though the nonlinearity is small. This occurs because of the very large value of l that can be used in optical fibres.

11.4.3 Resonant nonlinearities in semiconductors

In Section 11.2.2 we explained how the absorption coefficient of an absorbing medium is expected to depend on the intensity at high power levels. This phenomenon has been studied extensively in semiconductors, with a particular emphasis on developing nonlinear switching devices for applications in optical information processing.

The simplest mechanism that can cause saturable absorption in a semiconductor is illustrated in Fig. 11.8. This shows the band diagram of a direct gap semiconductor when a large number of electrons have been excited from the conduction band to the valence band. The electrons are excited by absorption of an intense laser pulse with photon energy greater than the band gap. The electrons fill up the states at the bottom of the conduction band, leaving empty states at the top of the valence band. This blocks further interband absorption transitions at photon energies close to the band gap such as the one shown in the figure, because there are no electrons in the valence band to be excited, and furthermore, the destination states in the conduction band are full. Therefore, as we turn up the intensity of the exciting laser, we will find that the absorption gradually saturates as the bands fill up. This effect is therefore called a **band-filling nonlinearity**.

Another mechanism which can cause the absorption to depend on the intensity is the saturation of the excitons at high carrier densities. This effect was described in Section 4.4. At high carrier densities the Coulomb force holding the excitons together is screened, and the electron and hole states near $k = 0$ required for the formation of excitons are filled. Both these effects cause bleaching of the exciton absorption. If the carriers are excited by absorption of a laser beam, the excitonic absorption will decrease with increasing laser intensity.

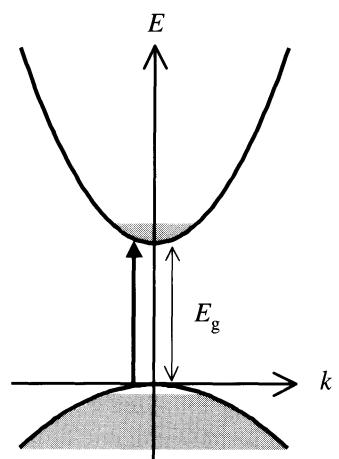


Fig. 11.8 Band-filling nonlinearity in an excited semiconductor. The interband absorption transition shown by the thick arrow in the diagram is blocked because there are no electrons in the valence band to absorb, and the destination states in the conduction band are filled.

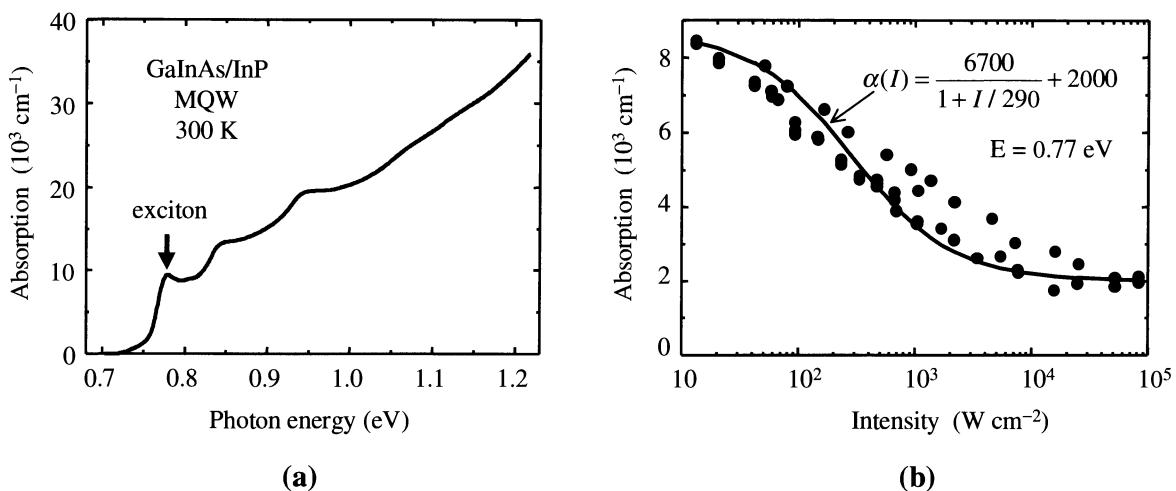


Fig. 11.9 Saturable excitonic absorption in a semiconductor multiple quantum well. The sample contained 30 $\text{Ga}_{0.53}\text{In}_{0.47}\text{As}$ quantum wells of width 15.4 nm with InP barriers. The temperature was 300 K. Part (a) shows the linear absorption of the sample, while (b) shows the intensity dependence of the absorption at 0.77 eV. This photon energy corresponds to the first excitonic peak in the linear absorption spectrum, and is identified by the arrow in (a). The solid line in (b) is a fit to the data using eqn 11.71. After [2] and [3], copyright 1987 American Institute of Physics, reprinted with permission.

Quantum wells are good materials to demonstrate excitonic nonlinearities because they show strong excitonic absorption even at room temperature. See Section 6.4.4 for further details.

Figure 11.9 shows the intensity dependence of the exciton absorption in a GaInAs multiple quantum well sample at room temperature. The linear absorption spectrum is given in Fig. 11.9(a). The step-like spectrum expected for the 2-D quantum wells is clearly apparent in the data, together with the peaks due to the excitons. Figure 11.9(b) shows the dependence of the absorption coefficient α on the intensity I at the first exciton peak in the linear absorption spectrum. The data show that α decreases as I increases, and fits reasonably well to a dependence of the form:

$$\alpha(I) = \alpha_b + \frac{\alpha_0}{1 + I/I_s}, \quad (11.71)$$

with $\alpha_b = 2000 \text{ cm}^{-1}$, $\alpha_0 = 6700 \text{ cm}^{-1}$ and $I_s = 290 \text{ W cm}^{-2}$. The α_b term in eqn 11.71 represents an unsaturable background absorption, while the second term represents a saturable absorption of the form given in eqn 11.40.

The intensity dependence of α shown in Fig. 11.9(b) gives rise to a relatively large change in the refractive index. (See Exercise 11.13.) The nonlinear refractive index n_2 associated with the exciton saturation was estimated to be $-3 \times 10^{-8} \text{ m}^2 \text{ W}^{-1}$. This is about 12 orders of magnitude larger than the n_2 value of an optical fibre at the same wavelength. This illustrates the point that resonant nonlinearities are much larger than non-resonant ones. However, the price that has to be paid for the large $|n_2|$ value is the high absorption at the operating wavelength. This restricts the length of the devices that can be used, and therefore puts a limit on the maximum nonlinear phase shift that can be obtained.

The resonant nonlinear refractive index of semiconductors can be used to make optical switching devices. Figure 11.10 shows the results obtained from an InSb sample at 5 K. InSb is a direct gap III-V semiconductor with $E_g = 0.23 \text{ eV}$. The band-filling nonlinearity at photon energies close to E_g can be

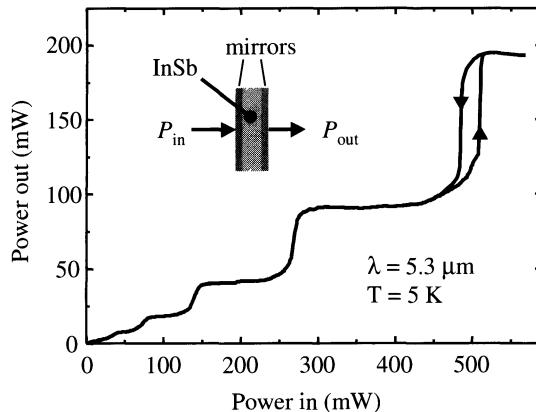


Fig. 11.10 Optical bistability in an InSb platelet of thickness 0.56 mm at 5 K. The laser wavelength was 5.277 μm . The platelet functions as a weak optical cavity, with the mirrors formed by the natural reflectivity of the air-InSb interface. The experiment consisted in measuring the output power as a function of the input power, as shown in the inset. After [4], copyright © 1981 IEEE, reprinted with permission.

exploited by using a carbon monoxide laser, which has several emission lines around 5.3 μm . The experiment consisted in measuring the power transmitted through a platelet sample as the input laser power was increased. The data clearly demonstrates that the output power is not a linear function of the input power, with sharp jumps observed at several power levels.

The nonlinear input–output characteristic of the device can be understood as follows. The front and back surfaces of the InSb platelet have a natural reflectivity R of 36 %, and the sample forms a weak optical cavity. In the absence of background absorption, the transmission T of the cavity is given by:

$$T = \frac{1}{1 + F \sin^2(\Phi)} , \quad (11.72)$$

where $F = 4R/(1 - R)^2$ is the ‘finesse’ of the cavity. The phase Φ is determined by the optical path length per round trip according to:

$$\Phi = \frac{2nl}{\lambda} 2\pi = \frac{4\pi nl}{\lambda} , \quad (11.73)$$

where n is the refractive index, l is the thickness of the platelet (0.56 mm), and λ is the laser wavelength (5.277 μm). Whenever $\Phi = 2\pi M$, where M is an integer, the etalon is on resonance and $T = 100$ %. On the other hand, if $\Phi = 2\pi \times (M + \frac{1}{2})$, T drops to 22 %. The phase varies with the input power because n decreases linearly with the intensity inside the etalon through the nonlinear refractive index, which has a value of $-3 \times 10^{-9} \text{ m}^2 \text{ W}^{-1}$ at 5.277 μm . The jumps in the input–output characteristic occur whenever the intensity I inside the etalon is such that $2(n_0 + n_2 I)l/\lambda = M$.

One striking aspect of the data shown in Fig. 11.10 is the hysteresis loop when the input power is between 480 mW and 510 mW. At these power levels we have two possible output powers for the same input power. This effect is called **optical bistability**. The hysteresis occurs because of the feedback mechanism between I and the output power. High output power corresponds to high I , and *vice versa*. In the right conditions, the high intensity inside the etalon can pull the cavity onto resonance by producing the correct nonlinear phase shift, which in turn further increases I . The two bistable states thus correspond to the high-intensity, on-resonance and the low-intensity, off-resonance

The optical cavities formed between two parallel mirrors are called Fabry–Perot etalons. The resonance condition $\Phi = \text{integer} \times 2\pi$ can be re-expressed from eqn 11.73 as $2nl = \text{integer} \times \lambda$. The light that bounces around the cavity is therefore in phase, and the fields add up constructively to give unity transmission. If the resonance condition is not satisfied, then some of the field will be out of phase and destructive interference will occur at the output. The transmission is thus less than unity. For further details, see Hecht (1998).

conditions. We can use these bistable states to store information, with one state representing binary ‘1’, and the other representing binary ‘0’. The device can be switched back and forth between 0 and 1 by altering the power of the input beam, and can act as a binary logic gate whose state is entirely controlled by optical beams.

Chapter summary

- Nonlinear optical effects become important at the high power levels available from lasers. Nonlinear effects cause the optical susceptibility, and all the properties that follow from it, to depend on the magnitude of the electric field of the light.
- The nonlinear polarization can be calculated from the nonlinear susceptibility tensor and the components of the electric field. The form of the tensor is determined by the symmetry class of the crystal. All the components of the second-order nonlinear susceptibility tensor of materials with inversion symmetry are zero.
- Non-resonant nonlinearities are caused by anharmonicity in the restoring force of the bound electrons.
- Resonant nonlinearities are caused by stimulated emission from the upper level when its population becomes significant. The saturation of interband absorption can be understood in terms of the blocking of the transitions due to the Pauli exclusion principle.
- The nonlinear polarization in a second-order effect is proportional to the product of two electric fields. This gives rise to a series of second-order effects such as frequency doubling, sum frequency generation, and difference frequency generation.
- Strong second-order nonlinear signals are only achieved when the phase matching condition is satisfied.
- The nonlinear polarization in a third-order effect is proportional to the product of three fields. Third-order effects are generally known as four-wave mixing.
- The third-order susceptibility is the lowest order non-zero term in the nonlinear response of isotropic media such as gases, liquids and glasses.
- Third-order effects give rise to the nonlinear refractive index, frequency tripling, and stimulated Raman scattering. Self-phase modulation describes the nonlinear phase shift induced by a beam through the nonlinear refractive index.
- The saturation of interband transitions and excitonic absorption in a semiconductor gives rise to large third-order nonlinear coefficients.

Further reading

More extensive introductory reading on nonlinear optics may be found in Yariv (1997). Butcher and Cotter (1990) give a comprehensive introduction to the theory of nonlinear optics.

A classic discussion of the effects of the crystal symmetry on the physical properties of materials is given by Nye (1957).

The optical nonlinearities due to excitons are reviewed in Chemla (1985). A more detailed review on excitonic nonlinearities in quantum wells is given by Schmitt-Rink *et al.* (1989).

References

- [1] Mollenauer, L.F. and Gordon, J.P. (1994). In *Nonlinear spectroscopy of solids* (ed. B. Di Bartolo), NATO ASI Series B, Vol. 339. Plenum Press, New York, p. 451.
- [2] Westland, D.J., Fox, A.M., Maciel, A.C., Ryan, J.F., Scott, M.D., Davies, J.I. and Riffat, J.R. (1987). *Appl. Phys. Lett.*, **50**, 839.
- [3] Fox, A.M., Maciel, A.C., Shorthose, M.G., Ryan, J.F., Scott, M.D., Davies, J.I. and Riffat, J.R. (1987). *Appl. Phys. Lett.*, **51**, 30.
- [4] Miller, D.A.B., Smith, S.D. and Seaton, C.T. (1981). *IEEE J. Quantum Electron.*, **17**, 313.

Exercises

- (11.1) Use the Bohr model of the atom to show that the electric field experienced by an electron bound in the n th shell of a hydrogenic atom of atomic number Z is equal to $(Z^3/n^4)e/4\pi\epsilon_0a_H^2$, where a_H is the Bohr radius of hydrogen. Make a rough estimate of this value for the valence electrons of silicon.
- (11.2) Estimate the magnitude of the electric field of the light wave for: (a) a 10 ns long pulse of energy 1 J from a Nd:glass laser with a beam diameter 5 mm; (b) a 1 mW continuous wave semiconductor laser focussed into an optical fibre with a core area of $20\ \mu\text{m}^2$ and a refractive index of 1.45.
- (11.3) A gas of atoms is subjected to a strong DC electric field. State with reasons whether you would expect to be able to observe any frequency doubling or not from the gas.
- (11.4) Which of the following materials would you expect to have non-zero components in the second-order nonlinear susceptibility: (a) NaCl, (b) GaAs, (c) water, (d) glass, (e) crystalline quartz, (f) ZnS (wurtzite).
- (11.5)* An intense laser is propagating through an absorbing medium as described in Section 11.2.2. The medium contains atoms with two non-degenerate levels 1 and 2, and the separation of the levels is resonant with the laser frequency. The atoms are initially in a state with

$N_1 = N_0$ and $N_2 = 0$, and the laser beam is turned on at time $t = 0$.

- (i) Explain why it is not possible to get population inversion between the two levels, no matter how intense the laser is.
- (ii) Show that if spontaneous emission is neglected, then the time dependence of the population difference $\Delta N = N_1 - N_2$ between the two levels is given by:

$$\Delta N(t) = N_0 \exp(-2B_{12}u_v t),$$

where u_v is the energy density of the beam, and B_{12} is the Einstein B coefficient for absorption. What does this relationship imply about the absorption coefficient?

- (11.6) The nonlinear optical coefficient tensor of a crystal from the orthorhombic crystal class is given by

$$\mathbf{d} = \begin{pmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & d_{36} \end{pmatrix}.$$

A laser beam is incident on the crystal. The beam travels in the xy plane, and is polarized with its electric field

* Exercises marked with an asterisk are more challenging.

in the xy plane. Show that a second harmonic beam is produced which is polarized along the z axis. Show also that the magnitude of the second harmonic beam is a maximum if the incoming beam travels at 45° to the x axis.

- (11.7) The electro-optic effect for a crystal with the cubic zinc blende structure can be written in the following form:

$$\begin{aligned} n_{x'} &= n_0 + \frac{1}{2}n_0^3 r_{41} \mathcal{E}_z \\ n_{y'} &= n_0 - \frac{1}{2}n_0^3 r_{41} \mathcal{E}_z \\ n_z &= n_0 \end{aligned}$$

where x' and y' are axes at 45° to the crystallographic axes x and y , and the field is applied along the z direction. n_0 is the refractive index at $\mathcal{E}_z = 0$, and r_{41} is the electro-optic coefficient. $n_{x'}$ and $n_{y'}$ are the modified refractive indices along the x' and y' directions when the field is applied.

- (i) Show that when light of air wavelength λ propagates through the crystal in the z direction, the phase difference introduced between the x' and y' polarizations is equal to

$$\Delta\Phi = \frac{\pi}{\lambda} r_{41} n_0^3 V$$

where V is the voltage applied across the crystal to produce the field along the z direction.

- (ii) Evaluate the voltage at which the phase change is equal to π for CdTe at $10.6 \mu\text{m}$, where $r_{41} = 6.8 \times 10^{-12} \text{ m/V}$ and $n_0 = 2.6$.

- (11.8) Calculate the phase matching angle for KDP at the wavelength of the ruby laser (694 nm). The relevant refractive indices are: $n_0(694 \text{ nm}) = 1.506$, $n_0(347 \text{ nm}) = 1.534$, $n_e(347 \text{ nm}) = 1.490$.

- (11.9) Explain why the imaginary part of the third-order nonlinear susceptibility must be non-zero in a saturable absorber medium.

- (11.10) Show that the third-order nonlinear polarization produced by a linearly polarized laser beam in an isotropic medium is always parallel to the \mathbf{E} vector of the laser.

- (11.11) A laser beam of wavelength $1.55 \mu\text{m}$ propagates down an optical fibre of length 10 m with a core diameter of $5 \mu\text{m}$. Calculate the power that must be launched into the fibre to produce a nonlinear phase shift of π , if the fibre has a nonlinear refractive index of $3 \times 10^{-20} \text{ m}^2 \text{ W}^{-1}$.

- (11.12)* A short laser pulse excites 10^{24} m^{-3} electron-hole pairs into a sample of GaAs at 4 K . Calculate the shift in the absorption edge at the band gap due to the presence of the photoexcited carriers. The band parameters for GaAs are given in Table C.2. Ignore excitonic effects.

- (11.13)* The $n = 1$ heavy hole exciton of a GaAs multiple quantum well sample has a peak absorption coefficient of $8 \times 10^5 \text{ m}^{-1}$ at 847 nm . By assuming that the exciton behaves like a classical dipole oscillator, estimate the change of the refractive index that can be obtained from this sample by completely saturating the absorption. The non-resonant refractive index is 3.5.

- (11.14) Calculate the saturation density for excitons in InP. If the carrier recombination time is 1 ns , and the absorption coefficient is 10^6 m^{-1} , estimate the saturation intensity when a laser is tuned to the exciton wavelength. The relevant band structure parameters for InP are given in Table C.2, and the dielectric constant is 12.5.

Electromagnetism in dielectrics

Appendix A

This appendix summarizes the principal results of electromagnetism that are used throughout the book. It is hoped that the reader will be familiar with this material. The main purpose of the appendix is to collect together the equations in a concise form for quick reference, and also to define the notation. SI units are used throughout. A short bibliography of suitable supplementary texts is given under Further Reading.

A.1 Electromagnetic fields and Maxwell's equations

The response of a dielectric material to an external electric field is characterized by three macroscopic vectors:

- the electric field strength \mathbf{E} ;
- the polarization \mathbf{P} ;
- the electric displacement \mathbf{D} .

The microscopic response of the material is determined primarily by the polarization. For this reason, the first task in all the examples treated by electromagnetism in this book is to calculate \mathbf{P} . The dielectric constant ϵ_r is then determined from \mathbf{P} , and the optical properties are deduced from ϵ_r .

The polarization is defined as the net dipole moment per unit volume. The application of a field produces a polarization by the forces exerted on the positive and negative charges of the atoms that are contained within the medium. If the atoms have permanent dipole moments, the field will apply a torque to these randomly orientated dipoles and tend to align them along the field direction. If there are no permanent dipoles, the field will push the positive and negative charges in opposite directions and induce a dipole parallel to the field. In either case, the end result is the same: the application of the field tends to produce many microscopic dipoles aligned parallel to the direction of the external field. This generates a net dipole moment within the dielectric, and hence a polarization.

The microscopic dipoles will all tend to align along the field direction, and so the polarization vector will be parallel to \mathbf{E} . This allows us to write:

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E}, \quad (\text{A.1})$$

where ϵ_0 is the electric permittivity of free space and χ is the electric susceptibility of the medium.

$\epsilon_0 = 8.854 \times 10^{-12} \text{ F m}^{-1}$ in SI units.

A.1 Electromagnetic fields and Maxwell's equations	255
A.2 Electromagnetic waves	258

Equation A.1 makes two assumptions that need a brief word of explanation.

- (1) We have assumed that the medium is isotropic, even though we know that some materials are anisotropic. In particular, anisotropic crystals have preferred non-equivalent axes, and \mathbf{P} will not necessarily be parallel to \mathbf{E} .
- (2) We have assumed that \mathbf{P} varies linearly with \mathbf{E} . This will not always be the case. In particular, if the optical intensity is very large, we can enter the realm of nonlinear optics, in which eqn A.1 is not valid.

Both of these qualifications introduce unnecessary complications at this stage, and are not considered further in this appendix. A discussion of how to treat non-isotropic materials may be found in Section 2.4, while nonlinear optics is the subject of Chapter 11.

The electric displacement \mathbf{D} of the medium is related to the electric field \mathbf{E} and polarization \mathbf{P} through:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}. \quad (\text{A.2})$$

This may be considered to be the definition of \mathbf{D} . By combining eqns A.1 and A.2, we can write:

$$\mathbf{D} = \epsilon_0 \epsilon_r \mathbf{E}, \quad (\text{A.3})$$

where

$$\epsilon_r = 1 + \chi. \quad (\text{A.4})$$

ϵ_r is the **relative dielectric constant** of the medium, and is an extremely important parameter in the understanding of the propagation of light through dielectrics.

In electrostatic problems we are frequently interested in calculating the spatial dependence of electric field, and hence the electric potential V , from the free charge density ϱ . This calculation can be performed by using the Poisson equation:

$$\nabla^2 V = -\frac{\varrho}{\epsilon_r \epsilon_0}. \quad (\text{A.5})$$

Poisson's equation is derived from Gauss's law of electrostatics:

$$\nabla \cdot \mathbf{E} = \frac{\varrho}{\epsilon_r \epsilon_0}. \quad (\text{A.6})$$

We recall that the electric field strength is the gradient of the potential:

$$\mathbf{E} = -\nabla V. \quad (\text{A.7})$$

Equation A.5 follows directly by substituting for \mathbf{E} in eqn A.6 using eqn A.7. Once we know V , we can then calculate \mathbf{E} from eqn A.7. This approach is also useful when we are treating devices in which the potential across the device is fixed by an external voltage source.

The response of a material to external magnetic fields is treated in a similar way to the response of dielectrics to electric fields. The magnetization \mathbf{M} of the medium is proportional to the magnetic field strength \mathbf{H} through the **magnetic susceptibility** χ_M :

$$\mathbf{M} = \chi_M \mathbf{H}. \quad (\text{A.8})$$

The magnetic flux density \mathbf{B} is related to \mathbf{H} and \mathbf{M} through:

$$\begin{aligned}\mathbf{B} &= \mu_0(\mathbf{H} + \mathbf{M}) \\ &= \mu_0(1 + \chi_M)\mathbf{H} \\ &= \mu_0\mu_r\mathbf{H},\end{aligned}\quad (\text{A.9})$$

where μ_0 is the magnetic permeability of the vacuum and $\mu_r = 1 + \chi_M$ is the relative magnetic permeability of the medium.

$\mu_0 = 4\pi \times 10^{-7} \text{ H m}^{-1}$ in SI units.

The laws that describe the combined electric and magnetic response of a medium are summarized in **Maxwell's equations** of electromagnetism:

$$\nabla \cdot \mathbf{D} = \varrho \quad (\text{A.10})$$

$$\nabla \cdot \mathbf{B} = 0 \quad (\text{A.11})$$

$$\nabla \wedge \boldsymbol{\mathcal{E}} = -\frac{\partial \mathbf{B}}{\partial t} \quad (\text{A.12})$$

$$\nabla \wedge \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, \quad (\text{A.13})$$

where ϱ is the free charge density, and \mathbf{j} is the current density. The first of these four equations is Gauss's law of electrostatics (eqn A.6) written in terms of \mathbf{D} rather than $\boldsymbol{\mathcal{E}}$. The second is the equivalent of Gauss's law for magnetostatics with the assumption that free magnetic monopoles do not exist. The third equation combines the Faraday and Lenz laws of electromagnetic induction. The fourth is a statement of Ampere's law, with the second term on the right hand side to account for the displacement current.

The second Maxwell equation naturally leads to the concept of the **vector potential**. This is defined through the equation

$$\mathbf{B} = \nabla \wedge \mathbf{A}. \quad (\text{A.14})$$

We see that the vector potential \mathbf{A} automatically satisfies eqn A.11, because $\nabla \cdot (\nabla \wedge \mathbf{A}) = 0$ for all \mathbf{A} . However, this definition does not define \mathbf{A} uniquely. We can add any vector of the form $\nabla\varphi$ to \mathbf{A} without changing \mathbf{B} . This follows because

$$\nabla \wedge (\mathbf{A} + \nabla\varphi) = \nabla \wedge \mathbf{A} + \nabla \wedge (\nabla\varphi) = \nabla \wedge \mathbf{A}. \quad (\text{A.15})$$

$\varphi(\mathbf{r})$ can be any scalar function of \mathbf{r} . For this reason, we have to give an additional definition, which specifies the **gauge** in which we are working. The **Coulomb gauge** is defined by

$$\nabla \cdot \mathbf{A} = 0. \quad (\text{A.16})$$

This gauge is a convenient one because it allows us to derive a simple relationship between $\boldsymbol{\mathcal{E}}$ and \mathbf{A} . By substituting for \mathbf{B} in the third Maxwell equation (eqn A.12) using eqn A.14, we see that:

$$\nabla \wedge \boldsymbol{\mathcal{E}} = -\frac{\partial}{\partial t}(\nabla \wedge \mathbf{A}) = \nabla \wedge \left(-\frac{\partial \mathbf{A}}{\partial t}\right). \quad (\text{A.17})$$

The solution is

$$\boldsymbol{\mathcal{E}} = -\frac{\partial \mathbf{A}}{\partial t} + \text{constant}, \quad (\text{A.18})$$

where the constant is any vector whose curl is zero. If the electrostatic potential is V , then we can combine eqn A.18 with eqn A.7 by writing

$$\boldsymbol{\epsilon} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V. \quad (\text{A.19})$$

This works because $\nabla \wedge \nabla V = 0$. By taking the divergence of eqn A.19, we can recover Poisson's equation (A.5) if we satisfy eqn A.16, that is, if we are in the Coulomb gauge. The more general definition of $\boldsymbol{\epsilon}$ given in eqn A.19 reduces to eqn A.7 when the magnetic field does not vary with time, and to

$$\boldsymbol{\epsilon} = -\frac{\partial \mathbf{A}}{\partial t} \quad (\text{A.20})$$

when the static potential is constant throughout space. The vector potential in the Coulomb gauge is used in the semiclassical treatment of the interaction of light with atoms discussed in Section B.2 of Appendix B.

A.2 Electromagnetic waves

Maxwell was able to show that eqns A.10–A.13 were consistent with wave-like solutions in a medium with no free charges or currents. To see this we first simplify eqns A.12 and A.13 by setting $\mathbf{j} = 0$ and eliminating \mathbf{B} and \mathbf{D} using eqns A.3 and A.9. This gives:

$$\nabla \wedge \boldsymbol{\epsilon} = -\mu_0 \mu_r \frac{\partial \mathbf{H}}{\partial t}, \quad (\text{A.21})$$

and

$$\nabla \wedge \mathbf{H} = \epsilon_0 \epsilon_r \frac{\partial \boldsymbol{\epsilon}}{\partial t}. \quad (\text{A.22})$$

We then take the curl of eqn A.21 and eliminate $\nabla \wedge \mathbf{H}$ using eqn A.22. This gives:

$$\nabla \wedge (\nabla \wedge \boldsymbol{\epsilon}) = -\mu_0 \mu_r \epsilon_0 \epsilon_r \frac{\partial^2 \boldsymbol{\epsilon}}{\partial t^2}. \quad (\text{A.23})$$

The left hand side can be simplified by using the vector identity

$$\nabla \wedge (\nabla \wedge \boldsymbol{\epsilon}) = \nabla(\nabla \cdot \boldsymbol{\epsilon}) - \nabla^2 \boldsymbol{\epsilon}. \quad (\text{A.24})$$

Equation A.6 with $\varrho = 0$ tells us that $\nabla \cdot \boldsymbol{\epsilon} = 0$. Therefore we obtain the final result:

$$\nabla^2 \boldsymbol{\epsilon} = \mu_0 \mu_r \epsilon_0 \epsilon_r \frac{\partial^2 \boldsymbol{\epsilon}}{\partial t^2}. \quad (\text{A.25})$$

Equation A.25 is of the same form as the wave equation:

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 y}{\partial t^2}, \quad (\text{A.26})$$

where v is the velocity of the wave. We therefore identify eqn A.25 as describing electromagnetic waves with a phase velocity v given by

$$\frac{1}{v^2} = \mu_0 \mu_r \epsilon_0 \epsilon_r. \quad (\text{A.27})$$

In free space $\epsilon_r = \mu_r = 1$ and the velocity of the wave is c , so we have:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 2.998 \times 10^8 \text{ m s}^{-1}. \quad (\text{A.28})$$

At the same time, we see from eqns A.27 and A.28 that the velocity in a medium is given by

$$v = \frac{1}{\sqrt{\epsilon_r \mu_r}} c. \quad (\text{A.29})$$

We define the **refractive index** n of the medium as the ratio of the velocity of light in free space to the velocity in the medium:

$$n = \frac{c}{v}. \quad (\text{A.30})$$

At optical frequencies we can set $\mu_r = 1$, and thus conclude:

$$n = \sqrt{\epsilon_r}. \quad (\text{A.31})$$

This allows us to relate the propagation constants of electromagnetic waves in a medium to the dielectric constant.

The solutions to eqn A.25 are of the form

$$\mathbf{E}(z, t) = \mathbf{E}_0 e^{i(kz - \omega t)}, \quad (\text{A.32})$$

where \mathbf{E}_0 is the amplitude of the wave, z is the direction of propagation, k is the wave vector, and ω is the angular frequency. The wave vector k is given by

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{v} = \frac{n\omega}{c}, \quad (\text{A.33})$$

where λ is the wavelength inside the medium. The first equality is the definition of k , the second follows by substitution of eqn A.32 into eqn A.25 with v given by eqn A.27, and the third follows from the definition of n given in eqn A.30.

The energy flow in an electromagnetic wave can be calculated from the **Poynting vector**:

$$\mathbf{I} = \mathbf{E} \wedge \mathbf{H}. \quad (\text{A.34})$$

This gives the power flow per unit area in W m^{-2} , which is equal to the **intensity** of the light wave. The intensity is defined as the energy crossing a unit area in unit time, and is therefore given by:

$$I = vu_v, \quad (\text{A.35})$$

where v is the velocity of the wave and u_v is the energy density per unit volume of the beam.

The Poynting vector given by eqn A.34 can be evaluated easily for the case of plane waves. Consider a wave polarized along the x axis of angular frequency ω propagating in the z direction. From eqn A.21 or A.22 we see that the magnetic field is perpendicular to the electric field. \mathbf{E} and \mathbf{H} therefore form a right handed system as depicted in Fig. A.1. Hence the components of the wave must satisfy:

$$\begin{aligned} \mathcal{E}_x(z, t) &= \mathcal{E}_{x0} e^{i(kz - \omega t)} \\ \mathcal{E}_y(z, t) &= 0 \\ H_x(z, t) &= 0 \\ H_y(z, t) &= H_{y0} e^{i(kz - \omega t)}. \end{aligned} \quad (\text{A.36})$$

The use of complex solutions of the type given in Eq. A.32 simplifies the mathematics and is used extensively throughout this book. Physically measurable quantities are obtained by taking the real part of the complex wave.

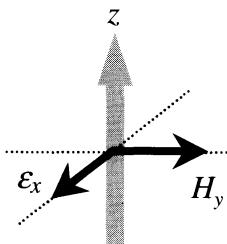


Fig. A.1 The electric and magnetic fields of an electromagnetic wave form a right handed system. The figure shows the directions of the fields in a wave polarized along the x axis and propagating in the z direction.

Z has a value of 377Ω in free space.

On substituting these fields into eqn A.21, we find that:

$$k \mathcal{E}_{x0} = \mu_0 \mu_r \omega H_{y0}, \quad (\text{A.37})$$

and hence that

$$H_{y0} = \frac{\mathcal{E}_{x0}}{Z} \quad (\text{A.38})$$

where

$$Z = \frac{k}{\mu_0 \mu_r \omega} = \sqrt{\frac{\mu_0 \mu_r}{\epsilon_0 \epsilon_r}} = \frac{1}{c \epsilon_0 n}. \quad (\text{A.39})$$

The second equality in eqn A.39 follows from eqns A.33 and A.27, while the third follows from Eq. A.28 and A.31 with $\mu_r = 1$. The quantity Z is called the **wave impedance**. On substituting eqns A.36–A.39 into eqn A.34, we obtain:

$$I = \frac{\langle \mathcal{E}(t)^2 \rangle_{\text{rms}}}{Z} = \frac{1}{2} c \epsilon_0 n \mathcal{E}_0^2 \quad (\text{A.40})$$

where $\langle \mathcal{E}(t)^2 \rangle_{\text{rms}}$ represents the root-mean-square time average. This shows that the intensity of a light wave is proportional to the square of the amplitude of the electric field. The relationship can be generalized for all light waves irrespective of the particular polarization of the beam.

In many topics covered in this book, it will be necessary to treat the refractive index as a complex number. A well-known example of how such a situation arises occurs when treating the propagation of electromagnetic waves through a conducting medium such as a metal. In a conductor, the current density is related to the electric field through the electrical conductivity σ according to:

$$\mathbf{j} = \sigma \mathbf{E}. \quad (\text{A.41})$$

Using this relationship to substitute for \mathbf{j} in eqn A.13, and eliminating \mathbf{D} , \mathbf{B} and \mathbf{H} in the same way that led to eqn A.25, we obtain:

$$\nabla^2 \mathbf{E} = \sigma \mu_0 \mu_r \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mu_r \epsilon_0 \epsilon_r \frac{\partial^2 \mathbf{E}}{\partial t^2}. \quad (\text{A.42})$$

We now look for plane wave solutions of the type given by eqn A.32. Substitution of eqn A.32 into eqn A.42 gives:

$$k^2 = i \sigma \mu_0 \mu_r \omega + \mu_0 \mu_r \epsilon_0 \epsilon_r \omega^2. \quad (\text{A.43})$$

This can be made compatible with the usual relationship between ω and k given in eqn A.33 by allowing n to be a complex number. The complex refractive index is usually written \tilde{n} , and is defined by

$$k = \tilde{n} \frac{\omega}{c}. \quad (\text{A.44})$$

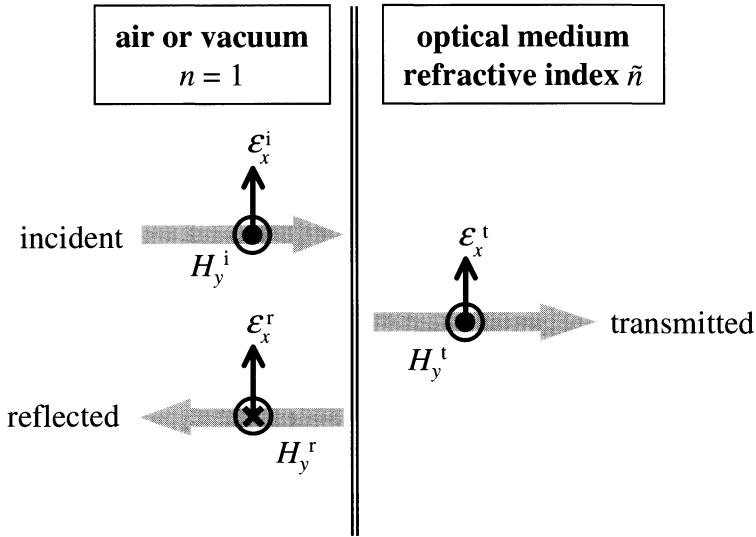


Fig. A.2 Transmission and reflection of light at an interface between air and a medium of refractive index \tilde{n} . The incident, transmitted and reflected rays are shown displaced from each other for clarity. All rays are normal to the interface. The symbol \odot for the magnetic fields of the incident and transmitted rays indicates that the field direction is out of the page, while the symbol \otimes for the reflected wave indicates that the field is pointing in to the page.

By combining eqns A.43 and A.44 we obtain:

$$\tilde{n}^2 = \frac{\mu_r \sigma}{\epsilon_0 \omega} i + \mu_r \epsilon_r, \quad (\text{A.45})$$

where we have made use of eqn A.28. This of course reduces to eqn A.31 if we set $\sigma = 0$ and $\mu_r = 1$. The physical significance of the complex refractive index implied by eqn A.45 is developed in more detail in Section 1.3.

The Maxwell equations also allow us to treat the transmission and reflection of light at an interface between two materials. This situation is depicted in Fig. A.2. Part of the beam is transmitted into the medium and the rest is reflected. The solution for an arbitrary angle of incidence was treated by Fresnel, and the resulting formulæ are known as **Fresnel's equations**. We restrict ourselves here to the simpler case when the angle of incidence is zero: that is, normal incidence.

We consider again an x -polarized light beam propagating in the z direction, with the field directions as shown in Fig. A.1. The electric and magnetic fields are given by eqn A.36. The beam is incident on a medium with complex refractive index \tilde{n} . The fields are related to each other through eqn A.38, with Z given by eqn A.39, although we now have to allow for the possibility that n may be complex.

The boundary conditions at the interface between two dielectrics tell us that the tangential components of the electric and magnetic fields are conserved. Applying this to the situation shown in Fig. A.2, we must have that both \mathcal{E}_x and H_y are conserved across the boundary. Hence we can write:

$$\mathcal{E}_x^i + \mathcal{E}_x^r = \mathcal{E}_x^t, \quad (\text{A.46})$$

and

$$H_y^i - H_y^r = H_y^t, \quad (\text{A.47})$$

where the superscripts i , r and t refer to the incident, reflected and transmitted beams respectively. By making use of the relationship between the magnetic

It is shown in Chapter 7 that eqn A.45 is only valid at low frequencies. This is because the AC conductivity at high frequencies is not the same as the DC conductivity that enters eqn A.41.

and electric fields given in eqns A.38–A.39, we can rewrite eqn A.47 as:

$$\mathcal{E}_x^i - \mathcal{E}_x^r = \tilde{n} \mathcal{E}_x^t, \quad (\text{A.48})$$

where we have assumed that the light is incident from air with $\tilde{n} = 1$ and that $\mu_r = 1$ at the optical frequencies of interest here. Equations A.46 and A.48 can be solved together to obtain

$$\frac{\mathcal{E}_x^r}{\mathcal{E}_x^i} = \frac{\tilde{n} - 1}{\tilde{n} + 1}. \quad (\text{A.49})$$

This can be rearranged to obtain the required result for the reflectivity R :

$$R = \left| \frac{\mathcal{E}_x^r}{\mathcal{E}_x^i} \right|^2 = \left| \frac{\tilde{n} - 1}{\tilde{n} + 1} \right|^2. \quad (\text{A.50})$$

This formula is used in many examples throughout the book.

Further reading

The subject matter of this appendix is standard electromagnetism, and there are numerous books on the market that cover the material, for example: Bleaney and Bleaney (1976), Duffin (1990), Good (1999), Grant and Phillips (1990), Lorrain, Corson and Lorrain (1988). The subject is also covered in many optics text books such as Hecht (1998) or Born and Wolf (1999).

Quantum theory of radiative absorption and emission

Appendix

B

The discussion of the absorption and emission of light by atoms throughout this book presupposes that the reader is familiar with the basic treatment of these processes found in all introductory quantum physics texts. The purpose of this appendix is to give a brief summary of the main results. We begin by discussing the Einstein coefficients in order to introduce the concepts of absorption and emission, and the connection between them. We then move on to discuss how the absorption and emission rates can be calculated using quantum mechanics. Finally, we briefly discuss the selection rules that apply to radiative transitions. The reader is encouraged to refer to a quantum physics textbook if any of the concepts are unfamiliar.

B.1 Einstein coefficients

The quantum theory of radiation assumes that light is emitted or absorbed whenever an atom makes a jump between two quantum states. These two processes are illustrated in Fig. B.1. Absorption occurs when the atom jumps to a higher level, while emission corresponds to the process in which a photon is emitted as the atom drops down to a lower level. It is customary to label the upper state as level ‘2’ and the lower one as level ‘1’. Conservation of energy requires that the frequency ν of the photon satisfies:

$$h\nu = E_2 - E_1, \quad (\text{B.1})$$

where E_2 is the energy of level 2 and E_1 is the energy of level 1.

Statistical physics tells us that atoms in excited states have a natural tendency to de-excite and lose their excess energy. Thus the emission of a photon by an atom in an excited state is a spontaneous process. The radiation of light by atoms in excited states is therefore called **spontaneous emission**. This process is illustrated in Fig. B.1(a). One of the electrons in the atom is in level 2 at the start of the process and drops to level 1 by emitting a photon. The frequency of the photon corresponds to the energy difference of the two levels according to eqn B.1. Hence each type of atom has a characteristic emission spectrum determined by its energy levels.

The process of **absorption** is illustrated in Fig. B.1(b). The atom is promoted to an excited state by absorbing the required energy from a photon. This promotes an electron from level 1 to level 2. Unlike emission, it is not a spontaneous process. The electron cannot jump to the excited state unless it is stimulated by an incoming photon.

In Section B.2 below we will explain how quantum mechanics enables us to calculate the spontaneous emission and absorption rates. At this stage we

B.1	Einstein coefficients	263
B.2	Quantum transition rates	266
B.3	Selection rules	269

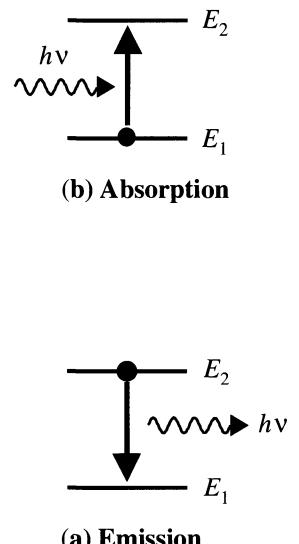


Fig. B.1 Optical transitions in an atom: (a) emission, (b) absorption.

restrict ourselves to a simpler analysis based on the **Einstein coefficients** for the transition.

Spontaneous emission is governed by the Einstein *A* coefficient. This gives the probability per unit time that the electron in level 2 will drop to level 1 by emitting a photon. The photon emission rate is therefore proportional to the number of atoms in the excited state and to the *A* coefficient for the transition. We thus write down the following rate equation for $N_2(t)$, the number of atoms in the excited state:

$$\frac{dN_2}{dt} = -A_{21}N_2. \quad (\text{B.2})$$

The subscript ‘21’ on the *A* coefficient in eqn B.2 makes it plain that the transition starts at level 2 and ends at level 1.

Equation B.2 can be solved for $N_2(t)$ to give:

$$\begin{aligned} N_2(t) &= N_2(0) \exp(-A_{21}t) \\ &= N_2(0) \exp(-t/\tau), \end{aligned} \quad (\text{B.3})$$

where

$$\tau = \frac{1}{A_{21}}. \quad (\text{B.4})$$

τ is the **natural radiative lifetime** of the excited state. Equation B.4 shows that the number of the atoms in the excited state decays exponentially with a time constant τ due to spontaneous emission. The value of τ for a transition can range from about 1 ns to several milliseconds. The selection rules that govern whether a particular transition is fast or slow are discussed in Section B.3 below.

The absorption rate between levels 1 and 2 is governed by the Einstein *B* coefficient. As mentioned above, this process must be stimulated by the incoming photon field. Following Einstein’s treatment, we write the rate of absorption transitions per unit time as:

$$\frac{dN_1}{dt} = -B_{12}N_1u(\nu), \quad (\text{B.5})$$

where $N_1(t)$ is the number of atoms in level 1 at time t , B_{12} is the Einstein *B* coefficient for the transition, and $u(\nu)$ is the energy density of the electromagnetic wave in J m^{-3} at frequency ν . By writing $u(\nu)$ we are explicitly stating that only the part of the spectrum of the incoming radiation at frequency ν , where $h\nu = E_2 - E_1$, can induce the absorption transitions. Equation B.5 may be considered to be the definition of the Einstein *B* coefficient.

The processes of absorption and spontaneous emission that we have described above are fairly intuitive. Einstein realized that the analysis was not complete, and introduced a third type of transition called **stimulated emission**. In this process, the incoming photon field can stimulate downward emission transitions as well as upward absorption transitions. The stimulated emission rate is governed by a second Einstein *B* coefficient, namely B_{21} . The subscript is now essential to distinguish the *B* coefficients for the two distinct processes of absorption and stimulated emission.

In analogy with eqn B.5, we write the rate of stimulated emission transitions by the following rate equation:

$$\frac{dN_2}{dt} = -B_{21}N_2u(\nu). \quad (\text{B.6})$$

In the full quantum treatment of radiative emission, the photon field is quantized, with a harmonic oscillator energy spectrum given by

$$E_n = (n + \frac{1}{2})h\nu,$$

n being the number of photons. The factor of $\frac{1}{2}$ corresponds to the zero-point fluctuations of the electromagnetic field. Spontaneous emission is then considered to be a stimulated emission process instigated by the ever-present zero-point fluctuations of the electromagnetic field.

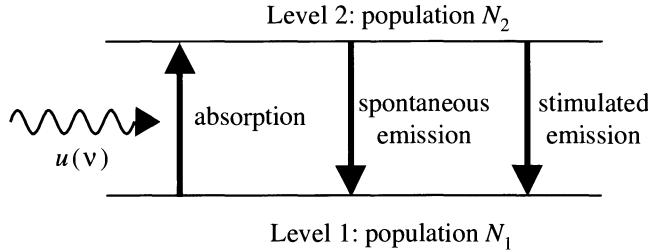


Fig. B.2 Absorption, spontaneous emission and stimulated emission transitions between two levels of an atom in the presence of electromagnetic radiation with energy density $u(v)$.

Stimulated emission is a coherent quantum mechanical effect. The photons emitted are in phase with the photons that induce the transition. Stimulated emission is the basis of laser operation, as the name suggests: ‘laser’ stands for ‘light amplification by stimulated emission of radiation’.

The three Einstein coefficients introduced above are not independent parameters: they are all related to each other. If we know one of them, we can work out the other two. To see how this works, we follow Einstein’s analysis.

We imagine that the atom is inside a box at temperature T with black walls. The atom will then be bathed in black body radiation. The black body radiation will induce both absorption and stimulated emission transitions, while spontaneous emission transitions will also be occurring at a rate determined by the Einstein A coefficient. The three types of transition are indicated in Fig. B.2. If we leave the atom for long enough, it will come to thermal equilibrium with the black body radiation. In these steady state conditions, the rate of upward transitions due to absorption must exactly balance the rate of downward transitions due to spontaneous and stimulated emission. From eqns B.2, B.5 and B.6, we must therefore have:

$$B_{12}N_1u(v) = A_{21}N_2 + B_{21}N_2u(v). \quad (\text{B.7})$$

Since the atoms are in thermal equilibrium with the radiation field at temperature T , the distribution of the atoms among the various energy levels will be governed by the laws of thermal physics. The ratio of N_2 to N_1 will therefore be given by Boltzmann’s law:

$$\frac{N_2}{N_1} = \frac{g_2}{g_1} \exp\left(-\frac{h\nu}{k_B T}\right), \quad (\text{B.8})$$

where g_1 and g_2 are the degeneracies of levels 1 and 2 respectively. Now the energy spectrum of a black body source is given by the Planck formula:

$$u(v) = \frac{8\pi h\nu^3}{c^3} \frac{1}{\exp(h\nu/k_B T) - 1}. \quad (\text{B.9})$$

The only way that eqns B.7–B.9 can be consistent with each other is if:

$$g_1 B_{12} = g_2 B_{21}, \quad (\text{B.10})$$

and

$$A_{21} = \frac{8\pi h\nu^3}{c^3} B_{21}. \quad (\text{B.11})$$

If the atom is embedded within an optical medium with a refractive index n , we replace c by c/n in eqn B.11 to account for the reduced velocity of light within the medium.

A moment's thought will convince us that it is not possible to get consistency between the equations without the stimulated emission term. This is what led Einstein to introduce the concept. Equation B.10 tells us that the probabilities for stimulated absorption and emission are the same apart from the degeneracy factors. Furthermore, the interrelationship of the Einstein coefficients tells us that transitions that have a high absorption probability will also have a high emission probability, both for spontaneous processes and stimulated ones.

The relationships between the Einstein coefficients given in eqns B.10 and B.11 have been derived for the case of an atom in equilibrium with black body radiation. However, once we have derived the inter-relationships, they will apply in all other cases as well. This is very useful, because we then only need to know one of the coefficients to work out the other two. For example, we can measure the radiative lifetime to determine A_{21} using eqn B.4, and then work out the B coefficients using eqns B.11 and B.10.

B.2 Quantum transition rates

The calculation of radiative transition rates by quantum mechanics is based on time-dependent perturbation theory. The light–matter interaction is described by transition probabilities, which can be calculated using **Fermi's golden rule**.

Fermi's golden rule gives the transition rate between levels 1 and 2 in terms of the **matrix element** M_{12} . Referring to the absorption process illustrated in Fig. B.1(b), we write the rate of transitions as

$$W_{1 \rightarrow 2} = \frac{2\pi}{\hbar} |M_{12}|^2 \delta(E_2 - E_1 - h\nu), \quad (\text{B.12})$$

or

$$W_{1 \rightarrow 2} = \frac{2\pi}{\hbar} |M_{12}|^2 g(h\nu). \quad (\text{B.13})$$

The first form is used for the transitions between discrete levels such as those found in individual atoms. The conservation of energy in such a process is contained explicitly in the Dirac delta function. The second form is used for transitions to a continuous bands of levels. This situation is the usual one in solids, and also occurs in atoms when considering transitions to the continuum of levels above the ionization threshold.

The significance of the **density of states** factor $g(h\nu)$ in eqn B.13 is illustrated in Fig. B.3. This diagram shows a transition from a discrete level to a continuum of levels. $g(h\nu)dE$ is the number of states that fall within the energy range E to $E + dE$, where $E = h\nu$. Transitions are possible to any of the states that fall within this energy range, and eqn B.13 is just the integral of eqn B.12 over the many levels that can participate. The density of states factor is very important in solid state problems, and will be discussed in detail wherever it occurs in the main text. It will frequently be the case that both the initial and final levels will be continuous bands. In this case, the density of states that has to be used in eqn B.13 is a suitably weighted joint density of states for both the initial and final bands.

The quantum optical treatment of the light–matter interaction tells us that the atoms oscillate between the initial and final states while the light field is present. This effect is called Rabi oscillations. We can reconcile this picture with the notion of discrete transitions by realizing that the oscillations can only be observed if the atoms do not scatter in any way. If scattering does occur, the coherence of the oscillations will be broken, and the atom will end up either in the initial or final state, depending on where it is when the scattering event occurs. This then gives the probability that the atom makes a transition to the final state due to the interaction with the light. In the systems we are studying in this book it will always be the case that the Rabi oscillations are heavily damped. Rabi oscillations at optical frequencies can only be observed in special conditions and do not concern us here.

The matrix element that appears in Fermi's golden rule can be written in the compact Dirac notation or explicitly in terms of the overlap integral as:

$$M_{12} = \langle 2 | H' | 1 \rangle = \int \psi_2^*(\mathbf{r}) H'(\mathbf{r}) \psi_1(\mathbf{r}) d^3\mathbf{r}, \quad (\text{B.14})$$

where H' is the perturbation caused by the light wave, \mathbf{r} is the position vector of the electron, and $\psi_1(\mathbf{r})$ and $\psi_2(\mathbf{r})$ are the wave functions of the initial and final states. To evaluate M_{12} , we need to know the wave functions of the states, and also the form of the perturbation due to the light wave.

The perturbation due to the light can be evaluated by calculating the effect of the electromagnetic field on the electron in the atom. From classical electromagnetism we know that the field changes the momentum of a charged particle from \mathbf{p} to $(\mathbf{p} - q\mathbf{A})$, where q is the charge and \mathbf{A} is the vector potential defined in eqn A.14 of Appendix A. The Hamiltonian for an electron with $q = -e$ in an electromagnetic field is therefore:

$$\begin{aligned} H &= \frac{1}{2m_0}(\mathbf{p} + e\mathbf{A})^2 + V(\mathbf{r}) \\ &= H_0 + \frac{e}{2m_0}(\mathbf{p} \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}) + \frac{e^2\mathbf{A}^2}{2m_0}, \end{aligned} \quad (\text{B.15})$$

where H_0 represents the Hamiltonian of the electron before the field was applied. It is given by:

$$H_0 = \frac{\mathbf{p}^2}{2m_0} + V(\mathbf{r}), \quad (\text{B.16})$$

where $V(\mathbf{r})$ is the potential energy of the electron in the atom. The perturbation H' due to the light field is thus:

$$H' = \frac{e}{2m_0}(\mathbf{p} \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}) + \frac{e^2\mathbf{A}^2}{2m_0}. \quad (\text{B.17})$$

The interaction Hamiltonian given by eqn B.17 can be simplified in two ways. Firstly, we can neglect the term in \mathbf{A}^2 compared to the other two terms on the right hand side because it is much smaller. Secondly, we can group together the first two terms of H' because the two operators commute. This occurs here because $\nabla \cdot \mathbf{A} = 0$ in the Coulomb gauge (cf. eqn A.16), and therefore if $\mathbf{p} = -i\hbar\nabla$, it follows that $\mathbf{p} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{p}$.

With these two simplifications, we can write the perturbation due to the light field as:

$$H' = \frac{e}{m_0}\mathbf{p} \cdot \mathbf{A}. \quad (\text{B.18})$$

This is now in a form that can be evaluated for the vector potential of an electromagnetic wave.

If the electric and magnetic fields \mathbf{E} and \mathbf{B} of the light wave both vary in time and space as $\exp i(\mathbf{k} \cdot \mathbf{r} - \omega t)$ with $\omega = ck$, it follows from eqns A.14 and A.20 that \mathbf{A} must also do so. Hence we may write:

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 \exp i(\mathbf{k} \cdot \mathbf{r} - \omega t), \quad (\text{B.19})$$

where from eqn A.20 we see that \mathbf{A}_0 must be pointing in the same direction as \mathbf{E} , namely along the polarization direction.

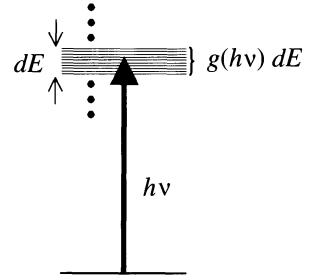


Fig. B.3 Optical transitions from a discrete level to a continuum of states.

The term in \mathbf{A}^2 corresponds to two-photon rather than one-photon interactions between the atom and the light field. These two-photon interactions are usually very weak, although they may become important with intense laser light fields.

The exponential in $\mathbf{k} \cdot \mathbf{r}$ that appears in eqn B.19 should be understood in terms of its Taylor expansion:

$$e^{i\mathbf{k} \cdot \mathbf{r}} = 1 + i\mathbf{k} \cdot \mathbf{r} + \frac{1}{2}(i\mathbf{k} \cdot \mathbf{r})^2 + \dots \quad (\text{B.20})$$

At optical frequencies the wavelength is around $1 \mu\text{m}$, and the dimension of a typical atom is $\sim 10^{-10} \text{ m}$. Thus $|\mathbf{k} \cdot \mathbf{r}| \approx 2\pi|\mathbf{r}|/\lambda \sim 10^{-3}$. This means that we need only consider the first term in eqn B.20 and can take $\exp i\mathbf{k} \cdot \mathbf{r} \sim 1$ to a very good approximation. This is called the **electric dipole** approximation for reasons that will become clear below. In this approximation we thus need to evaluate matrix elements of the type:

$$M_{12} = \frac{e}{m_0} \langle 2 | \mathbf{p} \cdot \mathbf{A}_0 | 1 \rangle. \quad (\text{B.21})$$

Equation B.21 can be simplified further by remembering that $\mathbf{p} = m_0 d\mathbf{r}/dt$. In the interaction picture we can write the equation of motion of a time-dependent operator \hat{O} as:

$$\frac{d}{dt} \hat{O}(t) = \frac{i}{\hbar} [H_0, \hat{O}] = \frac{i}{\hbar} (H_0 \hat{O} - \hat{O} H_0). \quad (\text{B.22})$$

Therefore, we can write:

$$\begin{aligned} \langle 2 | \mathbf{p} | 1 \rangle &= m_0 \left\langle 2 \left| \frac{d\mathbf{r}}{dt} \right| 1 \right\rangle \\ &= \frac{im_0}{\hbar} \langle 2 | [H_0, \mathbf{r}] | 1 \rangle \\ &= \frac{im_0}{\hbar} (E_2 - E_1) \langle 2 | \mathbf{r} | 1 \rangle \\ &= im_0 \omega_{21} \langle 2 | \mathbf{r} | 1 \rangle, \end{aligned} \quad (\text{B.23})$$

where $\hbar\omega_{21}$ is the transition energy. The intermediate step in the third line of eqn B.23 follows because we know from the Schrödinger equation that $H_0 \Psi_i = E_i \Psi_i$. Since the \mathbf{A}_0 that appears in eqn B.21 is just a simple vector and not an operator, we can use eqn B.23 to rewrite eqn B.21 as follows:

$$M_{12} = ie\omega_{21} \langle 2 | \mathbf{r} \cdot \mathbf{A}_0 | 1 \rangle. \quad (\text{B.24})$$

Finally, we note from eqns B.19 and A.20 that $\mathbf{E}_0 = i\omega \mathbf{A}_0$. Therefore, if $\omega = \omega_{21}$ as we have been presuming all along, the electric dipole matrix element is given by:

$$M_{12} = \langle 2 | e\mathbf{r} \cdot \mathbf{E}_0 | 1 \rangle. \quad (\text{B.25})$$

On comparing eqns B.14 and B.25 we see that in the electric dipole approximation, the interaction Hamiltonian is just

$$H' = -\mathbf{p}_e \cdot \mathbf{E}_0, \quad (\text{B.26})$$

where $\mathbf{p}_e = -e\mathbf{r}$ is the electric dipole moment of the electron. This is in fact exactly equal to the interaction energy we would expect for a dipole in an electric field, and explains why the transitions are called ‘electric dipole’ transitions.

The time dependence of \mathbf{A} does not appear here because it has already been assumed that the perturbation has a time dependence of the form $e^{\pm i\omega t}$ in the derivation of Fermi’s golden rule.

The result given in eqn B.25 allows us to evaluate the matrix elements for particular transitions if the wave functions of the initial and final states are known. We can then use Fermi's golden rule to work out $W_{1 \rightarrow 2}$ from eqns B.12 or B.13 as appropriate. This gives us the absorption rate, which must be the same as $B_{12}u(v)$ in eqn B.5. The energy density $u(v)$ of an electromagnetic wave is proportional to \mathcal{E}_0^2 (cf. eqns A.35 and A.40), and we can thus work out B_{12} from $W_{1 \rightarrow 2}$. We can then obtain the corresponding A coefficient using eqn B.11.

The final result for transitions between non-degenerate discrete atomic levels by absorption or emission of unpolarized light is:

$$B_{12} = \frac{\pi e^2}{3\epsilon_0 \hbar^2} |\langle 2|\mathbf{r}|1\rangle|^2, \quad (\text{B.27})$$

and

$$A_{21} = \frac{e^2 \omega_{21}^3}{3\pi \epsilon_0 \hbar c^3} |\langle 2|\mathbf{r}|1\rangle|^2. \quad (\text{B.28})$$

When the levels are degenerate, we must modify eqns B.27 and B.28 to allow for the different transition pathways. For example, if we consider the transitions between atomic levels with quantum numbers j and i , each of which consists of a manifold of degenerate levels labelled by additional quantum numbers m_j and m_i , then eqn B.28 is modified to:

$$A_{ji} = \frac{e^2 \omega_{ji}^3}{3\pi \epsilon_0 \hbar c^3} \frac{1}{g_j} \sum_{m_j, m_i} |\langle j, m_j | \mathbf{r} | i, m_i \rangle|^2, \quad (\text{B.29})$$

where g_j is the degeneracy of the upper state. In solid state systems, the summation over discrete levels is replaced with integrals over bands with known densities of states.

The matrix element for a transition is directly proportional to the **oscillator strength** f_{ij} introduced in the classical treatment of absorption of light by atoms in Section 2.2.2. For transitions between non-degenerate levels, the relationship between them is given by:

$$f_{ij} = \frac{2m\omega_{ji}}{3\hbar} |\langle j | \mathbf{r} | i \rangle|^2. \quad (\text{B.30})$$

The oscillator strength was introduced before quantum theory was developed to explain how some atomic absorption and emission lines are stronger than others. With the hindsight of quantum mechanics, it is easy to understand that this is simply caused by the different matrix elements for the transitions.

B.3 Selection rules

The electric-dipole matrix element given in eqn B.25 can be easily evaluated for simple atoms with known wave functions. This leads to the notion of **selection rules**. These are rules about the quantum numbers of the initial and final states. If the states do not satisfy the selection rules, the electric-dipole transition rate will be zero. For an electron in a hydrogenic system with quantum numbers l , m and m_s , the electric-dipole selection rules are:

- (1) The parity of the initial and final states must be different.
- (2) $\Delta m = -1, 0$ or $+1$.
- (3) $\Delta l = \pm 1$.
- (4) $\Delta m_s = 0$.

These rules can be generalized to many-electron atoms. The parity selection rule always applies. Another useful general rule is that the photon carries one unit of angular momentum, and so the total angular momentum of the atom must change by one unit in the transition. Furthermore, the photon does not interact with the electron spin, and so the spin quantum numbers never change in the transition.

Transitions that obey the electric-dipole selection rules are called **allowed** transitions, while those which do not are called **forbidden** transitions. Dipole-allowed transitions have high transition rates, and therefore have short radiative lifetimes, typically in the range 10^{-9} – 10^{-8} s. This fast radiative emission process is called **fluorescence**.

When electric-dipole transitions are forbidden, other types of processes such as **magnetic dipole** or **electric quadrupole** transitions may be possible. These have different selection rules to electric dipole transitions and arise from the higher order terms that we neglected in eqn B.20. Since they are higher order processes, they have smaller transition rates and longer radiative lifetimes, with values of τ ranging from about 10^{-6} s upwards. The slow emission by electric dipole-forbidden transitions is called **phosphorescence**. Phosphorescence causes ‘delayed’ emission in which the atoms re-emit light a substantial time after they have been excited. This contrasts with fluorescence, where the re-emission is ‘prompt’ in the sense that it takes place within a nanosecond or so after the atom has been excited.

Further reading

This material is covered in most quantum mechanics textbooks. See, for example: Gasiorowicz (1996) or Schiff (1969). More detailed information on atomic selection rules and transition rates may be found in atomic physics texts such as Corney (1977) or Woodgate (1980).

Magnetic dipole and electric quadrupole transitions can take place between states of the same parity.

Appendix



Band theory

The electronic states of crystals are described by the band theory of solids. This subject is covered extensively in all solid state physics texts, and this book presupposes a working knowledge of the main concepts. The purpose of this appendix is to give a brief review of the key points, and to give an overview of the band structure of a few representative materials. The reader is referred to the texts listed under Further Reading for a more extensive treatment.

C.1 Metals, semiconductors and insulators

The general concept of band formation is described in Section 1.5.2. The outer orbitals of the atoms in a densely packed solid overlap with each other as the chemical bonds that hold the crystal together are formed. This causes the discrete energy levels of the free atoms to be broadened into bands, as shown schematically in Fig. 1.9. Each band can contain $2N$ electrons, where N is the number of primitive unit cells in the crystal. Electrons fill the bands up to the **Fermi energy** E_F , which is determined by the total electron density.

Figure C.1(a) shows a schematic energy diagram for a monovalent metal such as sodium, or a trivalent metal such as aluminium. These have an odd number of electrons per atom, which means that the highest occupied band will only be half full. The Fermi energy will therefore lie in the middle of the highest occupied band. Electrons with energy just below E_F can easily be excited to empty states just above E_F . This makes it easy to accelerate the electrons with an electric field, and explains why metals are good conductors of electricity.

Figure C.1(b) shows the equivalent energy level diagram for a semiconductor such as silicon or an insulator such as diamond. These have an even number of electrons per atom, and the highest occupied band is therefore full of electrons. This highest occupied band is called the **valence band**, while the lowest unoccupied band is called the **conduction band**. The Fermi level lies somewhere within the energy gap between the valence and conduction bands. The first empty states for the electrons are in the conduction band, and it requires a minimum amount of energy equal to the **band gap** E_g to excite the electrons to available states. This makes it very difficult to accelerate the electrons when an external electric field is applied, and inhibits the flow of electrical current through the sample. Semiconductors and insulators therefore have much lower electrical conductivities than metals.

The distinction between an insulator and a semiconductor is related to the size of the band gap. Semiconductors have smaller band gaps than insulators. This makes it possible that a significant number of electrons are excited ther-

C.1 Metals, semiconductors and insulators	271
C.2 The nearly free electron model	273
C.3 Example band structures	276

The situation in divalent metals such as magnesium is slightly more complicated. These have an even number of electrons per atom, and so it might be expected that they would behave as semiconductors or insulators. This does not happen because the valence and conduction bands overlap with each other. Electrons below the Fermi level can therefore be excited to empty levels without having to cross an energy gap. This makes them good conductors like the monovalent or trivalent metals.

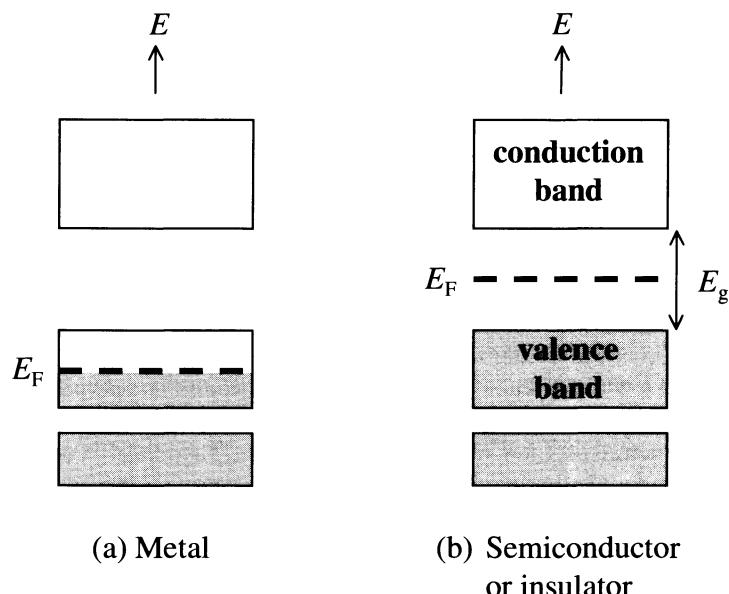


Fig. C.1 Energy level diagrams for (a) a monovalent or trivalent metal, and (b) a semiconductor or insulator. The bands are filled with electrons up to the Fermi level E_F . This is indicated by the shading.

mally from the valence band to the conduction band at room temperature. The free electrons in the conduction band can conduct electricity easily in the same manner to the free electrons in metals. Semiconductors therefore have a higher conductivity than insulators, but a smaller conductivity than metals because of the small number of free electrons.

Figure C.2 shows the conduction and valence bands of a semiconductor in more detail. Figure C.2(a) applies to a pure crystal. This is the same as the case considered in Fig. C.1(b), and the Fermi level lies about half way between the two bands. The thermal excitation of an electron to the conduction band leaves an empty state called a **hole** in the valence band. The hole is equivalent to the absence of an electron, and behaves like a free positive charge. Electrical current can be carried both by the electrons in the conduction band and by the holes in the valence band. The conduction by the thermally excited electrons and holes in a pure crystal is called **intrinsic**.

Figure C.2(b) applies to a crystal with **n-type** doping. In this case, impurities with extra electrons are deliberately introduced into the crystal. This is typically done by adding atoms from group V of the periodic table into the crystal while it is being grown. Semiconductor crystals such as silicon and germanium come from group IV in the periodic table, and have four valence electrons per atom. The impurities therefore donate one extra electron for each dopant atom. These extra electrons lie in **donor** levels just below the conduction band. The electron states are filled up to the donor levels, and so the Fermi energy must lie very close to the donor level energies, as indicated in Fig. C.2(b). The electrons in the donor levels can be easily excited into the conduction band at room temperature, and the electrical properties of the n-type material are determined by these **extrinsic** electrons arising from the impurity atoms.

Figure C.2(c) applies to a crystal with **p-type** doping. In this case, group III atoms are doped into the crystal during its growth. These atoms have a deficit of

Compound semiconductors such as the III–V or II–VI materials also have a valency of four because eight electrons are shared in the bond between the two atoms.

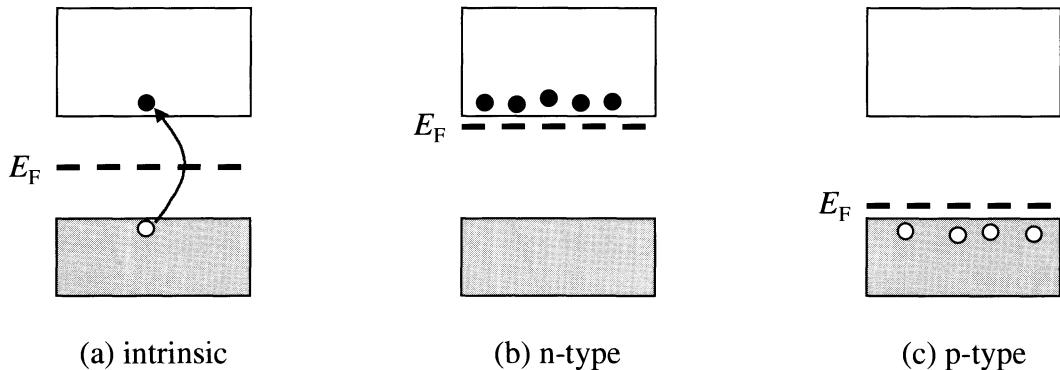


Fig. C.2 Valence and conduction bands for: (a) a pure (intrinsic) semiconductor; (b) a semiconductor with n-type doping; and (c) a semiconductor with p-type doping. The \bullet and \circ symbols represent free electrons and holes respectively. The Fermi energies of the n- and p-type samples lie very close in energy to the impurity levels arising from the donor or acceptor atoms, respectively.

one electron per atom compared to the four-valent atoms that make up the bulk of the crystal. Each impurity atom can therefore accept one electron from the valence band. The **acceptor** levels are just above the top of the valence band, and electrons can easily be excited to these empty states at room temperature. This creates a population of free holes in the valence band that determines the extrinsic electrical conductivity of the sample. The Fermi level approximately coincides with the acceptor energy because the states up to the acceptor levels are filled with electrons.

C.2 The nearly free electron model

The motion of free electrons and holes is determined by the E - k dispersion of the solid. If the electrons are completely free, they will only possess kinetic energy. In this case the E - k dispersion is given by:

$$E = \frac{p^2}{2m_0} = \frac{\hbar^2 k^2}{2m_0}, \quad (\text{C.1})$$

where $p = h/\lambda_e = \hbar k$ is the momentum of the electron, λ_e being the de Broglie wavelength of the electron.

In a crystal this E - k dispersion relationship is modified because the electrons are not really free. Each atom possesses a certain number of valence electrons. These are the electrons in the outermost atomic shells that determine the chemical properties. The **nearly free electron model** starts by assuming that the valence electrons are released from their atoms and move through the crystal. This leaves behind a regular lattice of positively charged ions. The potential of the ion cores perturbs the motion of the electrons and alters the E - k relationship.

Figure C.3 shows the typical band structure of a simple crystal. The E - k diagram is divided into different Brillouin zones, each of which spans a reciprocal lattice vector \mathbf{G} . This division of \mathbf{k} -space into Brillouin zones reflects the underlying periodicity of the crystal. The reciprocal lattice vectors are defined by

$$e^{i\mathbf{G}\cdot(\mathbf{r}+\mathbf{T})} = e^{i\mathbf{G}\cdot\mathbf{r}}, \quad (\text{C.2})$$

where \mathbf{T} is one of the primitive lattice translation vectors of the crystal.

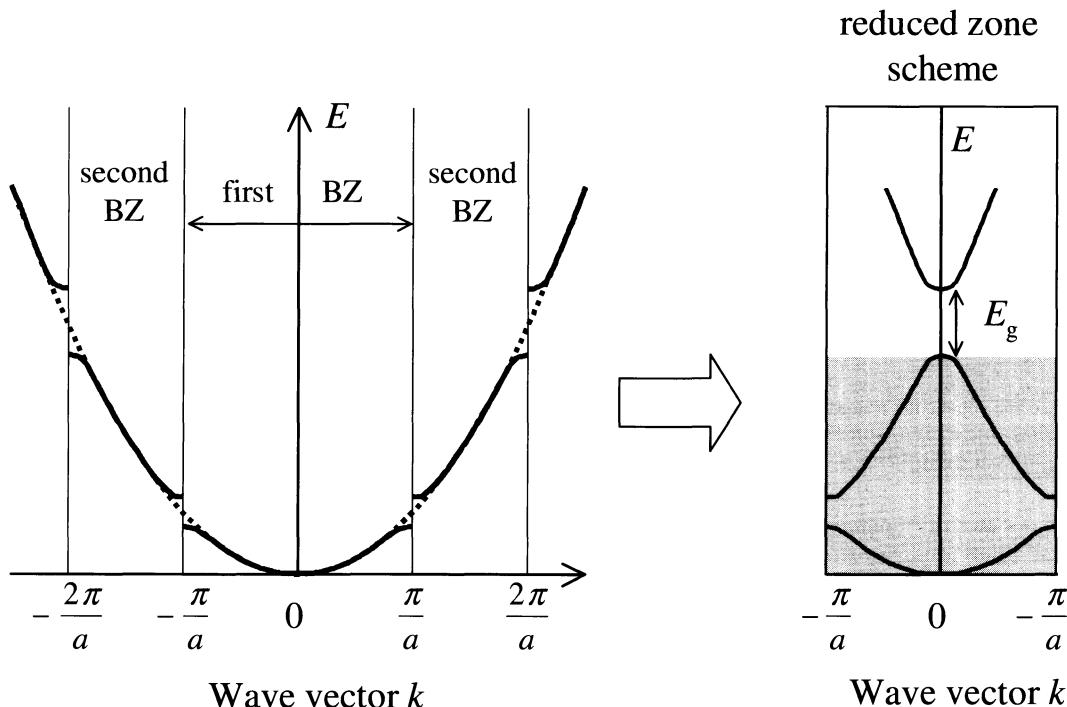


Fig. C.3 Band structure of a simple solid with a lattice constant of a . The left hand side of the figure shows the E - k relationship across several Brillouin zones, while the right hand side replots the same band structure in the reduced zone scheme. The dotted line indicates the parabolic dispersion of free electrons given by eqn C.1. Each band within a Brillouin zone can hold $2N$ electrons, where N is the number of unit cells in the crystal. The shading signifies that the band states are filled with electrons. The case shown applies to an atom such as silicon with four valence electrons.

The E - k relationship shown in Fig. C.3 applies to a simple cubic crystal with a lattice constant of a . In this case eqn C.2 is satisfied with

$$\mathbf{G} = \frac{2\pi}{a}(\mathbf{n}_x, \mathbf{n}_y, \mathbf{n}_z), \quad (\text{C.3})$$

where n_x , n_y and n_z are integers. The band dispersion is drawn for the (100) direction in \mathbf{k} -space. The central Brillouin zone enclosing the origin runs from $-\mathbf{G}/2$ to $+\mathbf{G}/2$, namely from $-\pi/a$ to $+\pi/a$. The next zone encloses the section of \mathbf{k} -space between $\pi/a \rightarrow 2\pi/a$ and $-\pi/a \rightarrow -2\pi/a$, and so on.

The dotted line in the left hand side of Fig. C.3 indicates the band dispersion for free electrons which obeys eqn C.1. The solid lines indicate the band dispersion for the nearly free electrons. The periodic lattice potential causes Bragg scattering of the electron waves. The scattering amplitude is small except in the region close to a Brillouin zone boundary. The band dispersion therefore only departs significantly from that of free electrons near a zone boundary. The group velocity of the electron wave is given by

$$v_g = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{dE}{dk}. \quad (\text{C.4})$$

Precisely at a zone boundary, the wave vector of the forward and backward waves are identical, and a standing wave with zero group velocity is set up. We

must therefore have $dE/dk = 0$ at a zone boundary. This means that the bands must bend over as they approach zone boundaries, as shown in Fig. C.3.

Band structure diagrams are usually plotted in the reduced zone scheme, as indicated by the right hand side of Fig. C.3. In this scheme we translate the electron wave vector by an integer number of reciprocal lattice vectors until it lies within the first Brillouin zone. We can do this because it follows from eqn C.2 that there is no physical difference between the wave vectors \mathbf{k} and $\mathbf{k} + \mathbf{G}$ in a periodic crystal. This conclusion also follows from Bloch's theorem, which is discussed below.

Each Brillouin zone contains N k -vector states, and can therefore hold $2N$ electrons due to the up-down spin degeneracy of each k -state. If each atom has four valence electrons, the first two bands will be full, as shown by the shading in Fig. C.3. This is the situation that applies to four-valent semiconductors such as silicon or germanium. The first available empty electron states are in the next band. This therefore corresponds to the case of the semiconductor or insulator shown in Fig. C.1(b), with an energy gap of E_g between the occupied electron states in the valence band and the first empty states in the conduction band.

The band dispersion near $k = 0$ for a semiconductor or insulator is shown in more detail in Fig. C.4. The top valence band and the lowest conduction band states are shown. The bands are parabolic for small k , and have dispersions given by:

$$\begin{aligned} E_c(k) &= E_g + \frac{\hbar^2 k^2}{2m_e^*} \\ E_v(k) &= -\frac{\hbar^2 k^2}{2m_h^*}, \end{aligned} \quad (\text{C.5})$$

where $E = 0$ corresponds to the top of the valence band. The subscripts on the energy identify the conduction and valence bands respectively. Equation C.5 shows that the band dispersion is determined by the effective mass m_e^* and m_h^* of the appropriate band.

In general, the effective mass is defined by the curvature of the E - k diagram according to:

$$m^* = \hbar^2 \left(\frac{d^2 E}{dk^2} \right)^{-1}. \quad (\text{C.6})$$

The effective mass is therefore a band structure parameter that quantifies the departure of the E - k relationship from the free electron dispersion as a consequence of the perturbation of the periodic lattice potential. It will generally be the case that neither m_e^* nor m_h^* are equal to the free electron mass m_0 , and that each material will be different. The negative curvature of the valence band indicates that it is a hole state: hence the 'h' subscript for the effective mass in the valence band. Electrons in the conduction band behave like negative free particles of mass m_e^* , while the holes in the valence band behave like free positive particles of mass m_h^* . Tables of effective masses are readily available. Tables C.1 and C.2 below give the values for a few important semiconductors.

The nearly free electron approach can be connected to the atomic states of the atoms from which the valence electrons are derived through Bloch's theorem (cf. Section 1.5.2):

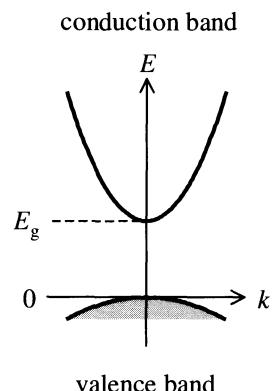


Fig. C.4 Band dispersion of a semiconductor or insulator near the top of the valence band and the bottom of the conduction band. The energy axis is defined so that $E = 0$ corresponds to the top of the valence band.

The eigenfunctions of the wave equation for a periodic potential are the product of a plane wave $\exp i\mathbf{k} \cdot \mathbf{r}$ and an envelope function $u(\mathbf{r})$ which has the periodicity of the crystal lattice.

This implies that the wave function of an electron in a periodic lattice takes the form:

$$\psi(\mathbf{r}) = u(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}}. \quad (\text{C.7})$$

where $u(\mathbf{r})$ must satisfy $u(\mathbf{r}) = u(\mathbf{r} + \mathbf{T})$. The Bloch functions are therefore modulated plane waves. The envelope function $u(\mathbf{r})$ is a wavelike periodic function that relates to the atomic character of the valence electrons. This link is formalized in the tight-binding approach to band structure calculations.

C.3 Example band structures

The band structure of aluminium is shown in Fig. 7.3 in Chapter 7. Aluminium is a trivalent metal with the three valence electrons in a configuration of $3s^23p^1$. The band structure looks much more complicated than Fig. C.3, but this is mainly a consequence of the way band diagrams are drawn. To understand how the diagram works, we first need to consider the shape of the Brillouin zone in three dimensions.

Aluminium has a face-centred cubic (f.c.c.) lattice. The cubic unit cell of an f.c.c. lattice is not the primitive unit cell as it contains four lattice points: one at the origin and three others at the centre of the cube faces with coordinates $(1/2, 1/2, 0)$, $(1/2, 0, 1/2)$ and $(0, 1/2, 1/2)$. The Brillouin zone of the f.c.c. lattice is therefore not a cube, but rather has the shape shown in Fig. C.5. The dispersion of the bands shown in Fig. 7.3 begins by plotting the energy for increasing \mathbf{k} outwards from the origin to the X-point. We then move across to the L-point via the W-point, and back to the origin. Finally we go out again from the origin to the X-point via the K-point.

The departure from the free electron dispersion is actually very small in aluminium. Most of the band diagram can be explained by taking the parabolic dispersion of free electrons shown by the dotted line in Fig. C.3 and folding it back into the complicated shape of the f.c.c. Brillouin zone. The changes in the curvature of the bands at the zone boundaries are then merely caused by the change of direction in which we are moving around the Brillouin zone. Note however that there are small gaps between most bands at the zone edges. These are the band gaps introduced by the lattice potential.

The band structure of the transition metal copper is shown in Fig. 7.5 in Chapter 7. Copper has an electronic configuration of $3d^{10}4s^1$ and crystallizes with an f.c.c. lattice. The band structure is more complicated than that of aluminium because of the need to include the dispersion of both the $3d$ and $4s$ bands, which overlap in energy. The $4s$ bands are approximately parabolic, but the $3d$ bands are fairly flat. This is a consequence of the strong localization of the d electrons, which means that their orbitals do not overlap much in the crystal. The low dispersion d -bands are the origin of the high density of states within a relatively narrow range of energies. These states are very important for both the optical and magnetic properties.

The band structure of the semiconductor silicon is shown in Fig. 3.12 in Chapter 3. Silicon has four valence electrons and crystallizes with the diamond

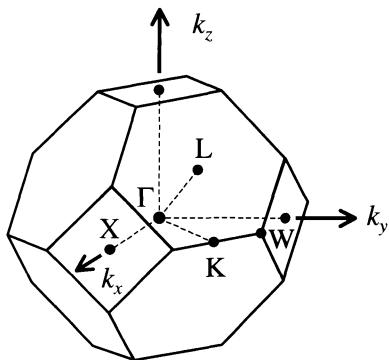


Fig. C.5 Brillouin zone of a face-centred cubic lattice. a is the size of the cubic unit cell in the crystal. High symmetry points within the Brillouin zone are given symbolic names derived from group theory. The origin at $\mathbf{k} = (0, 0, 0)$ is called the Γ -point. The X-point at $(2\pi/a)(1, 0, 0)$ identifies the zone edge along the six equivalent (100) directions. The L-point at $(\pi/a)(1, 1, 1)$ is at the zone edge along the eight equivalent (111) directions. An arbitrary point along the $\Gamma \rightarrow X$ direction is labelled Δ , while an arbitrary point along the $\Gamma \rightarrow L$ direction is given the symbol Λ . The K-point corresponds to the zone edge along the (110) direction, while the W-point is at the intersection of the square and hexagonal faces of the polygon. The X-point corresponds to a wave vector of $2\pi/a$ rather than π/a as Fig. C.3 would suggest because the cubic unit cell of the f.c.c. lattice is not primitive.

Table C.1 Band structure parameters for the indirect gap group IV elements diamond, silicon and germanium. All three materials crystallize with the diamond structure. E_g^{ind} : indirect band gap; k_{\min} : position of the conduction band minimum within the Brillouin zone; valley degeneracy: number of equivalent conduction band minima within the Brillouin zone; $m_{e\parallel}^*$: longitudinal electron effective mass; $m_{e\perp}^*$: transverse electron effective mass; E_g^{dir} : direct band gap at the Γ point; Δ : spin-orbit splitting at the Γ point; m_{hh}^* : heavy hole effective mass; m_{lh}^* : light hole effective mass; m_{so}^* : split-off hole effective mass. The effective masses are expressed in units of the free electron mass m_0 . The valence band parameters refer to the four-band model shown in Fig. 3.5. After [1].

Property	Diamond	Silicon	Germanium
E_g^{ind} (eV) (300 K)	5.47	1.12	0.66
E_g^{ind} (eV) (0 K)	5.5	1.17	0.74
k_{\min}	0.76 X	0.85 X	L
Valley degeneracy	6	6	4
$m_{e\parallel}^*$	1.4	0.92	1.58
$m_{e\perp}^*$	0.36	0.19	0.08
E_g^{dir} (eV) (300 K)	6.5	4.1	0.805
Δ (eV)	0.006	0.044	0.29
m_{hh}^*	1.08	0.54	0.3
m_{lh}^*	0.36	0.15	0.04
m_{so}^*	0.15	0.23	0.095

structure. The diamond structure consists of two identical interlocking f.c.c. lattices displaced from each other by $(a/4, a/4, a/4)$. The structure is f.c.c. with a basis of two atoms attached to each lattice point: one at the lattice point itself, and the other displaced by $(1/4, 1/4, 1/4)$ with respect to it. Silicon therefore has an f.c.c. Brillouin zone, as shown in Fig. C.5.

The real band structure of silicon can be compared to the schematic band dispersions shown in Figs C.3 and C.4. We see that the real material does show the general behaviour indicated by the schematic diagrams, although the actual band diagram is more complicated. One significant difference is the ‘camel back’ shape of the conduction band, which means that the minimum of the conduction band occurs near the X-point rather than at the Γ -point. The band gap of silicon is therefore indirect. This has very important consequences for the optical properties, as discussed in Chapters 3 and 5. Another important difference is the degeneracy of the valence band states at the Γ -point. This is usually described in terms of the four-band model shown in Fig. 3.5. The parameters needed to describe the valence band of silicon in this way are listed in Table C.1. The spin-orbit splitting of silicon is too small to be apparent in the low scale band diagram given in Fig. 3.12. Table C.1 also lists the effective masses that describe the conduction band minima near the X-point. Note that we must use two separate effective masses to parameterize the anisotropy of the conduction band minimum.

The band structure of germanium is given in Fig. 3.9 in Chapter 3. Germanium lies one line below silicon in the periodic table, and, like silicon, has the diamond crystal structure. It is not surprising, therefore, that the band structures are fairly similar. There are, however, a number of important differences. Most prominent among these is the fact that the conduction band minimum is at the L-point, rather than near the X-point. Moreover, the minimum at the

Table C.2 Band structure parameters for selected direct gap III–V semiconductors with the zinc blende structure. The parameters listed refer to the four-band model shown in Fig. 3.5. E_g : band gap; Δ : spin-orbit splitting; m_e^* : electron effective mass; m_{hh}^* : heavy hole effective mass; m_{lh}^* : light hole effective mass; m_{so}^* : split-off hole effective mass. The effective masses are expressed in units of the free electron mass m_0 . After [1] and [2].

Crystal	E_g (eV) (0 K)	E_g (eV) (300 K)	Δ (eV)	m_e^*	m_{hh}^*	m_{lh}^*	m_{so}^*
GaAs	1.519	1.424	0.34	0.067	0.5	0.08	0.15
GaSb	0.81	0.75	0.76	0.041	0.28	0.05	0.14
InP	1.42	1.34	0.11	0.077	0.6	0.12	0.12
InAs	0.42	0.35	0.38	0.022	0.4	0.026	0.14
InSb	0.24	0.18	0.85	0.014	0.4	0.016	0.47

Table C.3 Structure and band gap data for a number of common semiconductors. E_g is the band gap at 300 K, and the i/d label indicates whether the gap is indirect or direct. SiC crystallizes in more than 200 different modifications, and the data listed here is for the 6H polytype, which has a hexagonal unit cell. ZnS, ZnSe, CdS and CdSe can form stable crystals with either hexagonal or cubic unit cells, and the band gap may be slightly different for the two structural variations. The negative band gap of HgTe signifies that it is a semimetal: the top of the valence band is at higher energy than the bottom of the conduction band. After [1].

Compound	Crystal structure	E_g (eV)	Type
SiC	6H polytype	2.9	i
AlN	wurtzite	6.2	d
AlP	zinc blende	2.41	i
AlAs	zinc blende	2.15	i
AlSb	zinc blende	1.62	i
GaN	wurtzite	3.44	d
GaP	zinc blende	2.27	i
InN	wurtzite	1.89	d
ZnO	wurtzite	3.4	d
ZnS	wurtzite or zinc blende	3.8 or 3.7	d
ZnSe	wurtzite or zinc blende	2.8 or 2.7	d
ZnTe	zinc blende	2.3	d
CdS	wurtzite or zinc blende	2.5	d
CdSe	wurtzite or zinc blende	1.8	d
CdTe	zinc blende	1.5	d
HgTe	zinc blende	-0.14	semimetal
CuCl	zinc blende	3.17	d
Cu ₂ O	cuprite	2.2	d

Γ -point is only just above the one at the L-point. The band gap is therefore still indirect, but the optical transitions soon become direct as the photon energy is increased above E_g . The principal band structure parameters of germanium are listed in Table C.1.

The band structure of the III–V compound semiconductor gallium arsenide, which has the **zinc blende structure**, is given in Fig. 3.4 in Chapter 3. The zinc blende structure is similar to the diamond structure, except that the atom at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ is different to the one at $(0, 0, 0)$. The band structure is quite similar to that of germanium, except that the conduction band minimum now lies at the Γ -point. GaAs therefore has a direct band gap. This means that GaAs crystals

can emit light efficiently when excited, as discussed in Chapter 5. Table C.2 lists the most significant band parameters for GaAs, together with those of a few other direct gap III–V materials.

Less detailed band structure data on other important compound semiconductors is given in Table C.3. Some of these crystals have the zinc blende structure, while a number of others have the **wurtzite structure**, which has hexagonal symmetry. Several of the II–VI compounds can form stable crystals with either structure, and the band gap may be slightly different between the cubic and hexagonal forms. Cu₂O has its own particular structure, not surprisingly named the cuprite structure. The cuprite structure has cubic symmetry.

Further reading

An introductory treatment of band theory is given in Rosenberg (1988). More detailed accounts are given in Ashcroft and Mermin (1976), Burns (1985), Ibach and Luth (1995), Kittel (1996), or Singleton (2001), and many other texts on solid state physics.

References

- [1] Madelung, O. (1996). *Semiconductors, basic data* (2nd edn). Springer-Verlag, Berlin.
- [2] Madelung, O. (1982). *Semiconductors: physics of group IV elements and III–V compounds*, Landolt-Börnstein New Series, Vol. III/17a. Springer-Verlag, Berlin.

Appendix D

Semiconductor p-i-n diodes

The p-i-n structure is used extensively in semiconductor optoelectronic devices such as photodiodes, solar cells, light emitting diodes and optical modulators. The structure is shown schematically in Fig. D.1. It consists of a standard semiconductor p-n diode with a thin undoped i-region of thickness l_i inserted at the junction. This i-region is the optically active part of the diode. The purpose of the p-n junction is to control the number of electrons and holes injected into the active region, and to permit the application of strong electric fields.

In this appendix we discuss the band alignment and electrostatics of the p-i-n structure when an external bias voltage V_0 is applied to the device. The formation of the **depletion region** at the junction is an essential feature of the physics of the p-n diode. The external bias is dropped across the depletion region because it has a very high resistance compared to the heavily doped p- and n-regions. The width of the depletion region at a given voltage is determined by the doping levels in the p- and n-regions, with higher doping giving thinner depletion widths. In a p-i-n structure the residual doping level in the i-region is very small, and so the depletion region can extend across the whole i-region. The extension of the depletion region into the p- and n-regions is very small in comparison to l_i because of the heavy doping level in the contacts. This means that any external voltages that are applied will be dropped almost entirely across the i-region.

Fig. D.1 Schematic diagram of a p-i-n diode. The bias voltage V_0 is applied to the p-region, so that positive and negative V_0 correspond to forward and reverse bias respectively. The dimensions are not drawn to scale. The i-region is typically only a few microns thick.

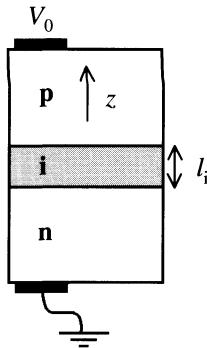


Figure D.2 shows a band diagram of a p-i-n structure. Figure D.2(a) shows the band alignments at zero bias, while Fig. D.2(b) applies to the situation with an external voltage applied. At zero bias the Fermi levels of the p- and n-regions align with each other. If we assume that the Fermi energies are small compared to the band gap, we see from Fig. D.2(a) that there is a voltage drop of magnitude E_g/e across the i-region in these conditions. This is equivalent to the built-in voltage V_{bi} that is important for the functioning of solar cells as discussed in Section 3.7. When the bias is applied, the energy difference between the Fermi levels of the p- and n-regions will be equal to $|eV_0|$. This is illustrated in Fig. D.2(b) for the case of reverse bias, that is, when a negative voltage is applied to the p-region with respect to the n-region. Reverse bias tends to increase the voltage drop across the i-region, while forward bias tends to reduce it.

In order to calculate the electric field across the i-region, we need to solve Poisson's equation (Appendix A, eqn A.5):

$$\nabla^2 V = -\frac{\varrho}{\epsilon_r \epsilon_0}, \quad (\text{D.1})$$

where V is the voltage and ϱ is the electric charge density. We set up axes so that z is the direction normal to the plane of the diode. We know from the

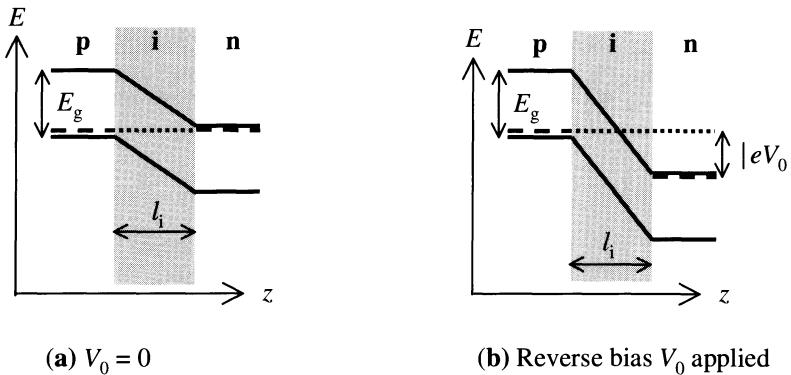


Fig. D.2 Band alignments in a p-i-n diode structure with an i-region thickness of l_i . (a) Bias = 0. (b) Reverse bias voltage V_0 applied. The thick dashed lines indicate the Fermi levels of the doped layers, which lie just above the valence band or just below the conduction band in the p- and n-regions respectively. E_g is the band gap of the semiconductor used for the p- and n-regions.

symmetry that the derivatives in the x and y planes must be zero. Poisson's equation therefore reduces to:

$$\frac{\partial^2}{\partial z^2} V(z) = -\frac{\varrho(z)}{\epsilon_r \epsilon_0}. \quad (\text{D.2})$$

We assume that $\varrho = 0$ in the i-region because it is undoped and is fully depleted of all free carriers. The solution to eqn D.2 in the i-region is therefore $V(z) = C_1 z + C_2$, where C_1 and C_2 are constants. The electric field strength can be calculated from eqn A.7 in Appendix A. This gives $\mathcal{E} = -dV/dz = -C_1$, which implies that the electric field is constant in the i-region.

We can see from Fig. D.2(b) that the magnitude of the voltage drop across the i-region is approximately equal to $|V_{bi} - V_0|$. The value of the constant electric field in the i-region is therefore given by:

$$\mathcal{E} = \frac{|V_{bi} - V_0|}{l_i}. \quad (\text{D.3})$$

This shows that negative bias increases the field across the i-region, while a small forward bias reduces \mathcal{E} . The field is zero for a forward bias of V_{bi} . At zero bias the field across the i-region is equal to V_{bi}/l_i . This can be a large value. For example, with a GaAs diode we might have $V_{bi} = 1.5$ V and $l_i = 1$ μm . The field at zero bias is therefore 1.5×10^6 m $^{-1}$.

In this book we frequently come across reverse-biased p-i-n structures in the context of studying the effects of electric fields on the optical properties of semiconductors. Section 3.3.5 discusses the effect of an electric field on the band edge absorption of a bulk semiconductor, while Sections 4.3.1 and 6.5 discuss the effects on the excitons in bulk semiconductors and quantum wells, respectively. In all of these examples, the calibration of the electric field strength from the applied bias is carried out by using eqn D.3.

Reverse-biased p-i-n structures are also used in semiconductor photodiodes and solar cells, as discussed in Section 3.7. Forward-biased p-i-n structures are used in electroluminescent devices: see Section 5.4.

Further reading

The physics of the p-n junction is discussed in most solid state physics texts. See, for example: Bleaney and Bleaney (1976), Rosenberg (1988), or Sze (1985). The p-i-n structure is described in detail in Sze (1981).

Solutions to exercises

Chapter 1

- (1.1) $R = 0.041$ and $T = 0.92$, assuming that $\alpha = 0$ because glass is transparent.
- (1.2) 2.1.
- (1.3) $v = 9.97 \times 10^7 \text{ m s}^{-1}$, $\alpha = 9.6 \times 10^6 \text{ m}^{-1}$ and $R = 25.6\%$.
- (1.4) $18 \mu\text{m}$.
- (1.5) $T = 0.034$, optical density = 1.1.
- (1.6) $\tilde{\epsilon}_r = 1.77 + i 9.2 \times 10^{-8}$.
- (1.7) Absorbing for blue light, but not for red and green light.
- (1.8) (ii) 1.6×10^{-3} . (iii) 0.04. (iv) The field ratio is important at normal or near-normal incidence when multiple beam interference occurs between the exiting beams. See Hecht (1998) for further details.
- (1.9) (ii) 5.9×10^{-4} , (iii) 0.024.
- (1.10) From eqn 1.6 we see that:

$$-\log_{10}(T) = -2 \log_{10}(1 - R) + \alpha l / \ln(10).$$

The optical density is found by comparing this to eqn 1.8.

If the medium is transparent at λ' then we will have that $(T_{\lambda'}) = (1 - R)^2$. The optical density at λ is therefore given by:

$$\text{O.D.}(\lambda) = -\log_{10}(T_\lambda) + \log_{10}(T_{\lambda'}).$$

This result will only hold if λ' is close to λ because we are assuming that R does not vary significantly with the wavelength.

- (1.11) 27 ns. The 1500 nm pulse takes less time.
- (1.12) 99.6 %.
- (1.13) 14 nm.
- (1.14) $\tilde{\epsilon}_r \approx 1$.
- (1.15) (i) 0.294 eV/atom; (ii) 8 W and 2 W; (iii) 4 W and 6 W.
- (1.16) 521 nm.
- (1.17) 81 %. The scattering cross-section is 11 times larger at 850 nm than at 1550 nm.
- (1.18) 3.5 m, 0.22 m.
- (1.19) Ice is a uniaxial crystal, but water is a liquid with no preferential axes.

Chapter 2

$$(2.1) \quad m_1 \ddot{x}_1 = -K_s(x_1 - x_2) \\ m_2 \ddot{x}_2 = -K_s(x_2 - x_1).$$

Divide these equations by m_1 and m_2 respectively and subtract them to obtain the equation of motion for an oscillator of frequency $(K_s/\mu)^{1/2}$:

$$\frac{d^2}{dt^2}(x_1 - x_2) = -K_s \left(\frac{1}{m_1} + \frac{1}{m_2} \right) (x_1 - x_2).$$

$$(2.2) \quad -\tan^{-1}[\omega\gamma/(\omega_0^2 - \omega^2)].$$

- (2.3) 6.3×10^{-4} .
- (2.4) 270 m^{-1} .
- (2.5) $\alpha(\omega_0) = Ne^2/n\epsilon_0 m_0 \gamma c$.
- (2.6) (i) 5.9; (ii) $5.0 \times 10^{12} \text{ Hz}$; (iii) 23 N; (iv) $3.0 \times 10^{28} \text{ m}^{-3}$; (v) about $6 \times 10^{12} \text{ s}^{-1}$; (vi) about $1 \times 10^6 \text{ m}^{-1}$. In part (iii), work out the spring constant K_s using eqn 2.2, remembering to use the correct reduced mass ($2.3 \times 10^{-26} \text{ kg}$). The answers to parts (v) and (vi) are only approximate

because the data does not exactly follow a simple Lorentzian line shape. The damping rate depends strongly on the frequency, which is why the resonance line is highly asymmetric.

- (2.7) Use $\omega = ck/n$ in eqn 2.25 to obtain:

$$v_g = \frac{d\omega}{dk} = \frac{c}{n} - \frac{ck}{n^2} \frac{dn}{dk}.$$

Then substitute $v = c/n$ to obtain eqn 2.26.

- (2.8) The dispersion in this undamped system is always normal, so that from eqn 2.26 we see that $v_g < v$. If $\omega < \omega_0$, $\epsilon_r > 1$ and therefore $v = c/n = c/\sqrt{\epsilon_r} < c$. Hence $v_g < c$ if $\omega < \omega_0$. For $\omega > \omega_0$ we must work out v_g explicitly:

$$v_g = nc \left(1 + \frac{Ne^2}{\epsilon_0 m_0} \frac{\omega_0^2}{(\omega_0^2 - \omega^2)^2} \right)^{-1}.$$

The denominator is greater than unity, and $n < 1$, so $v_g < c$.

- (2.9) This derivation is given in many solid state or electromagnetism textbooks. See, for example: Ibach and Luth (1996) or Kittel (1996).
- (2.10) Either when the density of absorbing atoms is small or when the frequency is far away from any resonance.

- (2.11) $\chi_a = 2.2 \times 10^{-29} \text{ m}^3$. The two field strengths are $0.8 \times 10^{11} \text{ V m}^{-1}$ and $1.4 \times 10^{11} \text{ V m}^{-1}$ respectively. It is not surprising that these values are of similar magnitude because the external field must work against the Coulomb forces in the molecule to induce a dipole.

- (2.12) (a) $\lambda_j = 2\pi c/\omega_{0j}$, $A_j = Ne^2 f_j \lambda_j^2 / 4\pi^2 \epsilon_0 m_0 c^2$.
 (b) $C_1 = (1 + A_1)^{\frac{1}{2}}$; $C_2 = A_1 \lambda_1^2 / 2(1 + A_1)^{1/2}$;
 $C_3 = A_1(4 + 3A_1)\lambda_1^4 / 8(1 + A_1)^{3/2}$.

- (2.13) (i) $C_1 = 1.5255$, $C_2 = 4824.7 \text{ nm}^2$. (ii) 1.5493 and 1.5369. (iii) 1.26° .

- (2.14) 0.17 ps if we assume a time-bandwidth product $\Delta\nu \Delta t = 1$.

- (2.15) Use $\epsilon_{11}/\epsilon_0 = \epsilon_{22}/\epsilon_0 = n_o$, $\epsilon_{33}/\epsilon_0 = n_e$, $x = 0$, $z/n(\theta) = \sin\theta$, and $y/n(\theta) = \cos\theta$ to derive the result.

- (2.16) $14 \mu\text{m}$.

- (2.17) (a) No (face-centred cubic); (b) no (cubic); (c) yes (hexagonal); (d) yes (hexagonal); (e) no (cubic); (f) no (face-centred cubic); (g) yes (orthorhombic). Sulphur is biaxial.

Chapter 3

- (3.1) $\mathbf{k} = (2\pi/L)(n_x, n_y, n_z)$, where n_x , n_y and n_z are integers. Each allowed \mathbf{k} -state occupies a volume of \mathbf{k} -space equal to $(2\pi/L)^3$, which implies that the number of states in a unit volume of \mathbf{k} -space is $L^3/(2\pi)^3$. Hence a unit volume of the material would have $1/(2\pi)^3$ states per unit volume of \mathbf{k} -space.

- (3.2) $dE/dk = \hbar^2 k/m^*$. Use this in eqn 3.14, and then substitute for k .

- (3.3) (i) The parity of a wave function is equal to ± 1 depending on whether $\psi(-\mathbf{r}) = \pm \psi(\mathbf{r})$. Atomic wave functions have well defined parities because atoms have inversion symmetry about $\mathbf{r} = 0$, and hence we must have that $|\psi(-\mathbf{r})|^2 = |\psi(\mathbf{r})|^2$.
 (ii) \mathbf{r} is an odd function, and so the integral will be zero unless the two wave functions have different parities.

- (iii) For z polarized light we have:

$$M \propto \int_{\phi=0}^{2\pi} e^{im'\phi} (r \cos\theta) e^{im\phi} d\phi,$$

which is zero unless $m' = m$. For x or y polarized light we have:

$$M \propto \int_{\phi=0}^{2\pi} e^{im'\phi} \left(r \sin\theta (e^{i\phi} \pm e^{-i\phi}) \right) e^{im\phi} d\phi,$$

which is zero unless $m' = m \pm 1$.

- (3.4) Same as Fig. 3.13, but with a scanning monochromator and an InSb detector.
- (3.5) Plot α^2 and $\alpha^{1/2}$ against $\hbar\omega$. Also investigate the temperature dependence of α .
- (3.6) Indirect band gap at 2.2 eV. Direct band gap at ~ 2.75 eV.
- (3.7) $\alpha \approx 1 \times 10^6 \text{ m}^{-1}$.
- (3.8) (i) $5.3 \times 10^8 \text{ m}^{-1}$ and $4.1 \times 10^8 \text{ m}^{-1}$. (ii) $3.0 \times 10^7 \text{ m}^{-1}$. This is more than an order of magnitude

- smaller than the electron wave vector. (iii) 2.1. (iv) 704 nm.
- (3.9) (i) 4.1 eV. This corresponds to transitions from the *p*-like valence band to the *s*-like conduction band. (ii) The discussion of the atomic character of bands given in Section 3.3.1 only applies at the Γ point. This means that electric dipole transitions can be allowed at the zone edges, even though they are forbidden at $k = 0$.
- (3.10) 0.75 eV.
- (3.11) $\mathcal{E} \approx 1.8 \times 10^6 \text{ V m}^{-1}$. This value is obtained by working out the field at which α drops by a factor e^{-1} between E_g and $(E_g - 0.01)$ eV.
- (3.12) The first part is easily derived by equating the central force for circular motion with the Lorentz force: $m\omega^2 r = e\omega r B$. The $\Delta n = 0$ selection rule follows from the orthogonality of harmonic oscillator wave functions φ_n , after using Landau level wave functions of the form $\psi_n(\mathbf{r}) \propto u(\mathbf{r}) \varphi_n(x, y) e^{ik_z z}$.
- (3.13) (i) $g_{1D}(E) = (2m/Eh^2)^{-1/2}$, where m is the particle mass. (ii) $\alpha \propto (\hbar\omega - E_g)^{-1/2}$. (iii) The magnetic field quantizes the motion in two dimensions. The absorption coefficient for transitions between Landau levels varies as $(\hbar\omega - E_n)^{-1/2}$, where $E_n = E_g + (n + \frac{1}{2})(e\hbar B/\mu)$. This follows from the 1-D density of states and the $\Delta n = 0$ selection rule. $\alpha(\hbar\omega)$ diverges each time the frequency crosses the threshold for a new value of n . These divergences are broadened by scattering. We therefore see dips in the transmission at each value of $\hbar\omega$ that satisfies eqn 3.32. (iv) $m_e^* \approx 0.035m_0$ and $E_g = 0.80$ eV. These values refer to the Γ point of the Brillouin zone.
- (3.14) 0.46 A W⁻¹ at 1.55 μm and 1.05 A W⁻¹ at 1.30 μm .
- (3.15) (i) The p- and n-regions are good conductors, whereas the i-region is depleted of free carriers and therefore acts like an insulator. (ii) 10 pF. (iii) 60 ps for the electrons and 200 ps for the holes. (iv) 0.2 V.
- ## Chapter 4
- (4.1) This is a standard result for any two particle system.
- (4.2) (i) Kinetic energy + Coulomb energy. (ii) $E = -\mu e^4 / 8\epsilon_r \epsilon_0^2 h^2 = -(\mu/m_0 \epsilon_r^2) R_H$, $a_0 = \epsilon_0 \epsilon_r h^2 / \pi \mu e^2 = (\epsilon_r m_0 / \mu) a_H$, $C = \pi^{-1/2} a_0^{-3/2}$.
- (4.3) a_0 . $\langle r \rangle = (3/2)a_0$.
- (4.4) (i) This is a spherically symmetric function, with a maximum value at $r = 0$. (ii) $\langle E \rangle = \hbar^2 / 2\mu \xi^2 - e^2 / 4\pi \epsilon_r \epsilon_0 \xi$. (iii) $\xi_{\min} = 4\pi \epsilon_0 \epsilon_r \hbar^2 / \mu e^2$, $\langle E \rangle_{\min} = -\mu e^4 / 8h^2 \epsilon_0^2 \epsilon_r^2$. (iv) ξ_{\min} and $\langle E \rangle_{\min}$ are the same as a_0 and E from Exercise 4.2. The variational method gives the energy and wave function exactly here because our ‘guess’ wave function had the correct functional form.
- (4.5) (ii) $E(n) = -\mu e^4 / 8h^2 \epsilon_0^2 \epsilon_r^2 n^2$, $r_n = 4\pi \epsilon_0 \epsilon_r \hbar^2 n^2 / \mu e^2$. (iii) $E(n)$ is identical to the solution of the hydrogen Schrödinger equation. (iv) r_1 corresponds to the peak in the radial probability density for the ground state 1s wave function.
- (4.6) $E(1) = -39.1$ meV, $r_1 = 2.3$ nm, stable. $E(2) = -9.8$ meV, $r_2 = 9.3$ nm, unstable.
- (4.7) 2.2 nm.
- (4.8) The refractive index has a maximum value of 3.60 at 1.5146 eV.
- (4.9) 394 μm .
- (4.10) Substitute $|E(n)| = \mu e^4 / 8(\epsilon_0 \epsilon_r h n)^2$ and $r_n = 4\pi \epsilon_0 \epsilon_r \hbar^2 n^2 / \mu e^2$ into $\mathcal{E} = e / 4\pi \epsilon_0 \epsilon_r r^2$ to obtain the result, using $|E(1)| = R_X$ and $r_1 = a_X$.
- (4.11) 1.5 meV and 31 nm. $V_0 = +0.55$ V.
- (4.12) 1.8 T.
- (4.13) $\mathbf{B} = \nabla \wedge \mathbf{A} = (0, 0, B_0)$. $\hat{H}' = e^2 \mathbf{A}^2 / 2m_0 = e^2 B_0^2 (x^2 + y^2) / 8m_0$. $\langle E \rangle = \langle \psi | \hat{H} | \psi \rangle$. Equation 4.7 follows by adding the contributions of the electron and hole, and remembering that spherical symmetry implies that $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle = \frac{1}{3} \langle r^2 \rangle$.
- (4.14) $\delta E = +4.9 \times 10^{-5}$ eV, $\delta \lambda = -0.026$ nm.
- (4.15) $8.1 \times 10^{24} \text{ m}^{-3}$ and $1.3 \times 10^{23} \text{ m}^{-3}$.
- (4.16) 0.50.
- (4.17) 17.2 K.
- (4.18) $r_1 = 0.85$ nm: invalid. $r_2 = 3.4$ nm: valid.

Chapter 5

- (5.1) See Section 5.2.2.
- (5.2) The relaxation within the bands is faster than the radiative recombination.
- (5.3) $A_{2p \rightarrow 1s} = 6.27 \times 10^8 \text{ s}^{-1}$. $\tau_R = 1.6 \text{ ns}$.
- (5.4) Faster non-radiative recombination at higher temperatures due to phonon emission. $\eta_R(300 \text{ K}) = 79 \%$, $\eta_R(350 \text{ K}) = 56 \%$.
- (5.5) ZnTe. (The device might also contain $\text{Ga}_x\text{In}_{1-x}\text{N}$ with $x \approx 0.26$: see Exercise 5.15.)
- (5.6) (i) This follows directly from the definition of α given in eqn 1.3. (ii) Set $\dot{N} = I\alpha/h\nu - N/\tau$ equal to zero. (iii) $6.6 \times 10^{20} \text{ m}^{-3}$.
- (5.7) (i) $1.9 \times 10^{24} \text{ m}^{-3}$. (ii) 0.62 ns. (iii) 3.5×10^{10} photons.
- (5.8) The emission rate is proportional to the probability that the upper level is occupied and that the lower level is empty, that is, $f_e \times f_h$. In the classical limit, $f_{e,h} \propto \exp(-E_{e,h}/k_B T)$, so
- $$\begin{aligned} f_e f_h &= \exp(-(E_e + E_h)/k_B T) \\ &= \exp(-(h\nu - E_g)/k_B T). \end{aligned}$$
- (5.9) (a) $E_F = -0.216 \text{ eV} = -8.4k_B T$, valid.
(b) $E_F = +0.021 \text{ eV} = +0.83k_B T$, invalid.
- (5.10) $f(E) = 1$ for $E < E_F$ and $f(E) = 0$ for $E > E_F$.
- (5.11) Electrons: (a) 0.36 meV, degenerate for $T \ll 4.2 \text{ K}$; (b) 36 meV, degenerate for $T \ll 420 \text{ K}$. Holes: (a) 0.073 meV, degenerate for $T \ll 0.9 \text{ K}$; (b) 7.3 meV, degenerate for $T \ll 85 \text{ K}$.
- (5.12) $k_F = (3\pi^2 N)^{1/3}$.
- (5.13) (i) Solid angle $\Omega = 3.9 \times 10^{-3}$.
(ii) $(1 - R)\Omega/n^2 = 4.2 \times 10^{-4}$.
(iii) $(1.61/2.41)\eta_R \times (1 - R) \text{ mW} = 0.53\eta_R \text{ mW}$.
(iv) $0.22\eta_R \mu\text{W}$.
- (5.14) (i) 0.14 eV. (ii) 0.012 eV. (iii) Electrons degenerate, but not holes. (iv) $E_g + E_F^c$ may be read from the spectrum at about 0.94 eV as the point where the luminescence falls to 50 % of its peak value. This agrees well with the estimate of E_F^c from the carrier density. (v) Read E_F^c from the data to find $N_e \approx 3 \times 10^{23} \text{ m}^{-3}$. $\tau \approx 0.13 \text{ ns}$.
- (5.15) $x = 0.39$.
- (5.16) (i) 31 %. (ii) $4.3 \times 10^{10} \text{ Hz}$. (iii) 610 m^{-1} .
- (5.17) (i) 150 mW. (ii) 26 %. (iii) 0.77 W A^{-1} , 51 %.

Chapter 6

- (6.1) About 0.01 K.
- (6.2) 9.3 nm and 30 nm.
- (6.3) The k vector must satisfy $k = \text{integer} \times 2\pi/L$, and hence the area per k -state in k -space is $(2\pi/L)^2$. The density of states in k -space is therefore $(1/2\pi)^2 \cdot g(k)dk = 2\pi k dk \times (1/2\pi)^2 \cdot g(E) = 2g(k)dk/dE$, where $dk/dE = m/\hbar^2 k$. The factor of 2 comes from the spin degeneracy.
- (6.4) The function on the right hand side of eqn 6.26 decreases from $+\infty$ at $x = 0$ to zero at $x = \sqrt{\xi}$. It will therefore always cross the $\tan x$ function between 0 and $\pi/2$, no matter how small ξ is.
- (6.5) 7.5 meV for the finite well, 11 meV for the infinite one.
- (6.6) (i) This result follows from the orthonormality of the wave functions. (ii) The initial and final states must have opposite parities.
- (6.7) First step at 1.679 eV due to the $n = 1$ heavy hole transition. Second step at 1.837 eV due to the light hole transition. The height of the two steps is in proportion to the reduced masses, that is $0.059 : 0.036$.
- (6.8) (i) The transition energies would be lower. Transitions such as $hh3 \rightarrow e1$ would be weakly allowed.
(ii) Peaks would appear below the steps due to exciton absorption.
- (6.9) (iii) $E_{\min} = -\mu e^4 / 8(\pi \epsilon_0 \epsilon_r \hbar)^2$. This is four times larger than the bulk exciton binding energy found in Exercise 4.4. (iv) $\xi_{\min} = 2\pi \hbar^2 \epsilon_0 \epsilon_r / \mu e^2 = a_X/2$, where a_X is defined in eqn 4.2.
- (6.10) At $d = \infty$ we have bulk GaAs, while at $d = 0$ we have bulk AlGaAs. As d is reduced from ∞ , the binding energy increases from 4 meV, going through a peak, and then dropping to 6 meV. The height of the peak would be about 17 meV, that is, four times larger than the binding energy of bulk GaAs.
- (6.11) (i) See Section 5.3.4. (ii) Heavy exciton and continuum absorption, followed by light hole exciton

and continuum absorption. (iii) The heavy hole continuum starts at 1.592 eV. This implies $d = 9.3$ nm in the infinite well model. The true width would be smaller, because the infinite well model overestimates the confinement energy. (iv) Heavy hole, 11 meV; light hole, 12 meV. A perfect 2-D GaAs quantum well would have $E_b = 4R_X = 16.8$ meV. The experimental binding energies are lower because a real quantum well is not a perfect 2-D system.

- (6.12) (i) $H' = -p_z \mathcal{E}_z$, where $p_z = -ez$. (ii) z is an odd function, while $\varphi^* \varphi$ is even. (iii) The result follows by taking just the first term in the perturbation, namely $|\langle 1 | H' | 2 \rangle|^2 / (E_1 - E_2)$, and substituting the wave functions and energies from eqns 6.11 and 6.13.

- (6.13) (i) 3.4 nm, assuming a quadratic Stark shift.
(ii) $\langle \delta z \rangle \approx 1.6$ nm.

- (6.14) The model works quite well for sample A, but not for sample B. The model breaks down when the size of the Stark shift becomes comparable to the energy splitting of the unperturbed hh1 and hh2 levels. This is essentially the same criterion as for the transition from the quadratic to the linear Stark effect in atomic physics. In sample B, we are in this regime at all the fields quoted.

- (6.15) At finite \mathcal{E}_z the inversion symmetry of the quantum well is broken. The states no longer have definite parities, and selection rules based on parity no longer hold.

- (6.16) The shift is about 0.02 eV, which is comparable to the linewidth observed in the data. A $\pm 5\%$ variation in d corresponds more or less to a fluctuation of one atomic layer.

- (6.17) 14 nm, assuming infinite barriers.
(6.18) (i) z is an odd function, and so the integral will be zero unless $\varphi_n^* \varphi_{n'}$ is also an odd function, which requires that the wave functions must have different parities.
(ii) The matrix elements are given by:

$$\begin{aligned}\langle 1 | z | 2 \rangle &= \frac{2}{d} \int_0^d \sin(\pi z/d) z \sin(2\pi z/d) dz \\ &= -(16/9\pi^2)d \\ \langle 1 | z | 4 \rangle &= \frac{2}{d} \int_0^d \sin(\pi z/d) z \sin(4\pi z/d) dz \\ &= -(4/45\pi^2)d.\end{aligned}$$

The transition strength is proportional to $|M^2|$, and so the $1 \rightarrow 4$ transition is weaker than the $1 \rightarrow 2$ transition by a factor $(1/20)^2 = 2.5 \times 10^{-3}$. $\lambda_{1 \rightarrow 2} = 29 \mu\text{m}$.

- (6.19) The electric field of the light wave in the medium is maximum for grazing incidence with $\theta = 90^\circ$, when the fractional power of the z component is $1/n^2$. The maximum possible fractional absorption is therefore equal to 9% if $n = 3.3$.
- (6.20) The quantized levels occur at energies of 3, 6, 9, 11, 12, 14, 17, ... in units of $\hbar^2/8m^*d^2$.

Chapter 7

(7.1) $E_F^3 = (9\epsilon_0^2 \hbar^2 / 8m_0)(\pi \hbar \omega_p/e)^4$.

(7.2) $N \sim 10^{11} \text{ m}^{-3}$.

- (7.3) $\delta \sim 0.5$ m. To obtain a strong signal in a submerged submarine it is necessary to use much lower frequencies. The data rate would then be very low due to the small carrier frequency.

(7.4) $m_e^* = 1.6m_0$.

(7.5) $R = 99.6\%$.

(7.6) $T = 0.16$.

- (7.7) The drop in the reflectivity for $\lambda < 600$ nm is caused by interband transitions. The energy gap between the d -bands and the Fermi energy can be read from the data as ~ 2.4 eV. The low

reflectivity for green and blue light causes the characteristic yellowish colour.

- (7.8) $\epsilon_r = 1$.
- (7.9) m_e^* increases from $0.020m_0$ at $3.5 \times 10^{23} \text{ m}^{-3}$ to $0.048m_0$ at $4 \times 10^{24} \text{ m}^{-3}$. The increase in m_e^* with N_e is caused by the non-parabolicity in the conduction band of InSb.
- (7.10) $\tau = 1.0 \text{ ps}$.
- (7.11) Calculate the carrier density as in Exercise 5.6, and the free carrier absorption using eqn 7.28. Then add the separate contributions of the electrons and holes together. The final answer is about 200 m^{-1} .

- (7.12) (i) $E_F = 0.032 \text{ eV}$. k_F is $6.5 \times 10^8 \text{ m}^{-1}$ and $2.6 \times 10^8 \text{ m}^{-1}$ for the heavy and light holes respectively.
(ii) (1): 0.03–0.17 eV, (2): 0.32–0.34 eV, (3) 0.34–0.42 eV.
- (7.13) (i) $m_e^* = 0.85m_0$. (ii) $R_0^* \approx 45 \text{ meV}$, $R_{\pm}^* \approx 25 \text{ meV}$.
- (7.14) $R^* = (m_e^*/m_0\epsilon_r^2) \times R_H$. $m_e^* = 0.036m_0$.

- (7.15) Transitions from acceptor levels above the valence band to the conduction band. Acceptor energy $E_A \sim 8 \text{ meV}$.
- (7.16) Raman scattering from plasmon modes: $N = 4.2 \times 10^{24} \text{ m}^{-3}$.
- (7.17) $7.2 \times 10^{23} \text{ m}^{-3}$.

Chapter 8

- (8.1) $E_1 = (1/2)\hbar\Omega$, $E_2 = (3/2)\hbar\Omega$, $E_3 = (5/2)\hbar\Omega$, $a = (\hbar/m\Omega)^{1/2}$.
- (8.2) $d \approx 6.7 \times 10^{-10} \text{ m}$, which corresponds to about seven carbon–carbon bonds.
- (8.3) 4×10^{-2} , 1.6×10^{-3} , 1.4×10^{-5} .
- (8.4) 10.2 eV. The ground state of the molecule is more strongly bound than the excited state, and hence the transition energy is larger.
- (8.5) (i) The van der Waals interaction energy varies as r^{-6} . (ii) Find the point where $dU/dr = 0$.
(iii)
- $$U(r) = U(r_0) + (1/2) \left(\frac{d^2U}{dr^2} \right)_{r=r_0} (r-r_0)^2 + \dots$$
- $$\Omega^2 = (18B^2/A\mu)(B/2A)^{1/3}.$$
- (8.6) S_1 at 4.64 eV, $\Omega/2\pi = 3 \times 10^{13} \text{ Hz}$.
- (8.7) The configuration diagram is similar to Fig. 8.7, but with two excited states. S_1 state: energy = 5.7 eV, vibrational splitting = 0.11 eV, turning point of $n = 6$ level aligned with Q_0 . S_2 state: energy 7.3 eV, vibrational splitting 0.13 eV, turning point of $n = 5$ level aligned with Q_0 .
- (8.8) Spin-orbit coupling mixes S and L , so that the triplet states contain a small admixture of singlet character through mixing with common L states.

- (8.9) Phosphorescence from a triplet state at 1.6 eV.
- (8.10) Both give $\hbar\Omega \approx 0.17 \text{ eV}$.
- (8.11) There will be a vibronic band of width $\sim 1 \text{ eV}$ extending from 3.1 eV downwards, with three or four peaks at energies $(3.1 - n\hbar\Omega)$, where $\hbar\Omega \approx 0.17 \text{ eV}$.
- (8.12) Broad vibronic band from 1.9 eV down to about 1.0 eV. Peaks at 1.9 eV and 1.7 eV.
- (8.13) The absorption and photoconductivity thresholds correspond to the exciton and band gap energies respectively, which implies $E_b = 1.1 \text{ eV}$ for the exciton.
- (8.14) 693 nm. (There will be other Raman lines in addition to this one.)
- (8.15) Optical excitation creates only singlets, whereas electrical injection creates both singlets and triplets with a probability determined by their statistical weights, namely 1:3. Only singlets emit efficiently, and the population of these is lower by a factor of four in the case of electrical injection.
- (8.16) (i) See Exercise 8.15. (ii) 5.6 mW. (iii) 11 %. The efficiency of a real device would be much lower, mainly due to the difficulty of collecting the photons, which are emitted in all directions.

Chapter 9

- (9.1) The solution for a 1-D infinite potential well is given in Section 6.3.2. In a cube the motion is quantized in three dimensions, and the energies for the x , y and z directions just add together.
- (9.2) Equation 9.4 predicts $E = 0.28/a^2$. The experimental energies are lower because a real F-centre is not a rigid cubic box.

- (9.3) Either calculate $h\nu = 2.6 \text{ eV}$ from eqn 9.5, or just read $h\nu \approx 2 \text{ eV}$ from Fig. 9.5.
- (9.4) $E = (\hbar^2\pi^2/2m_0b^2)(n_x^2 + n_y^2 + n_z^2/4)$ and $h\nu = 3h^2/32m_0b^2$. The frequency is half the value given in eqn 9.5 when $b = \sqrt{2}a$, which is appropriate for an F_2^+ centre. The experimental ratio is about 0.4, which is good agreement considering

the simplicity of the model.

- (9.5) (a) $r_{3d}/r_{4f} \sim (3^2/25)/(4^2/65) = 1.5$.
 (b) Transition metal ions have lost the outermost $4s$ electrons, whereas the $4f$ orbitals of the rare earths are inside the filled $5s$ and $5p$ orbitals.
- (9.6) (i) The x , y and z directions are all equivalent, and so the p_x , p_y and p_z orbitals must all experience the same interaction energy with the crystal.
 (ii) The z direction is now different, and so the p_z orbits will have a different energy to the p_x and p_y states.
 (iii) The singlet is at higher energy because of the greater repulsion from the closer negative charges.
- (9.7) The relative populations of the 11502 cm^{-1} and 11414 cm^{-1} levels of the ${}^4F_{3/2}$ term are proportional to $\exp(-\Delta E/k_B T)$. The relative population of the 11502 cm^{-1} level therefore increases from 0.19 at 77 K to 0.66 at 300 K, and the emission intensity increases in proportion to these factors.
- (9.8) The stimulated emission rate exceeds the absorption rate if population inversion is present: see Section B.1.
- (9.9) 0.3 J. The laser stops working when 10 % of the atoms in the upper level have transferred to

the lower level, at which point there is no more population inversion.

- (9.10) (i) The spectrum is proportional to $|E(\omega)|^2$, where
- $$E(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} E(t) e^{i\omega t} dt,$$
- and $E(t) = \exp(-t^2/2\tau^2) e^{-i\omega_0 t}$.
 (ii) $\Delta t = 2\sqrt{\ln 2}\tau$. $\Delta\nu = \sqrt{\ln 2}/\pi\tau$.
- (9.11) Inhomogeneities in the glass cause local variations in the environment leading to line broadening through the coupling of the laser levels to the local crystal field. $\Delta t = 60\text{ fs}$ for Gaussian pulses.
- (9.12) The transition is parity forbidden. Phosphorescence.
- (9.13) The probability for phonon-assisted non-radiative decay increases with T . Equation 5.5 gives $\eta_R(77) = 0.78$ and $\eta_R(300) = 0.03$. The radiative efficiency is too low at 300 K to allow lasing.
- (9.14) $P = 3.2\text{ W}$, assuming that all the pump power is absorbed, and the radiative quantum efficiency is unity. The remaining 1.8 W goes as heat in the crystal.

Chapter 10

- (10.1) (a) Yes, (b) No, (c) No, (d) Yes, (e) Yes. Germanium and argon are non-polar materials.
- (10.2) Solve eqn 10.16 with $\epsilon_r = 0$.
- (10.3) $15\text{--}33\text{ }\mu\text{m}$.
- (10.4) (a) 98 %; (b) 84 %.
- (10.5) (i) $\nu_{TO} = 9.5\text{ THz}$, $\nu_{LO} = 10\text{ THz}$, (ii) $\epsilon_\infty = 9.5$, $\epsilon_{st} = 11.8$, (iii) About 15 ps. Lyddane-Sachs-Teller predicts $\nu_{LO}/\nu_{TO} = 1.11$, but the experimental ratio is slightly smaller. This is not significant, given that the broadening causes some uncertainty in the experimental values.
- (10.6) (a) $3.3 \times 10^6\text{ m}^{-1}$; (b) $1.1 \times 10^7\text{ m}^{-1}$.
- (10.7) The phonon lifetime decreases with T as the probability for anharmonic decay increases.
- (10.8) $7.8 \times 10^5\text{ m}^{-1}$.
- (10.9) (i) $m^{**} = 0.097 m_0$. (ii) $m^* = 0.092 m_0$.
- (10.10) The diamond crystal would have only one peak in the Stokes or anti-Stokes spectrum.
- (10.11) Stokes and anti-Stokes peaks from the optical phonon at 15.5 THz. $I(501.2\text{ nm})/I(528.6\text{ nm}) = 0.08$.
- (10.12) IR active, but not Raman active.
- (10.13) GaAs: $\hbar\nu_{TO} = 32.5\text{ meV}$, $\hbar\nu_{LO} = 35.5\text{ meV}$; InP: $\hbar\nu_{TO} = 37.1\text{ meV}$, $\hbar\nu_{LO} = 42.3\text{ meV}$; AlSb: $\hbar\nu_{TO} = 38.7\text{ meV}$, $\hbar\nu_{LO} = 41.1\text{ meV}$; GaP: $\hbar\nu_{TO} = 45.1\text{ meV}$, $\hbar\nu_{LO} = 50.0\text{ meV}$. The small shift of a few wave numbers compared to the infrared data for GaAs in Fig. 10.5 is caused by the slight decrease of the optical phonon frequencies between 4 K and 300 K.
- (10.14) Apply conservation of momentum, with $k_1 = k_2 = n\omega/c$. $v_s = 810\text{ m s}^{-1}$.
- (10.15) (i) The negative term is the total Coulomb attraction, with the Madelung constant accounting

for the contributions of the positive and negative ions from the whole crystal. The positive term represents the short range repulsive force due to the Pauli exclusion principle when the electron wave functions overlap.

(ii) r_0 is the value for which $dU/dr = 0$.

(iii) The Taylor series about r_0 is:

$$U(r) = U(r_0) + (1/2)(d^2U/dr^2)_{r=r_0}(r - r_0)^2 + (1/6)(d^3U/dr^3)_{r=r_0}(r - r_0)^3 + \dots$$

Take $x = r - r_0$ to put this in the form of eqn 10.34, with $U(x)$ defined relative to the minimum at r_0 . $C_3 = -22\alpha e^2/3\pi\epsilon_0 r_0^4$.

(10.16) 6 ps, assuming a Lorentzian line shape.

Chapter 11

(11.1) $\mathcal{E} = Ze/4\pi\epsilon_0 r_n^2$. For the outer $3s$ and $3p$ electrons in silicon use $Z = 4$ and $n = 3$ to obtain a value of $\sim 5 \times 10^{11} \text{ V m}^{-1}$.

(11.2) (a) $6.2 \times 10^7 \text{ V m}^{-1}$, (b) $1.6 \times 10^5 \text{ V m}^{-1}$.

(11.3) Only with the field applied.

(11.4) (a) No, (b) yes, (c) no, (d) no, (e) yes, (f) yes. The second-order nonlinear susceptibility is zero if the material has an inversion centre.

(11.5) (i) N_2 cannot increase beyond $N_0/2$ because there is no net absorption when the populations are equal. (ii) The rate equations are $\dot{N}_1 = -B_{12}u_v(N_1 - N_2)$ and $\dot{N}_2 = B_{12}u_v(N_1 - N_2)$. Subtract these to obtain $\Delta\dot{N} = -2B_{12}u_v\Delta N$ where $\Delta N = N_1 - N_2$. Then integrate with $\Delta N(0) = N_0$ to obtain the required result, which implies that the populations will eventually equalize no matter how weak the laser beam is. This misleading conclusion arises from neglecting spontaneous emission and transitions to other levels.

(11.6) $P_x^{(2)} = d_{14}2\mathcal{E}_y\mathcal{E}_z = 0$, and $P_y^{(2)} = d_{25}2\mathcal{E}_z\mathcal{E}_x = 0$. Assume that the beam makes an angle θ with

the x axis and then maximize $P_z^{(2)} = d_{36}2\mathcal{E}_x\mathcal{E}_y$.

(11.7) (i) $\Delta\Phi_{x'} = -\Delta\Phi_{y'} = (2\pi L/\lambda)(\frac{1}{2}n_0^3r_{41}\mathcal{E}_z)$, where L is the length of the crystal. $\Delta\Phi = \Delta\Phi_{x'} - \Delta\Phi_{y'}$ gives the result with $\mathcal{E}_z L = V$.
(ii) 44 kV.

(11.8) 52°.

(11.9) $\Delta\alpha = (\alpha_0/I_s)I \propto \Delta\epsilon_2$, and $\Delta\epsilon_2 \propto \text{Im}(\chi^{(3)})I$. Hence $(\alpha_0/I_s) \propto \text{Im}(\chi^{(3)})$.

(11.10) Choose z as the direction of propagation and x as the polarization vector so that $\mathcal{E}_y = \mathcal{E}_z = 0$. The only non-zero term is $P_x^{(3)} = \epsilon_0\chi_{xxxx}\mathcal{E}_x^3$, which implies that \mathbf{P} is parallel to \mathbf{E} .

(11.11) 50 W.

(11.12) 0.06 eV.

(11.13) Follow Example 2.1 to work out the magnitude of the local maximum in n below the absorption line. We then find $|\Delta n| = 0.027$ if we assume that this local maximum is completely saturated.

(11.14) $1.8 \times 10^{23} \text{ m}^{-3}$. I_s is the intensity required to produce this carrier density, namely $4 \times 10^7 \text{ W m}^{-2}$.

Bibliography

Companion texts in the Oxford Masters Series

- Blundell, S. (2001). *Magnetism in condensed matter physics*. Clarendon Press, Oxford.
- Singleton, J. (2001). *Band structure and electrical properties of solids*. Clarendon Press, Oxford.

General solid state physics texts

- Ashcroft, N.W. and Mermin, N.D. (1976). *Solid state physics*. Saunders College Publishing, Philadelphia.
- Burns, G. (1985). *Solid state physics*. Academic Press, San Diego.
- Elliott, R.J. and Gibson, A.F. (1974). *An introduction to solid state physics and its applications*. Macmillan, New York.
- Ibach, H. and Luth, H. (1995). *Solid-state physics*. Springer-Verlag, Berlin.
- Kittel, Charles (1996). *Introduction to solid state physics* (7th edn). Wiley, New York.
- Rosenberg, H.M. (1988). *The solid state* (3rd edn). Clarendon Press, Oxford.

Electromagnetism and optics texts

- Bleaney, B.I. and Bleaney, B. (1976). *Electricity and magnetism* (3rd edn). Clarendon Press, Oxford. Reissued in two volumes in 1989.
- Born, M. and Wolf, E. (1999). *Principles of optics* (7th edn). Cambridge University Press, Cambridge.
- Duffin W.J. (1990). *Electricity and magnetism* (4th edn). McGraw-Hill, London.
- Good R.H. (1999). *Classical electromagnetism*. Saunders College Publishing, Fort Worth.
- Grant I.S. and Phillips W.R. (1990). *Electromagnetism* (2nd edn). Wiley, New York.
- Hecht, Eugene (1998). *Optics* (3rd edn). Addison-Wesley, Reading, Massachusetts.
- Klein, M.V. and Furtak, T.E. (1986). *Optics* (2nd edn). Wiley, New York.
- Lorrain P., Corson D.R. and Lorrain F. (2000). *Fundamentals of electromagnetic phenomena*. W.H. Freeman, Basingstoke.

Quantum mechanics and atomic physics texts

- Corney, Alan (1977). *Atomic and laser spectroscopy*. Clarendon Press, Oxford.
- Gasiorowicz, Stephen (1996). *Quantum physics* (2nd edn). Wiley, New York.
- Schiff, L.I. (1969). *Quantum mechanics*. McGraw-Hill, New York.
- Woodgate, G.K. (1980). *Elementary atomic structure* (2nd edn). Clarendon Press, Oxford.

Band theory and semiconductor physics texts

- Harrison, W. (1999). *Elementary electronic structure*. World Scientific, Singapore.
- Klingshirn, C.F. (1995). *Semiconductor optics*. Springer-Verlag, Berlin.
- Pankove, J.I. (1971). *Optical processes in semiconductors*. Dover, New York.
- Seeger, K. (1997). *Semiconductor physics* (6th edn). Springer-Verlag, Berlin.
- Yu, P.Y. and Cardona, M. (1996). *Fundamentals of semiconductors*. Springer-Verlag, Berlin.

Optoelectronics and laser physics texts

- Bhattacharya, P. (1997). *Semiconductor optoelectronic devices* (2nd edn). Prentice Hall, New Jersey.
- Chuang, S.L. (1995). *Physics of optoelectronic devices*. Wiley, New York.
- Silfvast, W.T. (1996). *Laser fundamentals*. Cambridge University Press, Cambridge.
- Svelto, O. (1998). *Principles of lasers* (4th edn). Plenum Press, New York.
- Sze, S.M. (1981). *Physics of semiconductor devices* (2nd edn). Wiley, New York.
- Sze, S.M. (1985). *Semiconductor devices*. Wiley, New York.
- Wilson, J. and Hawkes, J. (1998). *Optoelectronics* (3rd edn). Prentice Hall Europe, London.
- Yariv, Amnon (1997). *Optical electronics in modern communications* (5th edn). Oxford University Press, New York.

Further reading for specific chapters

Chapter 1: Introduction

- Nye, J.F. (1957). *Physical properties of crystals*. Clarendon Press, Oxford.

Chapter 2: Classical propagation

- Palik, E.D. (1985). *Handbook of the optical constants of solids*. Academic Press, San Diego.

Chapter 4: Excitons

- Dexter, D.L. and Knox, R.S. (1965). *Excitons*. Wiley, New York.
- Griffin, A., Snoke, D.W. and Stringari, S. (1995). *Bose-Einstein condensation*. Cambridge University Press, Cambridge.
- Ketterle, W. (1999). Experimental studies of Bose-Einstein condensation. *Physics Today*, **52**, No. 12, 30.
- Mandl, F. (1988). *Statistical physics* (2nd edn). Wiley, Chichester.
- Rashba E.I. and Sturge, M.D. (1982). *Excitons*. North Holland, Amsterdam.
- Reynolds, D.C. and Collins, T.C. (1981). *Excitons: their properties and uses*. Academic Press, New York.
- Song, K.S. and Williams, R.T. (1993). *Self-trapped excitons*. Springer-Verlag, Berlin.
- Wieman, C.E. (1997). The creation and study of Bose-Einstein condensation in a dilute atomic vapour. *Phil. Trans. R. Soc. Lond. A*, **355**, 2247.

Chapter 5: Luminescence

- Landsberg, P.T. (1991). *Recombination in semiconductors*. Cambridge University Press, Cambridge.
- Nakamura, F., Pearton S. and Fasol, G. (2000). *The blue laser diode* (2nd edn). Springer-Verlag, Berlin.
- Shah, J. (1999). *Ultrafast spectroscopy of semiconductors and semiconductor nanostructures* (2nd edn). Springer-Verlag, Berlin.
- Voos, M., Leheney, R.F. and Shah, J. (1980). Radiative recombination. In *Handbook on semiconductors*, Vol. 2 (ed. M. Balkanski). North Holland, Amsterdam, pp 329–416.

Chapter 6: Semiconductor quantum wells

- Bastard, G. (1990). *Wave mechanics applied to semiconductor heterostructures*. Wiley, New York.
- Bimberg, D., Grundmann M. and Ledentsov, N.N. (1999). *Quantum dot heterostructures*. Wiley, Chichester.
- Blood, P (1999). Visible-emitting quantum well lasers. In *Semiconductor quantum electronics* (ed. A. Miller, M. Ebrahimzadeh and D.M. Finlayson). Institute of Physics, Bristol, p. 193.
- Esaki, L. and Tsu, R. (1970). Superlattice and negative differential conductivity in semiconductors. *IBM Journal of Research and Development*, **14**, 61.
- Fox, A.M. (1996). Optoelectronics in quantum well structures. *Contemporary Physics*, **37**, 111.
- Grahn, H.T. (1995). *Semiconductor superlattices*. World Scientific, Singapore.
- Harrison, P. (1999). *Quantum wells, wires and dots*. Wiley, Chichester.
- Helm, M (2000). *Long wavelength infrared emitters based on quantum wells and superlattices*. Gordon and Breach, Amsterdam.
- Jaros, M. (1989). *Physics and applications of semiconductor microstructures*. Clarendon Press, Oxford.

- Kelly, M.J. (1995). *Low-dimensional semiconductors*. Clarendon Press, Oxford.
- Liu, H.C. and Capasso, F. (2000a). *Intersubband transitions in quantum wells: physics and device applications I*, Semiconductors and Semimetals, Vol. 62 (series eds R.K. Willardson and E.R. Weber). Academic Press, San Diego.
- Liu, H.C. and Capasso, F. (2000b). *Intersubband transitions in quantum wells: physics and device applications II*, Semiconductors and Semimetals, Vol. 66 (series eds R.K. Willardson and E.R. Weber). Academic Press, San Diego.
- Singh, J. (1993). *Physics of semiconductors and their heterostructures*. McGraw-Hill, New York.
- Weisbuch, C. and Vinter, B. (1991). *Quantum semiconductor structures*. Harcourt, San Diego.
- Woggon, U. (1996). *Optical properties of semiconductor quantum dots*. Springer-Verlag, Berlin.

Chapter 7: Free electrons

- Pidgeon, C.R. (1980). Free carrier optical properties of semiconductors. In *Handbook on Semiconductors*, Vol. 2 (ed. M. Balkanski). North Holland, Amsterdam, chapter 5.

Chapter 8: Molecular materials

- Banwell, C.N. and McCash, E.M. (1994). *Fundamentals of molecular spectroscopy* (4th edn). McGraw-Hill, London.
- Bradley, D. (1996). Electroluminescent polymers: materials, physics and device engineering. *Current Opinion in Solid State and Materials Science*, **1**, 789.
- Friend, R.H., Gymer, R.W., Holmes, A.B., Burroughes, J.H., Marks, R.N., Taliani, C., Bradley, D.D.C., Dos Santos, D.A., Bredas, J.L., Lögglund, M. and Salaneck, W.R. (1999). Electroluminescence in conjugated polymers. *Nature*, **397**, 121.
- Haken, H. and Wolf, H.C. (1995). *Molecular physics and elements of quantum chemistry*. Springer-Verlag, Berlin.
- Klessinger, M. and Michl, J. (1995). *Excited States and photochemistry of organic molecules*. VCH Publishers, New York.
- Mueller, G. (2000). *Electroluminescence II*, Semiconductors and Semimetals, Vol. 65 (series eds R.K. Willardson E.R. Weber). Academic Press, San Diego.
- Pope, M. and Swenberg, C.E. (1999). *Electronic processes in organic crystals and polymers* (2nd edn). Oxford University Press, New York.
- Schön, J.H., Kloc, Ch., Dodabalapur, A. and Batlogg, B. (2000). An organic solid state injection laser. *Science*, **289**, 599.
- Wright, J.B. (1995). *Molecular crystals* (2nd edn). Cambridge University Press, Cambridge.

Chapter 9: Luminescence centres

- Di Bartolo, B. (1992). *Optical properties of excited states in solids*, NATO ASI Series B, Vol. 301. Plenum Press, New York.
- Henderson, B. and Imbusch, G.F. (1989). *Optical spectroscopy of inorganic solids*. Clarendon Press, Oxford.

Chapter 10: Phonons

- Houghton, J.T. and Smith, S.D. (1966). *Infra-red physics*. Clarendon Press, Oxford.
- Madelung, O. (1978). *Introduction to solid-state theory*. Springer-Verlag, Berlin.
- Mooradian, A. (1972). Raman spectroscopy of solids. In *Laser Handbook* Vol. II (ed. F.T. Arecchi and E.O. Schulz-duBois). North Holland, Amsterdam, p. 1409.
- Pidgeon, C.R. (1980). Free carrier optical properties of semiconductors. In *Handbook on Semiconductors*, Vol. 2 (ed. M. Balkanski). North Holland, Amsterdam, chapter 5.
- Pope, M. and Swenberg, C.E. (1999). *Electronic processes in organic crystals and polymers* (2nd edn). Oxford University Press, New York.
- Shah, J. (1999). *Ultrafast spectroscopy of semiconductors and semiconductor nanostructures* (2nd edn). Springer-Verlag, Berlin.
- Song, K.S. and Williams, R.T. (1993). *Self-trapped excitons*. Springer-Verlag, Berlin.

Chapter 11: Nonlinear optics

- Butcher, P.N. and Cotter, D. (1990). *The elements of nonlinear optics*. Cambridge University Press, Cambridge.
- Chemla, D.S. (1985). Excitonic optical nonlinearities. *J. Opt. Soc. Am. B*, **2**, 1135–1243.
- Nye, J.F. (1957). *Physical properties of crystals*. Clarendon Press, Oxford.
- Schmitt-Rink, S., Chemla, D.S. and Miller, D.A.B. (1989). Linear and nonlinear optical properties of semiconductor quantum wells. *Adv. Phys.*, **38**, 89.

Symbols

A	area	I_{th}	threshold current
\mathbf{A}	magnetic vector potential	\mathbf{j}	current density
A_{ij}	Einstein A coefficient	j	angular momentum quantum number
a	unit cell dimension	\mathbf{k}	wave vector
a_X	exciton Bohr radius	k_F	Fermi wave vector
b	barrier thickness	K	Kerr constant
\mathbf{B}	magnetic flux density	K_s	spring constant
B_{ij}	Einstein B coefficient	l	length
C	capacitance	l_c	coherence length
d	period	l_i	intrinsic region thickness
d	quantum well thickness	L	length
d_{ij}	nonlinear optical coefficient tensor	l	orbital quantum number
\mathbf{D}	electric displacement	L	orbital quantum number
E	energy	m	magnetic quantum number
E_b	binding energy	m_J	magnetic quantum number
E_g	band gap energy	m_s	spin quantum number
E_F	Fermi energy	m	mass
\mathbf{E}	electric field	m^*	effective mass
$f_{e,h}$	electron or hole Fermi–Dirac distribution function	m^{**}	polaron mass
f_{BE}	Bose–Einstein occupancy factor	m_e^*	electron effective mass
f_j	oscillator strength	m_h^*	hole effective mass
F	finesse	m_{hh}^*	heavy hole effective mass
g_i	degeneracy of atomic level i	m_{lh}^*	light hole effective mass
$g(E)$	density of states in energy space	m_{so}^*	split-off hole effective mass
$g_c(E)$	conduction band density of states	M	matrix element
$g_v(E)$	valence band density of states	\mathbf{M}	magnetization
$g(k)$	density of states in wave vector space	n	refractive index
g_e	electron g-factor	\tilde{n}	complex refractive index
g_h	hole g-factor	n_o	ordinary refractive index
\mathbf{G}	reciprocal lattice vector	n_e	extraordinary refractive index
H	Hamiltonian	n_0	linear refractive index
H_0	unperturbed Hamiltonian	n_2	nonlinear refractive index
H'	perturbing Hamiltonian	n	quantum number
\mathbf{H}	magnetic field	N	number of atoms/particles per unit volume
I	intensity	N_e	electron density
I_s	saturation intensity	N_h	hole density
I_{pc}	photocurrent	N_{Mott}	Mott density
I_{in}	injection current		

N	photon number density	δ	skin depth
\hat{O}	quantum mechanical operator	Δ	split-off hole band energy
p	dipole moment	ϵ_r	relative dielectric constant
P	momentum	$\tilde{\epsilon}_r$	complex relative dielectric constant
\mathbf{P}	polarization	ϵ_1	real part of the complex relative dielectric constant
P	optical power	ϵ_2	dielectric constant
q	electric charge		imaginary part of the complex relative dielectric constant
\mathbf{q}	phonon wave vector		static relative dielectric constant
Q	generalized position coordinate	ϵ_{st}	high frequency relative dielectric constant
\mathbf{r}	position vector	ϵ_∞	quantum efficiency
r_p	polaron radius	η	radiative quantum efficiency
R	reflectivity	η_R	angle
R	electrical resistance	θ	imaginary part of the complex refractive index
R_X	exciton Rydberg constant	κ	wavelength
S	spin quantum number	λ	de Broglie wavelength
t	time	λ_{deB}	reduced mass
T	transmissivity	μ	relative magnetic permeability
T	temperature	μ_r	frequency
T_c	critical temperature	v	wave number
T_L	lattice temperature	\bar{v}	LO phonon frequency at $q = 0$
T_m	melting temperature	ν_{LO}	TO phonon frequency at $q = 0$
\mathbf{T}	lattice translation vector	ρ	density of states
u	envelope function within a Bloch function	ϱ	electrical charge density
$u(v)$	energy density of an electromagnetic wave at frequency v	σ	electrical conductivity
U	potential energy	σ_s	scattering cross-section
v	velocity of light in a medium	τ	lifetime
v_g	group velocity	τ_R	radiative lifetime
v_s	velocity of sound	τ_{NR}	non-radiative lifetime
\mathbf{v}	electron velocity	ϕ	azimuthal angle in spherical polar coordinates
V	volume	Φ	optical phase
V	voltage	φ	wave function
V_{bi}	built-in voltage	χ	electric susceptibility
W	transition rate	χ_a	electric susceptibility per atom
x	position coordinate	χ_M	magnetic susceptibility
y	position coordinate	ψ	wave function
z	position coordinate	Ψ	wave function
Z	impedance	ω	angular frequency
Z	atomic number	ω_c	cyclotron frequency
α	absorption coefficient	ω_p	plasma frequency
α_{ep}	electron–phonon coupling constant	Ω	phonon angular frequency
γ	damping rate	Ω_{LO}	LO phonon angular frequency at $q = 0$
γ_v	gain coefficient	Ω_{TO}	TO phonon angular frequency at $q = 0$
γ_{th}	threshold gain coefficient for laser oscillation		

Index

Γ point, 56, 276
 π bond, 167
 π conjugated materials, 13, 168
 π electron, 13, 89, 166–8
 π orbital, 13, 167–8, 177
 $\pi \rightarrow \pi^*$ transition, 168, 177, 180
 σ bond, 167

absorbance, 4
absorption
apparatus, 68
aromatic hydrocarbon, 177–9
atomic, 26–8, 33–4, 50
coefficient, 3
definition of, 2, 235, 263–4
dipole, 27–37
direct, 51–63
edge, 9, 50, 54–66, 127, 138, 160
electronic, 36
excitonic, 79–89, 130
extinction coefficient, relationship to, 5
F-centre, 190
far-infrared, 216–17
free carrier, 10, 147, 155–7
impurity, 158–60
in electric field, 60, 82, 132
in magnetic field, 61, 83
indirect, 63–6
infrared, 9–11, 20, 28, 38, 135, 157, 159, 212, 216–17
interband, 18, 38, 49–73, 151–4, 247
intersubband, 135
intervalence band, 157
lattice, 9, 205, 212
molecular, 169–77
nonlinear, 85, 234–6, 249–50
phonon, 9, 205, 212
optical density, relationship to, 4
polymer, 179–81
quantum dot, 137
quantum theory of, 263–70
quantum well, 124–33
rate, 264, 266, 269
saturable, 236, 249–50
two-dimensional, 127
ultraviolet, 38, 176
vibrational, 9, 36

vibronic, 170–7, 186–9
absorption spectra
ammonia, 176
CdS quantum dot, 137
GaAs, 64, 81, 83, 85, 130
GaAs quantum well, 129, 130, 132, 141
GaInAs quantum well, 250
GaN, 96
germanium, 65
InAs, 59
KF:F₂⁺, 191
LiF, 88
MeLPPP, 181
NaCl, 88
polyfluorene, 14
polydiacetylene (PDA), 180
pyrene (C₁₆H₁₀), 88
pyrromethene 567 (P-567) dye, 177
silica (SiO₂), 37
silicon, 64, 66
silicon, n-type, 159
Ti:sapphire, 196
AC conductivity, 146
acceptor, 105, 154, 273
acetylene (C₂H₂), 179
acoustic phonon, *see* phonon, acoustic active region, 103, 109, 133, 182, 280
Ag, *see* silver
AgCl, 216–17
Al₂O₃, *see* sapphire
AlGaAs, 97, 104
AlGaInP, 104
AlSb, 221, 225
Alferov, Zhores, 110
alkali halides, 87, 189–91
alkali metals, ultraviolet transmission 149
allowed transition, 55, 95, 126, 169–70, 178, 270
aluminium (Al), 149–52, 271, 276
ammonia (NH₃), 176
amorphous materials, 13, 18, 166, 182, 207
Ampere's law, 257
amplification of light, 107
amplification, nonlinear, 238, 246
anisotropic materials, 10, 228–9, 242–3, 256
anisotropy, optical, 17, 42, 256
anharmonic oscillator, 223, 231–4

anharmonicity, 223, 231
anion dopant, 192
anode, 182
anomalous dispersion, 41, 248
anthracene (C₁₄H₁₀), 88, 177–9
anti-bonding orbital, 55, 168
anti-Stokes scattering, 218–19
anticrossing, 214, 217
argon ion laser, 197
aromatic hydrocarbon, 13, 88, 166, 177–9
atomic oscillator, 26
attenuation of light, 2–7, 146
Au, *see* gold

back-scattering geometry, 219
band-filling nonlinearity, 249–50
band gap, 9–11, 50, 105, 217, 271–5
band theory of solids, 18, 271–9
band structure, 271–9
aluminium, 151, 276
copper, 153, 276
direct gap, 50, 56
four-band (Kane) mode, 57, 277–8
GaAs, 56, 278
germanium, 64, 277–8
III-V semiconductor, 57, 278
indirect gap, 50, 277
silicon, 67, 276–7
bands, electronic, 18, 50, 186, 271–6
bands, vibronic, 19, 186–9
barrier potential, 121
Beer's law, 3–6, 27
benzene (C₆H₆), 13, 167
benzene ring, 13, 167, 177
beryl (Be₃Al₂Si₆O₁₈), 195
biaxial crystal, 43
biexciton, 85
binary logic, optical, 252
bias voltage, 280–1
birefringence, 17, 42–5, 238, 242–3
birefringent materials, 10, 43, 242
bistability, optical, 251
black body radiation, 265
Bloch oscillations, 135–6
Bloch functions, 18, 52, 125, 276
Bloch waves, 20
Bloch's theorem, 18, 52, 275–6
Bohr radius, 78

Boltzmann statistics, 100
 Boltzmann's law, 265
 bonding orbital, 55, 168
 Born-Oppenheimer approximation, 171–4
 Bose-Einstein condensation, 86
 Bose-Einstein statistics, 65, 86
 bound electrons, 25–38, 230–4
 boundary conditions, quantum well, 120–2
 boundary conditions, optical, 261
 breathing mode, 188
 Brillouin, L., 222
 Brillouin scattering, 218, 222
 Brillouin zones, 273–6
 broadening, lifetime, 223
 broadening, thermal, 178, 194
 broadening due to damping, 30–3
 broadening due to monolayer fluctuations, 134
 broadening due to scattering, 63, 81
 built-in voltage, 72, 82, 132, 280–1

$[\text{CH}_2]_n$, *see* ethylene
 C_2H_2 , *see* acetylene
 C_2H_4 , *see* ethylene
 C_4H_8 , *see* cyclobutane
 C_6H_6 , *see* benzene
 $\text{C}_{14}\text{H}_{10}$, *see* anthracene
 $\text{C}_{16}\text{H}_{10}$, *see* pyrene
 $\text{C}_{18}\text{H}_{12}$, *see* tetracene
 cadmium selenide (CdSe), 9–11, 206, 212
 cadmium sulphide (CdS), 137
 calcite, 43–44
 carbon, 167
 carbon monoxide laser, 251
 cathode, 182
 cathode ray tube, 200
 cation vacancy, 189–91
 centrosymmetric crystal, 220, 239–40
 charge density, 256
 chromium ion (Cr^{3+}), 15, 187, 192, 195, 196, 198
 classical models, 20, 25–46, 143–5, 206–9, 230–4
 classical statistics, *see* Boltzmann statistics
 classification of optical processes, 1–5
 Clausius–Mossotti relationship, 40
 Co^{2+} , 198
 coefficient,
 absorption, 3
 extinction, 5
 gain, 108
 reflection, 2
 transmission, 3
 coherence length, nonlinear, 241
 collective excitations, 19, 160, 214, 216, 246
 colour centres, 189–91, 198–9
 colour glass filters, 11, 14, 138
 complex refractive index, 5–8, 68, 260–1

complex dielectric constant, 5–8, 31, 144, 208, 212
 conduction band, 50–7, 95–7, 271–3, 276
 conductivity, electrical, 146, 260
 conducting medium, 260–1, 271–3
 configuration coordinates, 171–4, 188
 configuration diagrams, 171–4, 188, 190
 confinement energy, 116, 120, 136, 168
 conjugated materials, 13, 89, 167–9, 177–83
 conjugated polymer, 13, 166, 168, 179–83
 copolymer, 182
 copper (Cu), 12, 152–3, 276
 copper chloride (CuCl), 85, 86
 copper oxide (Cu_2O), 86
 correlated photon pair, 238
 Coulomb gauge, 257, 267
 covalent bonds, 13, 55, 166–8, 206
 covalent crystal, 55, 206, 210
 Cr^{3+} , *see* chromium ion
 critical points, 68, 152
 crystal
 anisotropic, 10, 17, 43, 228–9, 242–3, 256
 biaxial, 43
 centrosymmetric, 220, 239–40
 class, 16, 43, 229, 239–40
 nonlinear, 229, 237–43
 uniaxial, 17, 43, 229, 240, 242–3
 crystal field effect, 15, 17, 192–6
 crystal symmetry, 16–8, 43, 220, 229, 239–40
 cubic crystal, 43, 105, 240, 276, 278
 cuprite (Cu_2O), 278–9
 current density, 146, 257, 260
 cyclic conjugated molecule, 168
 cyclobutane (C_4H_8), 179
 cyclotron energy, 83, 216
 cyclotron frequency, 61, 216
 cyclotron resonance, 216–17

d bands, 152–153, 276
 damping, 30, 145–6, 208, 211, 222, 232
 damping rate, 30, 146, 208, 222
 damping time, 146, 223
 de Broglie wavelength, 86, 116, 273
 defects, optically active, 186, 189–91
 degeneracy,
 atomic, 265, 269
 electronic, 101, 106
 lifting of, 17, 121, 131, 193
 spin 275
 degenerate four-wave mixing, 244–5
 degrees of freedom, 116, 175
 delayed fluorescence, 179, 270
 delocalized states, 19, 77, 166–8, 186
 density of states, 19, 51–54, 94, 99, 128, 133, 137, 152–3, 187, 234, 266–7
 density of states, 0-D, 137
 density of states, 2-D, 128, 137
 depletion region, 106, 280–1
 detector, *see* photodetector
 diamagnetism, 84
 diamond, 10, 210, 271, 277
 diamond structure, 56, 277
 diatomic molecule, 171–3
 dielectric constant, 6, 255
 bound electron, 35
 complex, 5–8
 free carrier, 144, 146, 155
 high frequency, 208
 lattice, 208
 linear, 227
 nonlinear, 228, 234
 relative, 6, 255–6
 static, 32, 208
 tensor, 44
 difference frequency mixing, 237–9
 dimer, 179
 diode, *see* p-n diode or p-i-n diode
 diode laser, 107, 134, 179, 197
 dipole, atomic, 26–8
 dipole, electric, 26, 131
 dipole, Hertzian, 27
 dipole moment, 26, 52, 126, 175, 255, 268
 dipole oscillator model, 25–40, 143–5, 154–6, 206–9
 Dirac notation, 52
 direct absorption, 51–63
 direct band gap, 50, 95, 98, 103–5, 277–9
 dispersion
 anomalous, 41, 248
 definition of, 3
 group velocity, 42
 glass, 38, 40–2
 microscopic origin, 40–2
 normal, 41, 241–3, 248
 optical fibre, 41–2, 248
 SiO_2 glass, 40–2
 displacement, *see* electric displacement
 displacement current, 257
 display technology, 182
 divalent metal, 271
 donor, 154, 158, 272–3
 donor levels, 158, 272–3
 doped glass, 14–5, 137–8
 doped insulator, 14–5, 192–6
 doped semiconductor, 29, 80, 143, 154–62, 272–3
 double refraction, 43
 double spectrometer, 221
 down conversion, 238–9
 Drude, Paul, 29, 147
 Drude model, 143, 148
 Drude-Lorentz model, 29, 143
 dye, 13, 176–7
 dynamic crystal field effect, 192–3

- effective charge, 207
 effective mass, 58, 149, 154, 159, 216–17,
 275, 277–8
 effective mass anisotropy, 159, 277
 Einstein *A* and *B* coefficients, 93, 234–6,
 263–6, 269
 Einstein–Podolsky–Rosen experiment, 238
 elastic light scattering, 2, 28, 218
 electric dipole,
 approximation, 268
 definition of, 26
 interaction, 52, 125, 268
 matrix element, 268
 moment, 26, 52, 126, 131, 175, 255,
 268
 selection rules, 269–70
 transition, 52, 54, 125–6, 169–70, 175,
 194, 268–70
 electric displacement, 31, 144, 209, 255–8
 electric field strength, 255–8, 281
 electric permittivity, 255
 electric quadrupole transition, 270
 electric susceptibility, 31, 39, 208, 227, 233,
 255–6
 electric susceptibility tensor, 42
 electrical confinement, 109
 electrical conductivity, 146, 271–3
 electrical injection, 103, 134, 182
 electroluminescence, 92, 103–10, 133–4,
 179, 181–3
 electroluminescent devices, 103, 134, 179,
 182
 electromagnetic induction, 257
 electromagnetic momentum, 267
 electromagnetic waves, 258–62
 electromagnetism, 255–62
 electron
 σ , 167
 π , 13, 167
 bound, 25–36, 143, 209, 230–1
 core, 38
 d, 152–3, 192–3, 195–6, 198, 276
 f, 192–4, 198, 200
 free, 29, 143–64
 p, 55, 57, 167, 276
 s, 55, 152–3, 167, 276
 valence, 55, 148, 154, 167, 271–6
 electron density of states, 54, 99, 128
 electron effective mass, 277–8
 electron injection, 98, 103, 134, 182
 electron-hole droplet, 85
 electron-hole overlap, 126, 132, 137
 electron-hole pair, 50, 76, 80, 95, 182
 electron-hole plasma, 85
 electron-hole recombination, 95, 101, 106,
 133, 182
 electron-phonon coupling, 98, 101, 136,
 186–7, 191, 193, 194, 215–18
 electron-phonon coupling constant, 216
 electronic bands, 18, 56–7, 271–9
 electronic states in quantum wells, 119–24
 electronic transition, 36
 electro-optic effect, 61, 131–3, 238–9, 246
 electroreflectance, 61
 electrostatics, 256–8
 electrostatic potential, 256–8
 emerald (Cr^{3+} : beryl), 195
 emission of light, *see* luminescence
 emission, spontaneous, 2, 4, 92, 107, 246,
 263–5
 emission, stimulated, 107, 264–5
 energy density of light wave, 259, 264
 energy independent relaxation time
 approximation, 156
 energy loss spectroscopy, 161
 envelope function, 52, 126, 276
 epitaxial growth, 103–5, 117, 138
 epitaxy, 104
 equation of motion, quantum mechanical,
 268
 equipartition of energy, 116
 Er^{3+} , 198–9
 Esaki, L., 115, 118, 135
 ethylene (C_2H_4), 167, 179
 Eu^{2+} , 200
 Eu^{3+} , 200
 excitations, elementary, 20, 161, 214, 216,
 218, 246
 exciton, 76–91
 absorption, 59, 79–82, 130, 250
 binding energy, 77
 Bose–Einstein condensation, 86
 collective excitation, 20
 diamagnetic shift, 84
 free, 76–86
 field ionization, 82
 Frenkel, 76, 86–9, 181, 218
 high densities, 84–6
 impurity screening, 80
 in electric field, 82
 in magnetic field, 83
 indirect, 80, 85
 interactions, 84
 luminescence, 100
 molecular, 88, 180
 nonlinear absorption, 85, 249–50
 phonon collisions, 81
 quantum dot, 138
 quantum well, 130–3
 quantum confined Stark effect, 131
 radius, 77
 Rydberg, 78
 saturation, 249–50
 self-trapped, 87, 218
 stability, 80
 tightly bound, 76, 86–9, 180
 Wannier–Mott, 76
- exciton data
 alkali halides, 87
 CdS quantum dot, 137
 GaAs, 81, 83, 85
 GaAs quantum well, 129, 130, 132
 GaInAs quantum well, 250
 II–VI semiconductors, 78
 III–V semiconductors, 78
 LiF, 88
 NaCl, 88
 polydiacetylene (PDA), 180
 pyrene ($\text{C}_{16}\text{H}_{10}$), 88
 rare gas crystals, 87
 experimental techniques, 68, 102, 220–1
 extraordinary ray, 43, 242–3
 extinction coefficient, 5
 extrinsic semiconductor, 272–3
- F8, *see* polyfluorene
 F-band, 190
 F-centres, 189–91
 F_2^+ centre, 191, 199
 Fabry–Perot etalon, 251
 Fabry–Perot interferometer, 222
 face-centred cubic crystal, 56, 276–7
 Faraday's law, 257
 far-infrared absorption, 216–7
 Fermi–Dirac statistics, 99, 236
 Fermi energy, 99, 101, 106, 151–3, 156–7,
 271–3, 280–1
 Fermi surface, 151
 Fermi's golden rule, 51, 94, 125, 152, 175,
 234, 266
 Feynman diagrams, 21, 220, 237, 243–4
 fibre laser, 198
 fibre amplifier, erbium-doped, 198–9
 fibre optics, 11, 42, 104–5, 134, 198–9, 248
 field ionization of excitons, 82, 132
 filter, colour glass, 11, 14, 138
 finesse, 251
 finite potential well, 121
 flash lamp pumping, 197
 fluorescence, 169, 178–9, 200, 270
 fluorescence, delayed, 179, 270
 fluorescence, prompt, 179, 270
 fluorescent lighting, 200
 forbidden transition, 132, 169, 194, 196,
 200, 270
 forward bias, 280–1
 four-wave mixing, 244–5
 Fourier transform spectrometer, 68
 fourth harmonic generation, 237–8
 Franck–Condon factor, 175, 189
 Franck–Condon principle, 173–6, 189
 Franz–Keldysh effect, 61, 83
 Franz–Keldysh oscillations, 62
 free carrier, *see* free electron

- free electron, 143–64
 absorption, 10, 155–7
 band dispersion, 274
 conductivity, 145–7
 Drude model, 148
 oscillator, 25, 29, 143–5
 plasma, 143
 reflectivity, 12, 143–5, 154–6
 free exciton, 76–86, 130–3, 250
 Frenkel exciton, 76, 86–9, 180–1, 218
 frequency doubling, 2, 237–9, 241
 frequency tripling, 238, 244
 frequency mixing, nonlinear, 237–9, 243
 Fresnel's equations, 261
 fundamental absorption edge, 9, 54–66, 80, 160
 gain coefficient, 108
 gallium antimonide (GaSb), 278
 gallium arsenide (GaAs)
 absorption, 64, 81, 83, 85, 129, 130, 132
 absorption in electric field, 83, 132
 absorption in magnetic field, 83
 band structure, 56, 278–8
 band gap, 63, 278
 crystal structure, 278
 effective mass, 217, 278
 electroluminescence, 104, 107, 134
 electron-phonon coupling, 216
 electronic configuration, 55
 exciton, 59, 79–85, 130–1
 Franz–Keldysh effect, 83
 infrared reflectivity, 211
 laser, 109–10, 134, 197
 light-emitting diode, 107, 134
 n-type, 161
 p-i-n diode, 82, 107, 281
 phonon, 211, 220–1
 phonon lifetime, 224
 photodiode, 71
 photoluminescence, 100
 plasmon, 161
 polaron, 216–7
 quantum well, 117, 123, 129–32, 135, 141
 Raman spectrum, 162, 221
 Raman active modes, 220
 gallium indium arsenide (GaInAs), 101, 134, 250
 gallium nitride (GaN), 96, 105, 201
 gallium phosphide (GaP), 103, 214–15, 221
 gauge, electromagnetic, 257
 Gauss's law, 256–7
 gemstones, 9
 germanium (Ge)
 absorption, 65
 absorption in magnetic field, 62
 band structure, 64, 277–8
 band gap, 65, 277–8
 crystal structure, 64, 277
 electronic configuration, 55
 effective mass, 277
 exciton, 85
 luminescence, 97
 melting point, 177
 n-type, 272
 phonon energies, 65, 210
 photodiode, 70
 transmission, 62
 glass,
 doped, 14–15, 137–8, 247
 nonlinear response, 247–9
 optical properties, 11–12, 37–8, 41
 refractive index, 11, 38, 41
 silica, 11–12, 37–8, 41
 stained, 11, 14, 138
 types, 12
 gold (Au), 12, 153, 164
 group theory, 54, 193, 195, 204, 220, 276
 group velocity, 38, 80, 274
 group velocity dispersion, 42, 248
 halophosphate phosphor, 200
 heavy hole band, 57, 157
 heavy hole transition, 57, 129, 132
 Heisenberg uncertainty principle, 115
 Hertz, Henrick, 26
 Hertzian dipole, 27
 heterojunction laser, 109–10
 heterostructure, 110, 117
 hexagonal crystal, 43, 105, 279
 hole, 50, 57, 154, 180, 190, 272–3, 275
 hole effective mass, 277–8
 hole injection, 105, 182
 HOMO level, 168, 181
 hopping, 86, 182, 217
 hot carrier effect, 101
 Hund's rules, 194
 hybridization, sp^2 , 167
 hydrogen molecule, 172
 hysteresis, optical, 251
 Huygens point source, 28
 Iceland spar, *see* calcite
 idler beam, 238
 impurities, 60, 80, 94, 100, 154, 158–60, 186, 192–6, 272–3
 impurities, optically active, 186, 192–6
 impurity scattering, 157
 impurity level, 158, 272–3
 indirect absorption, 51, 63–6
 indirect band gap, 50, 63, 96, 103–5, 277–8
 indium antimonide (InSb), 155–6, 217, 250–1, 278
 indium arsenide (InAs), 59, 138, 211, 278
 indium phosphide (InP), 221, 278
 indium-tin oxide (ITO), 182
 inelastic light scattering, 2, 218–22
 infinite potential well, 119, 128
 infrared absorption, 9–11, 20, 28, 38, 135, 157, 159, 212, 216–17
 infrared active phonon, 204
 infrared emission, 135, 136
 infrared reflectivity, 155, 206–11, 219, 225
 infrared resonance, 29, 35, 38, 204
 injection current, 108
 injection of carriers, 93, 95, 98, 107, 133, 182
 InSb: *see* indium antimonide
 insulator
 band structure, 271–3
 doped, 14–15
 excitons, 86–88
 optical properties, 8–11
 refractive index, 10
 intensity, optical, 3, 259–60
 interaction picture, 268
 interband absorption, 49–73, 95–7, 124–31, 151–4, 236
 interband luminescence, 49, 95–111, 133–4
 interband matrix element 51–6, 95, 126
 interface, optical, 261
 intersubband transition, 135
 intersystem crossing, 178–9
 intervalence band absorption, 157
 intrinsic region, 82, 107, 132, 280–1
 intrinsic semiconductor, 272–3
 inversion symmetry, 121, 123, 220, 239–40, 243
 ionic crystal, 28, 189–91, 206
 ionization field, 82, 132
 ionosphere, 145
 isotropic materials, 31, 43, 125, 243, 247–9, 256
 ITO, *see* indium-tin oxide
 joint density of states, 53, 128, 266
 K point, 276
 Kane four-band model, 57
 KDP (KH_2PO_4), 229, 240, 242–3
 KF: F_2^+ , 191, 199
 Kerr effect, optical, 244–7
 Kerr effect, DC, 246
 Kerr constant, 246
 Kleinman symmetry, 247
 Kramers–Kronig relationship, 40, 61, 68
 Kroemer, Herbert, 110
 L point, 56, 276
 LA phonon, *see* phonon, LA

- ladder polymer, 181
 Landau levels, 61, 216
 large polaron, 217
 laser
 cavity, 107
 colour centre, 198–9
 erbium-doped fibre, 198–9
 frequency doubling, 237–9
 infrared, 217
 modes, 108
 Nd:YAG, 194, 197–9, 237–8
 organic, 179
 principle of operation, 107–9, 265
 quantum cascade, 135
 quantum well, 134
 pumping mechanisms, 197–9
 ruby, 196, 246
 semiconductor, 107, 134
 short pulse, 195, 199, 224, 248
 solid state, 15, 107–10, 134, 196–9
 Raman, 246–7
 Ti:sapphire, 197–9
 threshold, 109
 vertical cavity surface emitting, 109
 vibronic, 197
 wavelengths, table of, 198
 laser diode, 103, 107–10, 134, 179
 lattice absorption, *see* phonon absorption
 lattice constant, 105
 lattice matching, 104–5, 118, 134, 138
 lattice vibration, *see* phonon
 LED, *see* light emitting diode
 Lenz's law, 257
 lifetime, non-radiative, 94, 196
 lifetime, phonon, 222–4
 lifetime, radiative, 93–4, 134, 169, 178–9,
 194, 196, 264, 270
 light emitting diode (LED), 10, 13–14,
 103–7, 134, 179, 181–3, 200–1,
 280
 light emitting polymer, 182
 light hole band, 57, 157
 light hole transition, 57, 129, 132
 light scattering, 2, 4, 218–22
 line broadening mechanisms, 30–3, 63; 81,
 134, 178, 194, 223
 linear conjugated molecule, 168
 linear electro-optic effect, 238–9
 liquid phase epitaxy (LPE), 104
 lithium fluoride (LiF), 87
 LO phonon, *see* phonon, LO
 local field, 39, 208
 localized electronic states, 77, 86, 165–6,
 177, 182, 186–7
 localized phonon, 165, 174, 178, 187–8
 longitudinal acoustic phonon, *see* phonon,
 LA
 longitudinal electric field oscillations, 161,
 209
 longitudinal laser modes, 108
 longitudinal optic phonon, *see* phonon, LO
 longitudinal phonon, *see* phonon,
 longitudinal
 Lorentz, Henrick Antoon, 26
 Lorentz model, 29–34, 143
 Lorentz local field correction, 39
 Lorentz oscillator, 29–34
 Lorentzian absorption, 32
 low-dimensional physics, 116
 LPE, *see* liquid phase epitaxy
 luminescence, 92–114, 133–4, 177–83,
 186–203
 apparatus, 103
 aromatic hydrocarbon, 177–9
 colour-centre, 189–91
 definition of, 2, 92, 263–4
 direct, 95
 efficiency, 94
 electroluminescence, 92, 103–10, 134,
 179, 181–3
 fluorescence, 169, 178, 270
 indirect, 96
 interband, 95–111, 133–4
 phosphor, 199–201
 phosphorescence, 169, 178–9, 200,
 270
 photoluminescence, 92, 98–103, 181
 polymer, 181–3
 quantum well, 133
 rare earth ions, 193–5
 rate, 264, 269–70
 semiconductor, 10, 95–111, 133, 182
 Stokes shift, 4, 171, 181, 188, 191
 time-resolved, 101–3
 transition metal ions, 195–6
 vibronic, 170–83, 186–9
 luminescence data,
 GaN, 96
 GaAs, 100, 107
 GaInAs, 102
 KF:F₂⁺, 191
 MeLPPP, 181
 Nd:YAG, 195
 PPV, 182
 pyrromethene 567 dye, 177
 Ti:sapphire, 196
 tricolour lamp, 200
 ZnCdSe quantum well, 133
 LUMO level, 168, 181
 Lyddane–Sachs–Teller (LST) relationship,
 209–10
 magnetic dipole transition, 270
 magnetic field strength, 256–7
 magnetic flux density, 257
 magnetic permeability, 257
 magnetic susceptibility, 256–7
 magnetization, 256–7
 matrix element, 51–4, 93, 95, 125, 175, 234,
 266–8
 Maxwell's equations, 257
 MBE, *see* molecular beam epitaxy
 MeLPPP, *see* poly(paraphenylenes)
 mercury discharge, 200
 metal-organic chemical vapour deposition
 (MOCVD), 104, 117
 metal-organic vapour phase epitaxy
 (MOVPE), 104, 117
 metals, 12, 147–54
 alkali, 149
 aluminium (Al), 149–51
 band structure, 151–4, 271–2
 colours, 12, 153–4
 copper (Cu), 152–3
 excitonic effects, 80
 free carrier effects, 29, 143–54
 gold (Au), 153, 164
 interband absorption, 151–4
 mirrors, 12, 149, 154
 optical properties, 12, 147–54
 plasma frequency, 148
 plasmons, 161
 reflectivity, 12, 147–54
 silver (Ag), 12, 153
 ultraviolet transparency, 12, 149
 valency, 148
 microscopic models, 20–1, 255
 mirror symmetry rule, 176, 181, 189, 191,
 196
 MOCVD, *see* metal-organic chemical
 vapour deposition
 modulator, optical, 133, 238, 246
 molecular beam epitaxy (MBE), 104, 117
 molecular materials, 13–14, 88–9, 165–185
 aromatic hydrocarbons, 13, 88, 166,
 177–9
 conjugated, 13, 89, 166–9, 177–83
 crystals, 166, 177, 180
 electronic states, 167–9
 exciton, 88, 180
 electroluminescence, 179, 181–3
 energy levels, 168–9
 laser, 179
 light emitting diode, 182
 optical properties, 13–14, 165–85
 photoluminescence, 181
 polaronic effects, 87, 217
 polymer, 89, 166–9, 179–83
 saturated, 13, 166, 179
 thin film, 166, 181–3
 molecular configuration, 171–3
 molecular spectra, 169–177
 molecular vibrations, 166, 170–6, 207, 212,
 246
 molecular Raman scattering, 220, 246
 molecule, conjugated, 167–9

molecule, diatomic, 171–3
 molecule, saturated, 168
 momentum, electromagnetic, 267
 momentum conservation, nonlinear, 242
 momentum scattering time, 146
 monoclinic crystal, 43
 monolayer fluctuations, 134
 monomer, 179
 monovalent metal, 271–2
 Mott density, 84
 MOVPE: metal-organic vapour phase epitaxy
 MQW, *see* multiple quantum well
 multiple quantum well, 117–18
 multiple resonance, 34–7
 multiplicity, spin, 168, 195
 n-type doping, 135, 154–61, 272–3
 NaCl, *see* sodium chloride
 Nakamura, Shuji, 105
 natural lifetime, *see* radiative lifetime
 natural resonant frequency, 26–37, 208–9,
 212, 230–1, 247
 nearly free electron model, 273–6
 Nd^{3+} , 194–5, 197–8,
 Nd:YAG, 194–5, 197–9
 Nd:YLF, 195
 neon, 87
 Neumann's principle, 16, 239–40
 NH_3 , *see* ammonia
 nitride compounds, 103–5, 134, 201
 nodes, wave function, 121, 123
 non-radiative lifetime, 94
 non-radiative relaxation, 94, 98, 170–1,
 173–4, 187, 197–8
 non-resonant dielectric constant, 35, 155
 non-resonant polarization, 31, 35
 non-resonant susceptibility, 208
 non-resonant nonlinear response, 230–4
 nonlinear
 absorption, 234–6, 249–50
 amplification, 238, 246
 coherence length, 241
 conversion efficiency, 241, 244
 crystal, 237–43
 dielectric constant, 228, 245
 frequency mixing, 237–39, 243–6
 optical coefficient tensor, 239–40
 optics, 2, 85, 227–54, 256
 phase shift, 247–9
 polarization, 228–9, 231–4, 237–40,
 243
 refractive index, 245–52
 susceptibility, 228, 234, 236–40,
 243–7
 susceptibility tensor, 227–30, 247
 switching device, 249–52
 nonlinear optics, 2, 227–54, 256

excitonic effects, 85, 249–50
 physical origin, 230–6
 non-resonant, 230, 234–6
 phase matching, 230, 241–3, 244–5
 resonant, 230–4, 249–52
 second-order, 228, 231–4, 236–43
 semiconductor, 249–52
 third-order, 228, 236, 243–52
 normal dispersion, 41
 normal incidence, 261
 occupancy factors, 93, 99, 236
 octahedral crystal environment, 192–6
 one-dimensional materials, 116, 168
 optic axis, 17, 43, 242–3
 optical
 amplification, 107, 198–9, 238
 amplifier, 196–9
 anisotropy, 17, 42–5
 bistability, 251
 coefficients, 2–8
 confinement, 109
 density, 4, 96, 118, 137, 181, 191
 fibre, 41, 198–9, 248–9
 Kerr effect, 244–7
 injection, 93, 133
 materials, 8–15
 modulator, 133, 238, 246, 280
 phonon, *see* phonon, optical
 processes, classification of, 1–5
 pumping, 197–8
 rectification, 238–9
 optoelectronics, 10, 18, 103–10, 133, 179,
 181–3, 280
 ordinary ray, 43, 242–3
 organic materials, *see* molecular materials
 organic optoelectronics, 179, 181–3
 orthorhombic crystal, 43
 oscillator
 anharmonic, 230–4
 atomic, 26–8
 bound electron, 26–8, 35
 classical, 25–37, 143–5, 206–9
 damped, 30, 32, 146, 208, 232
 dipole, 26–34
 free electron, 29, 143–5
 harmonic, 173, 189, 223
 Lorentz, 29
 strength, 36, 269
 vibrational, 28, 206–9
 output coupler, 108
 overlap integral, 126, 175, 189
 P-567, *see* pyrromethene 567 dye
 p-i-n diode, 69, 82, 132, 280–1
 p-n diode, 61, 69, 103, 106, 134, 280–1
 p-type doping, 105, 154, 157, 160, 272–3
 parabolic bands, 57–8, 127, 274–5
 parallel band effect, 152
 paramagnetic impurities, 192–6
 parametric amplification, 238, 246
 parametric oscillation, 238
 parity, 121, 123, 126, 195–6, 220
 parity-forbidden transitions, 132
 Pauli exclusion principle, 50, 93, 236
 PDA, *see* polydiacetylene
 perspex, 13
 phase lag due to scattering, 28
 phase matching, 230, 238, 241–3, 244–5
 phase modulator, 238
 phase velocity, 38
 phonon, 204–26
 absorption, 9, 206, 212
 acoustic, 204–5, 218, 222–3, 246
 anharmonic process, 223
 Brillouin scattering, 218, 222, 244,
 246
 delocalized modes, 20
 dispersion, 205, 214–15
 inelastic light scattering, 218–22
 infrared active, 204–5, 220
 ionization of excitons, 81
 LA (longitudinal acoustic), 206
 lifetime, 208, 212, 222–4
 LO (longitudinal optic), 81, 205–6,
 209–18, 221, 224, 246
 localized, 165–6, 174, 178, 187–8
 longitudinal, 204–5
 Lyddane-Sachs-Teller (LST)
 relationship, 209–10
 optical, 204–5, 218, 223, 246
 oscillator, 29, 206–9
 parity, 220
 polariton, 214–15
 polaron, 215–18
 Raman active, 220, 246
 Raman scattering, 20, 161, 214–5,
 218, 220–1, 244, 246
 reflectivity, 206–12, 214
 restrahlen, 210–2
 scattering, 81, 98, 157, 223
 TA (transverse acoustic), 65, 206
 TO (transverse optic), 205–15, 220–1,
 246
 transverse, 204–5
 virtual, 216
 phonon-phonon interaction, 223
 phosphor, 105, 199–201
 phosphorescence, 169, 178–9, 200, 270
 phosphorous, 159
 photoconductivity, 71
 photocurrent, 69, 82, 132, 180–1
 photodetector, 69–72, 133, 135
 photodiode, 69, 280
 photoexcitation, 99
 photoluminescence, 92–103, 133, 181

- photoluminescence spectroscopy, 102
 photoluminescence excitation spectroscopy (PLE), 102, 141
 photoluminescence, time-resolved, 101–3
 photomultiplier, 103, 221
 photovoltaic device, 71
 Planck formula, 265
 plasma, 85, 143, 161
 plasma edge, 155
 plasma frequency, 12, 144, 155, 161
 plasma oscillations, 160
 plasma reflectivity, 143–5, 155
 plasmon, 20, 160–2, 209, 218
PLE, see photoluminescence excitation spectroscopy
 plexiglas, *see* perspex
 Pockels effect, 238–9, 246
 point group symmetry, 16
 Poisson's equation, 256, 280–1
 polar molecule, 28
 polar bond, 28, 206
 polar solid, 206–13
 polariton, 214–15, 220
 polarization, dielectric, 30, 144, 227, 255–6
 polarization, nonlinear, 228–9, 231–4, 237–40, 243
 polarization, optical, 125, 135, 259–60
 polaron, 87, 215–18
 polaron effective mass, 216–17
 polaron radius, 217
 polyacetylene, 179–80
 polydiacetylene (PDA), 89, 166, 180–1, 217
 polyethylene, *see* polythene
 polyfluorene, 13–14, 181
 polymer, 13, 89, 166, 168, 179–83
 polymer, luminescent, 181–3
 poly(paraphenylene), 181
 poly-phenylenevinylene (PPV), 182
 polythene, 13, 179
 population inversion, 107, 197
 potassium dihydrogen phosphate, *see* KDP
 potential, electrostatic, 256–8
 potential, vector, 257
 potential well, 119–24
 Poynting vector, 259
 PPV, *see* poly-phenylenevinylene
 prompt fluorescence, 179, 270
 propagation of light, 1, 25–46
 prism, 41
 pulse broadening, 41, 248
 pulsed laser, 199
 pyrene ($C_{16}H_{10}$), 88
 pyrromethene 567 dye, 176–7
- quadratic electro-optic effect, 246
 quantization energy, 120
 quantum box, *see* quantum dot
 quantum cascade laser, 135
- quantum confined Stark effect, 131
 quantum confinement, 115–17, 168, 176
 quantum dot, 116, 136–8
 quantum efficiency, 69, 94, 109
 quantum optics, 21, 238
 quantum size effect, 14, 116, 137, 168
 quantum theory of radiative transitions, 263–70
 quantum well, 115–42
 absorption, 124–31, 135, 250
 definition, 116
 density of states, 128
 electronic states, 119–24
 electroluminescence, 106, 133–4
 exciton, 130–3, 250
 laser, 134, 197
 non-lattice matched, 118, 134
 nonlinear optics, 250
 photoluminescence, 133
 quantum confined Stark effect, 131
 radiative efficiency, 134
 structure, 117
 quantum wire, 116
 quartz, 44
 quasi-equilibrium, 99
 quaternary alloy, 104–5
- Rabi oscillations, 266
 radiative efficiency, 94, 196
 radiative lifetime, 93–6, 134, 169, 178–9, 194, 196, 264, 270
 radiative transition, 93, 263–70
 Raman, C.V., 220
 Raman scattering, 161, 218, 220–1, 246
 AlSb, 221
 anti-Stokes, 218–19
 apparatus, 220–1
 frequency shift, 161, 218
 GaP, 215, 221
 GaAs, 221
 InP, 221
 lifetime broadening, 223
 n-type GaAs, 161–2
 molecular, 220
 optical phonons, 20, 161, 220–1
 plasmon, 161–2
 polariton, 214–15
 selection rules, 220
 spin-flip, 246
 spontaneous, 246
 stimulated, 244, 246
 Stokes, 218–19, 221
- rare earth ion, 192–6, 200
 rare gas crystals, 87
 Rayleigh scattering, 4
 reciprocal lattice vector, 273–4
 recombination, electron-hole, 95, 101, 106, 133, 182
- rectification, optical, 238
 reduced mass, 26, 58, 78, 127, 173, 208
 reduced zone scheme, 274–5
 reflection, 1, 261
 reflection coefficient, 2, 262
 reflectivity
 definition of, 2, 262
 free carrier, 12, 143–56
 infrared, 155, 206–11
 lattice, 206–14
 plasma, 143–5
 relationship to refractive index, 7, 68, 262
 restrahlen, 210–12
- reflectivity spectra
 aluminium, 149
 AlSb, 225
 copper, 153
 GaAs, 211
 gold, 164
 InAs, 211
 InSb, n-type, 156
 silver, 12
- refraction, 2
 refractive index,
 complex, 4–8, 260–1
 definition of, 3, 259
 dipole oscillator, 29–37
 dispersion, 40–2
 frequency dependence, 36
 microscopic origin, 28
 nonlinear, 245–52
 relation to dielectric constant, 6, 259
- refractive index data
 fused silica, 11, 37, 41
 insulators, 10
 NaCl, 47
 semiconductors, 11
 uniaxial crystals, 43
- relative dielectric constant, 6, 256
 relative magnetic permeability, 257
 relaxation mechanisms, 93, 98, 173–4, 187
 residual ray, 211
 resonance enhancement, 234, 246
 resonant nonlinear response, 230, 234–6, 246, 249–52
- resonant frequency, 26–37, 208, 230–1, 247
 resonant polarization, 31, 35
 responsivity, 69
 restrahlen, 210–12, 214
 reverse bias, 280–1
 rhombohedral crystal, 43
 rigid lattice model, 216
 ruby ($Cr^{3+}: Al_2O_3$), 2, 15, 186, 192, 196
 rule of mutual exclusion, 220
 Russell, John Scott, 248
 Russell-Saunders coupling, 193
 Rutherford, Ernest, 26
 Rydberg constant, 78

- s bands, 152–3, 276
 sapphire (Al_2O_3), 8–10, 15, 105, 195–6, 206, 212
 saturable absorber, 236, 249–50
 saturated bond, 13, 168, 179
 saturation intensity, 236
 scattering of light, 2–4, 28, 161, 218–22
 scattering cross-section, 4
 second harmonic generation, 2, 231–2, 237–9, 241–3
 second-order nonlinear optics, 228–34, 236–43
 second-order nonlinear susceptibility, 228, 234, 236–40
 second-order nonlinear susceptibility tensor, 229
 selection rules,
 electric dipole, 269–70
 electric quadrupole, 270
 infrared activity, 220
 Landau level, 62
 magnetic dipole, 270
 parity, 126, 270
 quantum well, 125, 135
 spin, 169, 178, 270
 Raman scattering, 220
 wave vector, 53
 self-organized quantum dots, 138
 self-phase modulation, 244, 247–9
 self-trapping, 87, 217
 semiclassical model, 20, 52
 semiconductor
 band gaps, tables of, 277–8
 band structure, 56–8, 271–3, 277–9
 conductivity, 272
 crystallites, 14
 doped, 29, 135, 143, 154–62, 246, 272–3
 doped glass, 14, 115, 137
 effective mass, 277–8
 extrinsic, 272–3
 GaAs, *see* gallium arsenide
 Ge, *see* germanium
 heterojunction, 110, 117
 II–VI, 78, 105, 137, 154, 216–17, 272, 278
 III–V, 55–57, 78, 105, 138, 154, 157, 206, 210, 216–17, 221, 250, 272, 278
 intrinsic, 272–3
 light-emitting diode, 106, 134, 200
 laser, 107–10, 134, 135
 n-type, 135, 143, 154–61, 272–3
 nonlinear effects, 85, 249–52
 optical properties, summary of, 8–11
 organic, 181–3
 p-i-n diode, 82, 280–1
 p-n junction, 106, 280
 p-type, 143, 154, 157, 160, 272–3
 photodetectors, 69
 quantum dot, 116, 136–38
 quantum well, 106, 115–42
 quantum wire, 116
 Si, *see* silicon
 superlattice, 116–18, 135
 semimetal, 278
 signal beam, 238
 signal velocity, 38
 silica (SiO_2), 11–12, 37, 41–2, 207, 248
 silicon (Si)
 absorption, 64, 66
 band structure, 67, 271, 276–7
 band gap, 63, 277
 effective mass, 277
 exciton, 85
 luminescence, 97
 n-type, 159, 272
 phonon, 210, 220
 photodiode, 70
 Raman active modes, 220
 solar cell, 72
 silicon carbide (SiC), 103
 silver (Ag), 12, 153
 simple harmonic oscillator, 173, 189, 223
 single quantum well, 117
 singlet states, 168–9, 176, 177–9
 singlet-singlet transition, 169, 177–81
 singlet-triplet transition, 169, 178–9
 skin depth, 147
 skin effect, 147, 156
 slope efficiency, 109
 small polaron, 217
 Snell's law of refraction, 2
 sodium, 271
 sodium chloride (NaCl), 28, 47, 87, 177, 207, 213
 soft condensed matter, 166
 solar cells, 71, 280
 solid state, characteristic optical physics, 15–20
 solid state laser, 15, 107–10, 134, 196–9
 solitons, 248
 sound velocity, 205, 222
 spin coating, 166, 182
 spin multiplicity, 168, 195
 spin quantum number, 168, 177, 270
 spin selection rule, 169, 178, 270
 split-off hole band, 57, 157, 277–8
 spin-orbit interaction, 57, 169, 179, 193–5, 200, 277–8
 spontaneous emission, 2, 4, 92, 107, 246, 263–5
 stained glass, 14, 138
 Stark effect, 131
 static crystal field effect, 192–3
 states, Bloch, 18, 52, 125, 276
 states, delocalized, 19, 77, 166–8, 186
 states, localized, 77, 86, 165–6, 174, 177, 182, 186, 215–18
 stimulated Brillouin scattering, 244, 246
 stimulated emission, 107, 234–6, 264–5
 stimulated Raman scattering, 244, 246
 Stokes, George, 4
 Stokes scattering, 218–19
 Stokes shift, 4, 14, 171, 181, 188, 191, 196
 streak camera, 103
 substrate, 103–5, 117
 sum frequency mixing, 237–9
 superlattice, 116–18, 135
 susceptibility, *see* electric susceptibility
 susceptibility, nonlinear, 227–34, 236–40, 243–7
 symmetry,
 class, 16, 43, 229, 239–40
 Kleinman, 247
 lowering, 17, 121, 131, 132, 193
 point group, 16
 points of Brillouin zone, 56, 276
 octahedral, 192–6
 translational, 16, 276
 symmetry, effects of, 16–18
 birefringence, 43
 crystal field splitting, 193, 195
 electronic bands, 57
 infrared activity, 220
 nonlinear susceptibility, 229, 239–40, 247
 Raman activity, 220
 TA phonon, *see* phonon, TA
 Tb^{3+} , 200
 telecommunications, fibre optic, 41, 104, 198–9, 248
 ternary alloy, 104
 tetracene ($\text{C}_{18}\text{H}_{10}$), 179
 tetragonal crystal, 43, 229, 240
 thermal equilibrium, 99, 101, 265
 thin film, 166, 181–3, 212
 third harmonic generation, 237–8, 244
 third-order nonlinear optics, 228–9, 236, 243–52
 third-order nonlinear susceptibility, 229, 243–7
 Thomson, J.J., 26
 threshold current, 108
 Ti^{3+} , *see* titanium
 Ti:sapphire laser, 196–9
 tight-binding band theory, 276
 tightly bound exciton, *see* Frenkel exciton
 time-bandwidth product, 41, 199, 248
 time-resolved photoluminescence
 spectroscopy, 101–3
 time-resolved Raman spectroscopy, 224
 titanium (Ti^{3+}), 195–9
 TO phonon, *see* phonon, TO

- translational symmetry, 16, 276
 transition,
 allowed, 55, 95, 126, 169–70, 178, 270
 electric dipole, 52, 54, 125–6, 169–70, 175, 194, 268–70
 electric quadrupole, 270
 forbidden, 132, 169, 194, 196, 200, 270
 magnetic dipole, 270
 probability, 36, 266
 radiative, 263–70
 rate, 51, 266–70
 selection rules, 269–70
 transition metal, 152
 transition metal ion, 192–6
 transmission, 1, 261
 transmission, coefficient of, 3
 transmission spectra
 CdSe, 9
 germanium, 62
 ruby, 15
 sapphire, 9, 15
 transmissivity, 3
 transparency range, 9–11, 36
 transverse acoustic phonon, *see* phonon, TA
 transverse light wave, 260
 transverse optic phonon, *see* phonon, TO
 transverse phonon, *see* phonon, transverse traps, 94
 triclinic crystal, 43, 240
 tricolour lamp, 200
 trigonal crystal, 43
 triplet states, 168–9, 178–9
 triplet-singlet transition, 169, 178–9
 trivalent metal, 271–2
 Tsu, R., 115, 118, 135
 tunnelling, 121
 two-dimensional materials, 115–36
 two-photon absorption, 244, 267
 ultraviolet absorption, 38
 ultraviolet transparency of metals, 149
 uncertainty principle, 41, 115, 199
 up-conversion, 103
 uniaxial crystal, 43, 229, 240
 vacancies, crystal, 189–91
 vacuum photon, 246
 valence band, 50–7, 95–7, 157, 271–3
 valence electrons, 55, 148, 154, 167, 271–6
 valency, 148, 271
 van der Waals interaction, 13, 88, 166, 177
 Van Hove singularities, 68
 vector potential, 257–8, 267
 velocity of sound, 205, 222
 velocity of light, 3, 38, 258–9
 vertical cavity surface emitting laser, 109
 vibrational absorption, 9, 36, 212
 vibrational-electronic, *see* vibronic
 vibrational levels, 170, 173
 vibrational oscillator, 25, 28, 170, 206–9
 vibronic
 absorption, 170–81, 186–9
 band, 19, 170, 186–9, 196
 coupling, 15, 19, 187
 emission, 170–7, 186–91
 luminescence, 177, 181–2, 191, 195–6, 200
 material, 165–83, 186–201
 mirror symmetry rule, 176, 181, 191, 196
 states, 165, 175, 186–9
 transition, 13, 169–83, 186–201
 wave function, 175
 virtual transition, 234
 W point, 276
 Wannier-Mott exciton, *see* free exciton
 wave, electromagnetic, 258–62
 wave equation, 258
 wave function, Bloch, 18, 52, 125, 276
 wave function, quantum well, 119–24
 wave function, vibronic, 175
 wave function nodes, 121, 123
 wave impedance, 260
 wave number units, 161, 211
 wave vector, 20, 52, 205, 214, 259
 waveguide, 110
 wurtzite structure, 105, 278–9
 X point, 56, 276
 xenon, 87
 YAG (yttrium aluminium garnet), 194
 YLF (YLiF_4), 195
 Zeeman effect, 17, 193
 zero-dimensional materials, 116, 136
 zero-phonon lines, 189–90, 196
 zero-point fluctuations, 246, 264
 zinc blende structure, 56, 105, 240, 278–9
 zinc cadmium telluride (ZnCdTe), 133
 zinc selenide (ZnSe), 216

Fundamental constants

Bohr radius for hydrogen	$4\pi\epsilon_0\hbar^2/m_0e^2 =$	a_{H}	$5.292 \times 10^{-11} \text{ m}$
Velocity of light in free space		c	$2.9979 \times 10^8 \text{ m s}^{-1}$
Electronic charge		e	$1.6022 \times 10^{-19} \text{ C}$
Gravitational constant		G	$6.673 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$
Planck's constant		h	$6.626 \times 10^{-34} \text{ J s}$
Boltzmann's constant	$h/2\pi =$	\hbar	$1.0546 \times 10^{-34} \text{ J s}$
Electron rest mass		k_{B}	$1.3807 \times 10^{-23} \text{ J K}^{-1}$
Proton rest mass		m_0	$9.109 \times 10^{-31} \text{ kg}$
Avogadro's number		m_p	$1.6726 \times 10^{-27} \text{ kg}$
Molar gas constant		N_A	$6.022 \times 10^{23} \text{ mol}^{-1}$
Rydberg's constant	$\alpha^2 m_0 c / 2h =$	R	$8.315 \text{ J mol}^{-1} \text{ K}^{-1}$
Rydberg's constant for hydrogen		R_{∞}	$1.0974 \times 10^7 \text{ m}^{-1}$
Standard molar volume		R_H	13.606 eV
Fine structure constant	$e^2 / 4\pi\epsilon_0\hbar c =$	V_m	$22.414 \times 10^{-3} \text{ m}^3 \text{ mol}^{-1}$
Electric permittivity of free space		α	$(137.036)^{-1}$
Bohr magneton		ϵ_0	$8.854 \times 10^{-12} \text{ F m}^{-1}$
Nuclear magneton		μ_B	$9.274 \times 10^{-24} \text{ A m}^2 \text{ or JT}^{-1}$
Magnetic flux quantum	$h/2e =$	μ_N	$5.051 \times 10^{-27} \text{ A m}^2 \text{ or JT}^{-1}$
Stefan's constant		Φ_0	$2.0678 \times 10^{-15} \text{ T m}^2$
Magnetic permeability of free space		σ	$5.671 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$
		μ_0	$4\pi \times 10^{-7} \text{ H m}^{-1}$

Energy equivalents for photons $E = h\nu = k_{\text{B}}T = \frac{hc}{\lambda}$

	E	ν	T	λ	λ^{-1}	
(J)	(eV)	(Hz)	(K)	(m)	(m ⁻¹)	(cm ⁻¹)
1	6.242×10^{18}	1.509×10^{33}	7.243×10^{22}	1.486×10^{-25}	5.034×10^{24}	5.034×10^{22}
1.602×10^{-19}	1	2.418×10^{14}	1.160×10^4	1.240×10^{-6}	8.066×10^5	8.066×10^3
6.626×10^{-34}	4.136×10^{-15}	1	4.799×10^{-11}	2.998×10^8	3.336×10^{-9}	3.336×10^{-11}
1.381×10^{-23}	8.617×10^{-5}	2.084×10^{10}	1	1.439×10^{-2}	69.50	0.6950
1.987×10^{-25}	1.240×10^{-6}	2.998×10^8	1.439×10^{-2}	1	1.0	0.01
1.987×10^{-25}	1.240×10^{-6}	2.998×10^8	1.439×10^{-2}	1.0	1	0.01
1.987×10^{-23}	1.240×10^{-4}	2.998×10^{10}	1.439	0.01	100	1