

Quantum Processes in Semiconductors

FIFTH EDITION

OXFORD

B. K. RIDLEY

QUANTUM PROCESSES IN SEMICONDUCTORS

This page intentionally left blank

Quantum Processes in Semiconductors

Fifth Edition

B. K. RIDLEY FRS

*Professor Emeritus of Physics
University of Essex*

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

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

© B. K. Ridley, 1982, 1988, 1993, 1999, 2013

The moral rights of the author have been asserted

First Edition published in 1982
Second Edition published in 1988
Third Edition published in 1993
Fourth Edition published in 1999
Fifth Edition published in 2013

Impression: 1

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, by licence 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 work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2013941293

ISBN 978-0-19-967721-4 (hbk.)
ISBN 978-0-19-967722-1 (pbk.)

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Preface to the Fifth Edition

Semiconductor physics is of fundamental importance in understanding the behaviour of semiconductor devices and for improving their performance. Among the more recent devices are those exploiting the properties of III–V nitrides, and others that explore the technical possibilities of manipulating the spin of the electron. The III–V nitrides, which have the hexagonal structure of wurtzite (ZnO), have properties that are distinct from those like GaAs and InP, which have the cubic structure of zinc blende (ZnS). Moreover, AlN and GaN have large band gaps, which make it possible to study electron transport at very high electric fields without producing breakdown. This property, combined with an engineered large electron population, makes GaN an excellent candidate for high-power applications. In such situations the role of hot phonons and their coupling with plasmon modes cannot be ignored. This has triggered a number of recent studies concerning the lifetime of hot phonons, leading to the discovery of new physics. An account of hot-phonon effects, the topic of the first of the new chapters, seemed to be timely.

In the new study of spintronics, a vital factor is the rate at which an out-of-equilibrium spin population relaxes. The spin of the electron scarcely enters the subject matter of previous editions of this book other than in relation to the density of states, so an account of spin processes has been overdue, hence the second of the new chapters in this edition. The rate of spin relaxation is intimately linked to details of the band structure, and in describing this relationship I have taken the opportunity to describe the band structure of wurtzite and the corresponding eigenfunctions of the bands, from which the cubic results are deduced. There are several processes that relax spin in bulk material, and these are described.

The properties of semiconductors extend beyond the bulk. All semiconductors have surfaces and, when incorporated into devices, they have interfaces with other materials. The physics of metal–semiconductor interfaces has been studied ever since the discovery of rectifying properties in the early part of the 20th century. More recently, the advent of so-called low-dimensional devices has highlighted problems connected with the physics of interfaces between different semiconductors, so an account of the properties of surfaces and interfaces was, it seemed to me, no longer timely, but long overdue. Hence, the third new chapter.

This new edition is therefore designed to expand (rather than replace) the physics of bulk semiconductors found in the previous edition. The

Preface to the Fifth Edition

expansion has been motivated by the subject matter of my own research and that of colleagues at the Universities of Essex and Cornell. I am particularly indebted to Dr Angela Dyson for her insightful collaboration in these studies.

Thorpe-le-Soken, 2013

B.K.R.

Preface to the Fourth Edition

This new edition contains three new chapters concerned with material that is meant to provide a deeper foundation for the quantum processes described previously, and to provide a statistical bridge to phenomena involving charge transport. The recent theoretical and experimental interest in fundamental quantum behaviour involving mixed and entangled states and the possible exploitation in quantum computation meant that some account of this should be included. A comprehensive treatment of this important topic involving many-particle theory would be beyond the scope of the book, and I have settled on an account that is based on the single-particle density matrix. A remarkably successful bridge between single-particle behaviour and the behaviour of populations is the Boltzmann equation, and the inclusion of an account of this and some of its solutions for hot electrons was long overdue. If the Boltzmann equation embodied the important step from quantum statistical to semi-classical statistical behaviour, the drift-diffusion model completes the trend to fully phenomenological description of transport. Since many excellent texts already cover this area I have chosen to describe only some of the more exciting transport phenomena in semiconductor physics such as those involving a differential negative resistance, or involving acoustoelectric effects, or even both, and something of their history.

A new edition affords the opportunity to correct errors and omissions in the old. No longer being a very assiduous reader of my own writings, I rely on others, probably more than I should, to bring errors and omissions to my attention. I have been lucky, therefore, to work with someone as knowledgeable as Dr N.A. Zakhleniuk who has suggested an update of the discussion of cascade capture, and has noted that the expressions for the screened Bloch–Grüneisen regime were for 2-D systems and not for bulk material. The update and corrections have been made, and I am very grateful for his comments.

My writing practically always takes place at home and it tends to involve a mild autism that is not altogether sociable, to say the least. Nevertheless, my wife has put up with this once again with remarkable good humour and I would like to express my appreciation for her support.

Preface to the Third Edition

One of the topics conspicuously absent in the previous editions of this book was the scattering of phonons. In a large number of cases phonons can be regarded as an essentially passive gas firmly anchored to the lattice temperature, but in recent years the importance to transport of the role of out-of-equilibrium phonons, particularly optical phonons, has become appreciated, and a chapter on the principal quantum processes involved is now included. The only other change, apart from a few corrections to the original text (and I am very grateful to those readers who have taken the trouble to point out errors) is the inclusion of a brief subsection on exciton annihilation, which replaces the account of recombination involving an excitonic component. Once again, only processes taking place in bulk material are considered.

Thorpe-le-Soken
December 1992

B.K.R.

Preface to the Second Edition

This second edition contains three new chapters—‘Quantum processes in a magnetic field’, ‘Scattering in a degenerate gas’, and ‘Dynamic screening’—which I hope will enhance the usefulness of the book. Following the ethos of the first edition I have tried to make the rather heavy mathematical content of these new topics as straightforward and accessible as possible. I have also taken the opportunity to make some corrections and additions to the original material—a brief account of alloy scattering is now included—and I have completely rewritten the section on impact ionization. An appendix on the average separation of impurities has been added, and the term ‘third-body exclusion’ has become ‘statistical screening’, but otherwise the material in the first edition remains substantially unchanged.

Thorpe-le-Soken 1988

B.K.R.

Preface to the First Edition

It is a curious fact that in spite of, or perhaps because of, their overwhelming technological significance, semiconductors receive comparatively modest attention in books devoted to solid state physics. A student of semiconductor physics will find the background theory common to all solids well described, but somehow all the details, the applications, and the examples—just those minutiae which reveal so vividly the conceptual cast of mind which clarifies a problem—are all devoted to metals or insulators or, more recently, to amorphous or even liquid matter. Nor have texts devoted exclusively to semiconductors, excellent though they are, fully solved the student's problem, for they have either attempted global coverage of all aspects of semiconductor physics or concentrated on the description of the inhomogeneous semiconducting structures which are used in devices, and in both cases they have tended to confine their discussion of basic physical processes to bare essentials in order to accommodate breadth of coverage in the one and emphasis on application in the other. Of course, there are distinguished exceptions to these generalizations, texts which have specialized on topics within semiconductor physics, such as statistics and band structure to take two examples, but anyone who has attempted to teach the subject to postgraduates will, I believe, agree that something of a vacuum exists, and that filling it means resorting to research monographs and specialist review articles, many of which presuppose a certain familiarity with the field.

Another facet to this complex and fascinating structure of creating, assimilating, and transmitting knowledge is that theory, understandably enough, tends to be written by theoreticians. Such is today's specialization that the latter tend to become removed from direct involvement in the empirical basis of their subject to a degree that makes communication with the experimentalist fraught with mutual incomprehension. Sometimes the difficulty is founded on a simple confusion between the disparate aims of mathematics and physics—an axiomatic formulation of a theory may make good mathematical sense but poor physical sense—or it may be founded on a real subtlety of physical behaviour perceived by one and incomprehended by the other, or more usually it may be founded on ignorance of each other's techniques, of the detailed analytic and numerical approximations propping up a theory on the one hand, and of the detailed method and machinery propping up an experimental result on the other. Certainly, experimentalists cannot avoid being theoreticians from time to time, and they have to be aware of the basic theoretical structure of their subject. As students of physics operating in an area where physical intuition is more

important than logical deduction they are not likely to appreciate a formalistic account of that basic structure even though it may possess elegance. Intuition functions on rough approximation rather than rigour, but too few accounts of theory take that as a guide.

This book, then, is written primarily for the postgraduate student and the experimentalist. It attempts to set out the theory of those basic quantum-mechanical processes in homogeneous semiconductors which are most relevant to applied semiconductor physics. Therefore the subject matter is concentrated almost exclusively on electronic processes. Thus no mention is made of phonon–phonon interactions, nor is the optical absorption by lattice modes discussed. Also, because I had mainstream semiconductors like silicon and gallium arsenide in mind, the emphasis is on crystalline materials in which the electrons and holes in the bands obey non-degenerate statistics, and little mention is made of amorphous and narrow-gap semiconductors. Only the basic quantum mechanics is discussed; no attempt is made to follow detailed applications of the basic theory in fields such as hot electrons, negative-differential resistance, acousto-electric effects, etc. To do that would more than triple the size of the book. The theoretical level is at elementary first- and second-order perturbation theory, with not a Green's function in sight; this is inevitable, given that the author is an experimentalist with a taste for doing his own theoretical work. Nevertheless, those elementary conceptions which appear in the book are, I believe, the basic ones in the field which most of us employ in everyday discussions, and since there is no existing book to my knowledge which contains a description of all these basic processes I hope that this one will make a useful reference source for anyone engaged in semiconductor research and device development.

Finally, a word of caution for the reader. A number of treatments in the book are my own and are not line-by-line reproductions of standard theory. Principally, this came about because the latter did not exist in a form consistent with the approach of the book. One or two new expressions have emerged as a by-product, although most of the final results are the accepted ones. Where the treatment is mine, the text makes this explicit.

This page intentionally left blank

Contents

1	Band structure of semiconductors	1
1.1.	The crystal Hamiltonian	1
1.2.	Adiabatic approximation	1
1.3.	Phonons	2
1.4.	The one-electron approximation	3
1.5.	Bloch functions	4
1.6.	Nearly-free-electron model	6
1.6.1.	Group theory notation	8
1.7.	Energy gaps	10
1.8.	Spin–orbit coupling and orbital characteristics	12
1.9.	Band structures	14
1.10.	Chemical trends	18
1.11.	$\mathbf{k} \cdot \mathbf{p}$ perturbation and effective mass	22
1.11.1.	Oscillator strengths	27
1.12.	Temperature dependence of energy gaps	28
1.13.	Deformation potentials	30
1.14.	Alloys	34
	References	35
2	Energy levels	36
2.1.	The effective-mass approximation	36
2.2.	Electron dynamics	39
2.3.	Zener–Bloch oscillations	41
2.4.	Landau levels	44
2.5.	Plasma oscillations	48
2.6.	Excitons	49
2.7.	Hydrogenic impurities	51
2.8.	Hydrogen molecule centres	56
2.9.	Core effects	57
2.10.	Deep-level impurities	59
2.11.	Scattering states	64
2.12.	Impurity bands	64
	References	68
3	Lattice scattering	69
3.1.	General features	69
3.2.	Energy and momentum conservation	73
3.2.1.	Spherical parabolic band	73
3.2.1.1.	Absorption	73
3.2.1.2.	Emission	75

3.2.2.	Spherical non-parabolic band	77
3.2.3.	Ellipsoidal parabolic bands	78
3.2.4.	Equivalent valleys	78
3.2.5.	Non-equivalent valleys	79
3.3.	Acoustic phonon scattering	79
3.3.1.	Spherical band: equipartition	81
3.3.2.	Spherical band: zero-point scattering	83
3.3.3.	Spheroidal parabolic bands	84
3.3.4.	Momentum and energy relaxation	87
3.4.	Optical phonon scattering	89
3.4.1.	Inter-valley scattering	92
3.4.2.	First-order processes	93
3.5.	Polar optical mode scattering	95
3.5.1.	The effective charge	98
3.5.2.	Energy and momentum relaxation	99
3.6.	Piezoelectric scattering	101
3.7.	Scattering-induced electron mass	106
3.8.	Mobilities	108
3.9.	Appendix: Acoustic waves in the diamond lattice	112
	References	115
4	Impurity scattering	116
4.1.	General features	116
4.2.	Charged-impurity scattering	119
4.2.1.	Conwell–Weisskopf approximation	119
4.2.2.	Brooks–Herring approach	121
4.2.3.	Uncertainty broadening	124
4.2.4.	Statistical screening	126
4.3.	Neutral-impurity scattering	129
4.3.1.	Hydrogenic models	129
4.3.2.	Square-well models	130
4.3.3.	Sclar’s formula	132
4.3.4.	Resonance scattering	133
4.3.5.	Statistical screening	136
4.4.	Central-cell contribution to charged-impurity scattering	137
4.5.	Dipole scattering	144
4.6.	Electron–hole scattering	145
4.7.	Electron–electron scattering	147
4.8.	Mobilities	150
4.9.	Appendix: Debye screening length	151
4.10.	Appendix: Average separation of impurities	154
4.11.	Appendix: Alloy scattering	155
	References	156
5	Radiative transitions	157
5.1.	Transition rate	157
5.1.1.	Local field correction	160
5.1.2.	Photon drag	160

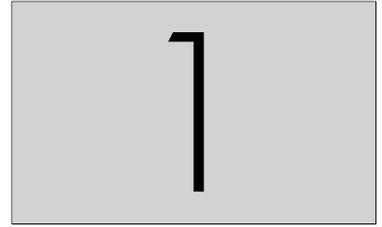
5.2.	Photo-ionization and radiative capture cross-sections	161
5.3.	Wavefunctions	162
5.4.	Direct interband transitions	165
5.4.1.	Excitonic absorption	168
5.5.	Photo-deionization of a hydrogenic acceptor	169
5.6.	Photo-ionization of a hydrogenic donor	171
5.7.	Photo-ionization of quantum-defect impurities	173
5.8.	Photo-ionization of deep-level impurities	178
5.9.	Summary of photo-ionization cross-sections	180
5.10.	Indirect transitions	181
5.11.	Indirect interband transitions	183
5.12.	Free-carrier absorption	186
5.12.1.	Energy and momentum	189
5.12.2.	Scattering matrix elements	189
5.12.3.	Electron scattering by photons	191
5.12.4.	Absorption coefficients	193
5.13.	Free-carrier scattering of light	195
5.13.1.	Scattering of laser light	199
5.14.	Appendix: Justification of effective-mass approximation in light scattering	201
	References	202
6	Non-radiative processes	203
6.1.	Electron–lattice coupling	203
6.2.	The configuration coordinate diagram	205
6.2.1.	Semi-classical thermal broadening	207
6.3.	Semi-classical thermal generation rate	208
6.4.	Thermal broadening of radiative transitions	211
6.5.	Thermal generation and capture rates	217
6.6.	Electron–lattice coupling strength	222
6.7.	Selection rules for phonon–impurity coupling	228
6.8.	Phonon-cascade capture	231
6.9.	The Auger effect	233
6.10.	Impact ionization	240
6.11.	Appendix: The multiphonon matrix element	242
	References	244
7	Quantum processes in a magnetic field	246
7.1.	Introduction	246
7.2.	Collision-free situation	247
7.2.1.	Quantum states in a magnetic field	247
7.2.2.	Magnitudes	249
7.2.3.	Density of states	249
7.2.4.	Spin	251
7.2.5.	Phenomenological quantities	251
7.3.	Collision-induced current	252
7.3.1.	Expression for the scattering rate in the extreme quantum limit	252

7.3.2.	Energy and momentum conservation	254
7.3.3.	Integrations	255
7.3.4.	General expression for the drift velocity	256
7.3.5.	Diffusion	259
7.4.	Scattering mechanisms	259
7.4.1.	Acoustic phonon scattering	259
7.4.2.	Piezoelectric scattering	262
7.4.3.	Charged-impurity scattering	263
7.4.4.	Statistical weighting for inelastic phonon collisions	266
7.5.	Transverse Shubnikov–de Haas oscillations	267
7.5.1.	Magnetoconductivity in the presence of many Landau levels	267
7.5.2.	The oscillatory component	270
7.5.3.	Collision broadening	271
7.5.4.	Thermal broadening	272
7.5.5.	Spin-splitting	272
7.5.6.	Shubnikov–de Haas formula	273
7.6.	Longitudinal Shubnikov–de Haas oscillations	274
7.7.	Magnetophonon oscillations	276
	References	280
8	Scattering in a degenerate gas	282
8.1.	General equations	282
8.2.	Elastic collisions	284
8.3.	Acoustic phonon scattering	284
8.3.1.	Low-temperature limit	287
8.3.2.	High-temperature limit	288
8.3.3.	Strong screening	288
8.4.	Energy relaxation time	290
	References	291
9	Dynamic screening	292
9.1.	Introduction	292
9.2.	Polar optical modes	293
9.3.	Plasma modes	295
9.4.	Coupled modes	296
9.5.	The Lindhard dielectric function	302
9.6.	Fluctuations	305
9.7.	Screening regimes	308
	References	308
10	Phonon processes	310
10.1.	Introduction	310
10.2.	Three-phonon processes	312
10.2.1.	Coupling constants	312
10.2.2.	Selection rules for acoustic phonons	313

Contents	xvii
10.2.3. Rates for LA modes via normal processes	314
10.2.4. Rates for TA modes via normal processes	317
10.2.5. Rates for umklapp processes	318
10.2.6. Higher-order processes	319
10.2.7. Lifetime of optical phonons	320
10.3. Scattering by imperfections	321
10.4. Scattering by charged impurities	323
10.5. Scattering by electrons	325
10.6. Other scattering mechanisms	326
References	327
11 Quantum transport	328
11.1. The density matrix	328
11.2. Screening	330
11.3. The two-level system	333
11.4. Fermi's Golden Rule	334
11.5. Wannier–Stark states	336
11.6. The intracollisional field effect	337
11.7. The semi-classical approximation	338
References	338
12 Semi-classical transport	339
12.1. The Boltzmann equation	339
12.2. Weak electric fields	344
12.3. Electron–electron scattering	347
12.4. Hot electrons	350
12.5. Hot electron distribution functions	353
12.5.1. Scattering by non-polar acoustic phonons	355
12.5.2. Scattering by non-polar optical modes	357
12.5.3. The drifted Maxwellian	358
References	362
13 Space-charge waves	364
13.1. Phenomenological equations	364
13.2. Space-charge and acoustoelectric waves	366
13.3. Parametric processes	370
13.4. Domains and filaments	371
13.5. Recombination waves	375
References	378
14 Hot phonons	379
14.1. Introduction	379
14.2. Rate equations	380
14.3. Lifetime	381
14.4. Dependence on lattice temperature	382
14.5. Coupled modes	383

14.6.	Lifetime dispersion	387
14.7.	Migration	390
14.8.	Role of daughter modes	390
	References	393
15	Spin processes	394
15.1.	Introduction	394
15.2.	Band structure	395
15.3.	Valence band eigenfunctions	398
15.4.	Conduction band eigenfunctions	399
15.5.	The Elliot–Yafet process	401
15.6.	The D’yakonov–Perel process	404
15.7.	The Rashba mechanism	406
15.8.	The Bir–Aranov–Pikus mechanism	407
15.9.	Hyperfine coupling	409
15.10.	Summary	409
15.11.	Optical generation	410
	References	412
16	Surfaces and interfaces	413
16.1.	Introduction	413
16.2.	The Kronig–Penney model	414
16.3.	Tamm states	415
16.4.	Virtual gap states	416
16.5.	The dielectric band gap	418
16.6.	The Schottky contact	419
	References	423
	Author Index	424
	Subject Index	427

Band structure of semiconductors



1.1. The crystal Hamiltonian

For an assembly of atoms the classical energy is the sum of the following:

- (a) the kinetic energy of the nuclei;
- (b) the potential energy of the nuclei in one another's electrostatic field;
- (c) the kinetic energy of the electrons;
- (d) the potential energy of the electrons in the field of the nuclei;
- (e) the potential energy of the electrons in one another's field;
- (f) the magnetic energy associated with the spin and the orbit.

Dividing the electrons into core and valence electrons and leaving out magnetic effects leads to the following expression for the crystal Hamiltonian:

$$H = \sum_l \frac{\mathbf{p}_l^2}{2M_l} + \sum_{l,m} U(\mathbf{R}_l - \mathbf{R}_m) + \sum_i \frac{\mathbf{p}_i^2}{2m} + \sum_{i,l} V(\mathbf{r}_i - \mathbf{R}_l) + \sum_{i,j} \frac{e^2/4\pi\epsilon_0}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1.1)$$

where l and m label the ions, i and j label the electrons, \mathbf{p} is the momentum, M is the ionic mass, m is the mass of the electron, $U(\mathbf{R}_l - \mathbf{R}_m)$ is the interionic potential, and $V(\mathbf{r}_i - \mathbf{R}_l)$ is the valence-electron-ion potential.

The Schrödinger equation determines the time-independent energies of the system:

$$H\mathcal{E} = E\mathcal{E} \quad (1.2)$$

where H is now the Hamiltonian operator.

1.2. Adiabatic approximation

The mass of an ion is at least a factor of 1.8×10^3 greater than that of an electron, and for most semiconductors the factor is well over 10^4 . For comparable energies and perturbations ions therefore move some 10^2 times slower than do electrons, and the latter can be regarded as

instantaneously adjusting their motion to that of the ions. Therefore the total wavefunction is approximately of the form

$$\mathcal{E} = \Psi(\mathbf{r}, \mathbf{R})\Phi(\mathbf{R}) \quad (1.3)$$

where $\Phi(\mathbf{R})$ is the wavefunction for all the ions and $\Psi(\mathbf{r}, \mathbf{R})$ is the wavefunction for all the electrons instantaneously dependent on the ionic position.

The Schrödinger equation can be written

$$\Psi(\mathbf{r}, \mathbf{R})H_L\Phi(\mathbf{R}) + \Phi(\mathbf{R})H_e\Psi(\mathbf{r}, \mathbf{R}) + H'\Psi(\mathbf{r}, \mathbf{R})\Phi(\mathbf{R}) = E\Psi(\mathbf{r}, \mathbf{R})\Phi(\mathbf{R}) \quad (1.4)$$

where

$$H'\Psi(\mathbf{r}, \mathbf{R})\Phi(\mathbf{R}) = H_L\Psi(\mathbf{r}, \mathbf{R})\Phi(\mathbf{R}) - \Psi(\mathbf{r}, \mathbf{R})H_L\Phi(\mathbf{R}) \quad (1.5)$$

$$H_L = \sum_l \frac{\mathbf{p}_l^2}{2M_l} + \sum_{l,m} U(\mathbf{R}_l - \mathbf{R}_m) \quad (1.6)$$

$$H_e = \sum_i \frac{\mathbf{p}_i^2}{2m} + \sum_{i,l} V(\mathbf{r}_i - \mathbf{R}_l) + \sum_{i,j} \frac{e^2/4\pi\epsilon_0}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (1.7)$$

The relative contribution of H' is of the order m/M_l . The adiabatic approximation consists of neglecting this term. In this case eqn (1.4) splits into a purely ionic equation

$$H_L\Phi(\mathbf{R}) = E_L\Phi(\mathbf{R}) \quad (1.8)$$

and a purely electronic equation

$$H_e\Psi(\mathbf{r}, \mathbf{R}) = E_e\Psi(\mathbf{r}, \mathbf{R}). \quad (1.9)$$

1.3. Phonons

Provided that the ions do not move far from their equilibrium positions in the solid their motion can be regarded as simple harmonic. If the equilibrium position of an ion is denoted by the vector \mathbf{R}_{l0} and its displacement by \mathbf{u}_l , the Hamiltonian can be written

$$H_L = \sum_l \frac{\mathbf{p}_l^2}{2M_l} + \sum_{l,m} D_{l,m}(\mathbf{R}_l - \mathbf{R}_m)\mathbf{u}_l\mathbf{u}_m + H_{L0}(\mathbf{R}_{l0}) + H'_L \quad (1.10)$$

where $D_{lm}(\mathbf{R}_l - \mathbf{R}_m)$ is the restoring force per unit displacement, $H_{L0}(\mathbf{R}_{l0})$ is an additive constant dependent only on the equilibrium separation of the ions, and H'_L represents the contribution of anharmonic forces. The displacements can be expanded in terms of the normal modes of vibration of the solid. The latter take the form of longitudinally polarized and transversely polarized acoustic waves plus, in the case of lattices with a basis, i.e. more than one atom per primitive unit cell, longitudinally and transversely polarized 'optical' modes. (See Section 3.9 for an account of the theory for long-wavelength acoustic modes.) Ionic motion therefore manifests itself in terms of travelling plane waves

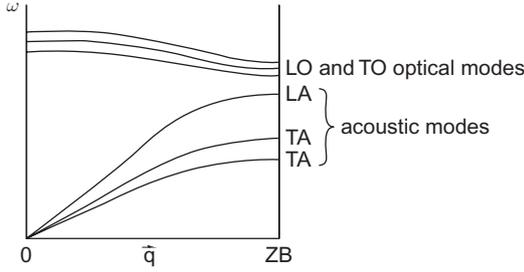


FIG. 1.1.
Dispersion of lattice waves.

$$\mathbf{u}(\omega, \mathbf{q}) = \mathbf{u}_0 \exp\{i(\mathbf{q} \cdot \mathbf{r} - \omega t)\} \quad (1.11)$$

which interact weakly with one another through the anharmonic term H'_L . Figure 1.1 shows the typical dispersion relation between ω and \mathbf{q} .

The energy in a mode is given by

$$E(\omega, \mathbf{q}) = \left\{ n(\omega, \mathbf{q}) + \frac{1}{2} \right\} \hbar \omega \quad (1.12)$$

where $n(\omega, \mathbf{q})$ is the statistical average number of phonons, i.e. vibrational quanta, excited. At thermodynamic equilibrium $n(\omega, q) = n(\omega)$ is given by the Bose–Einstein function for a massless particle

$$n(\omega) = \frac{1}{\exp(\hbar\omega/k_B T) - 1}. \quad (1.13)$$

The following points should be noted.

1. The limits of \mathbf{q} according to periodic boundary conditions are $2\pi/Na$ and the Brillouin zone boundary, where N is the number of unit cells of length a along the cavity.
2. The magnitude of a wavevector component is $2\pi l/Na$, where l is an integer. The curves in Fig. 1.1 are really closely spaced points.
3. An impurity or other defect may introduce localized modes of vibration in its neighbourhood if its mass and binding energy are different enough from those of its host.
4. For long-wavelength acoustic modes $\omega = v_s q$. For others it is often useful to approximate their dispersion by $\omega = \text{constant}$.

1.4. The one-electron approximation

If the electron–electron interaction is averaged we can regard any deviation from this average as a small perturbation. Thus we replace the repulsion term as follows:

$$\sum_{i,j} \frac{(e^2/4\pi\epsilon_0)}{|\mathbf{r}_i - \mathbf{r}_j|} = H_{e0} + H_{ee} \quad (1.14)$$

where H_{e0} contributes a constant repulsive component to the electronic energy and H_{ee} is a fluctuating electron–electron interaction which can be regarded as small. If H_{ee} is disregarded each electron reacts independently with the lattice of ions. Consequently we can take

$$\Psi(\mathbf{r}, \mathbf{R}) = \prod_i \psi_i(\mathbf{r}_i, \mathbf{R}) \quad (1.15)$$

with the proviso that the occupation of the one-electron states is in accordance with the Pauli exclusion principle. We obtain the one-electron Schrödinger equation

$$H_{ei} \psi_i(\mathbf{r}_i, \mathbf{R}) = E_{ei} \psi_i(\mathbf{r}_i, \mathbf{R}) \quad (1.16)$$

where

$$H_{ei} = \frac{\mathbf{p}_i^2}{2m} + \sum_l V(\mathbf{r}_i - \mathbf{R}_l). \quad (1.17)$$

This Hamiltonian still depends on the fluctuating position of ions, and it is useful to reduce the Hamiltonian into one that depends on the interaction with the ions in their equilibrium positions with the effect of ionic vibrations taken as a perturbation. Thus we take

$$H_{ei} = H_{e0i} + H_{ep} \quad (1.18)$$

$$H_{e0i} = \frac{\mathbf{p}_i^2}{2m} \sum_l V(\mathbf{r}_i - \mathbf{R}_{l0}) \quad (1.19)$$

where the H_{ep} is the electron–phonon interaction. The electronic band structure is obtained from (dropping the subscripts i and e)

$$\left\{ \frac{\mathbf{p}^2}{2m} + \sum_l V(\mathbf{r} - \mathbf{R}_{l0}) \right\} \psi(\mathbf{r}) = E \psi(\mathbf{r}). \quad (1.20)$$

1.5. Bloch functions

In the case of a perfectly periodic potential the eigenfunction is a Bloch function:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \mathbf{u}_{n\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (1.21)$$

$$u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}) \quad (1.22)$$

where \mathbf{R} is a vector of the Bravais lattice, n labels the band and \mathbf{k} is the wavevector of the electron in the first Brillouin zone (Fig. 1.2). From eqns (1.21) and (1.22) it follows that

$$\psi_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = \psi_{n\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{R}). \quad (1.23)$$

If a macroscopic volume V is chosen whose shape is a magnified version of the primitive cell, then we can apply the periodic boundary condition

$$\psi_{n\mathbf{k}}(\mathbf{r} + N\mathbf{a}) = \psi_{n\mathbf{k}}(\mathbf{r}) \quad (1.24)$$

where \mathbf{a} is a vector of the unit cell and N is the number of unit cells along the side of V in the direction of \mathbf{a} . This decouples the properties of the wavefunction from the size of a crystal, provided the crystal is macroscopic. Equations (1.23) and (1.24) constrain \mathbf{k} such that

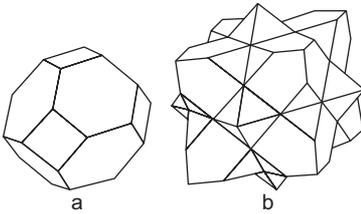


Fig. 1.2.

The first and second zones for a face-centred cubic lattice. The first has half the volume of the cube that is determined by extending the six square faces. The second has the same volume as this cube.

$$\exp(i\mathbf{k} \cdot N\mathbf{a}) = 1.$$

Therefore

$$\mathbf{k} \cdot N\mathbf{a} = 2\pi n \quad (1.25)$$

where n is an integer. In terms of reciprocal lattice vectors \mathbf{K} , defined by

$$\mathbf{K}_i \cdot \mathbf{a}_j = 2\pi \delta_{ij} \quad (1.26)$$

the electronic wavevector assumes the values

$$\mathbf{k} = \frac{n_1}{N_1} \mathbf{K}_1 + \frac{n_2}{N_2} \mathbf{K}_2 + \frac{n_3}{N_3} \mathbf{K}_3. \quad (1.27)$$

Thus the volume of an electronic state in \mathbf{k} -space is given by

$$\Delta k_1 \Delta k_2 \Delta k_3 = (2\pi)^3 / V. \quad (1.28)$$

If \mathbf{q} is any vector that satisfies the periodic boundary conditions then the wavefunction can be written generally as an expansion in plane waves:

$$\psi(\mathbf{r}) = \sum_{\mathbf{q}} c_{\mathbf{q}} \exp(i\mathbf{q} \cdot \mathbf{r}). \quad (1.29)$$

This general expansion can be related to the Bloch form by putting $\mathbf{q} = \mathbf{k} - \mathbf{K}$ where \mathbf{k} is not necessarily confined to the first Brillouin zone:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r}) \sum_{\mathbf{K}} c_{\mathbf{k}-\mathbf{K}} \exp(-i\mathbf{K} \cdot \mathbf{r}) \quad (1.30)$$

and thus

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{K}} c_{\mathbf{k}-\mathbf{K}} \exp(-i\mathbf{K} \cdot \mathbf{r}). \quad (1.31)$$

Yet another form for a Bloch function can be formed out of functions $\phi_n(\mathbf{r} - \mathbf{R})$ which are centred at the lattice points \mathbf{R} . These are known as Wannier functions. The relation between Bloch and Wannier functions is

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} \phi_n(\mathbf{r} - \mathbf{R}) \exp(i\mathbf{k} \cdot \mathbf{R}). \quad (1.32)$$

This is a useful formulation for describing narrow energy bands when the Wannier function can be approximated by atomic orbitals in the tight-binding approximation.

Since the Bloch functions are eigenfunctions of the one-electron Schrödinger equation they are orthogonal to one another, *viz.*

$$\int \psi_{n'\mathbf{k}'}^* \psi_{n\mathbf{k}} \mathbf{d}\mathbf{r} = \delta_{n'n} \delta_{\mathbf{k}'\mathbf{k}} \quad (1.33)$$

with

$$\psi_{n\mathbf{k}} = \frac{1}{V^{1/2}} u_{n\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}). \quad (1.34)$$

1.6. Nearly-free-electron model

When the periodic potential is very weak the valence electron is almost free, and hence

$$E_{\mathbf{k}} \approx \hbar^2 k^2 / 2m. \quad (1.35)$$

In the cases of semiconductors with diamond and sphalerite structure there are two atoms in each primitive cell and eight valence electrons. Therefore there have to be four valence bands with two electrons of opposing spin in each state. By allowing \mathbf{k} to extend beyond the first zone, we can work out the total width of the four valence bands by equating it with the Fermi energy E_F for a free-electron gas of the same density as the valence electrons. Observations of soft X-ray emission confirm that the width of the valence band in these semiconductors is indeed close to E_F . Thus it is reasonable to assume that the valence electrons are almost free, and eqn (1.35) is a good approximation to the energy provided we take into account the effect of the lattice.

Restricting \mathbf{k} to the first Brillouin zone (Fig. 1.2) we obtain

$$E_{\mathbf{k}} \approx \hbar^2 q^2 / 2m \quad (1.36)$$

$$\mathbf{q} = \mathbf{k} + \mathbf{K}. \quad (1.37)$$

The first band is obtained for $\mathbf{K} = (0, 0, 0)2\pi/a$, and is obviously parabolic. At the zone boundary there is an energy gap in general. The second band is obtained from the smallest non-zero reciprocal lattice vectors, which are $\mathbf{K}_1 = (1, 1, 1)2\pi/a$ and its cubic fellows (e.g. $(-1, 1, 1)2\pi/a$) and $\mathbf{K}_2 = (2, 0, 0)2\pi/a$ and its cubic fellows (e.g. $(0, 2, 0)2\pi/a$). At the zone boundary along the $\langle 100 \rangle$ direction $q = K_2/2 = 2\pi/a$ and $k = -K_2/2 = -2\pi/a$. As q increases k moves towards zero, reaching it when $q = K_2 = 4\pi/a$. At the zone boundary along the $\langle 111 \rangle$ direction $q = K_1/2 = \sqrt{3}\pi/a$ and $k = -K_1/2 = -\sqrt{3}\pi/a$. As q increases, k moves to zero, reaching it when $q = K_1 = 2\sqrt{3}\pi/a$. The band continues to be parabolic in both directions, except close to the zone boundaries.

The first and second bands are parabolic directions because the appropriate reciprocal lattice vector simply subtracts from q . Bands 3 and 4 are not that simple because \mathbf{K}_1 and \mathbf{K}_2 are neither parallel nor anti-parallel in this case. The region in reciprocal lattice space which contains the first four Brillouin zones is the Jones zone (Fig. 1.3).

Bands 1 and 2 reach the surface of the Jones zone at the points $(2, 0, 0)$ and $(1, 1, 1)$. Bands 3 and 4 are associated with combinations of \mathbf{k} , \mathbf{K}_1 , and \mathbf{K}_2 which keep \mathbf{q} close to the zone boundary for all \mathbf{k} . The smallest \mathbf{q} corresponds to the centre of a face $\mathbf{q} = \mathbf{K}_1 - \mathbf{K}_2/2$ ($q = 2\sqrt{2}\pi/a$). With \mathbf{k} along the $\langle 100 \rangle$ direction the band is described by $\mathbf{q} = \mathbf{K}_1 - \mathbf{k}$. When $\mathbf{k} = 0$, $\mathbf{q} = \mathbf{K}_1$ ($q = 2\sqrt{3}\pi/a$). Thus q changes by an amount $\sqrt{3} - \sqrt{2}$ in units of $2\pi/a$ as \mathbf{k} sweeps through the zone in the $\langle 100 \rangle$ direction, and hence the energy changes very little with \mathbf{k} . This band is far from being free-electron-like. The other band is also flat, for again q changes comparatively little with \mathbf{k} because \mathbf{k} is more or less perpendicular to the reciprocal vector.

The free-electron model predicts different energies at the \mathbf{q} corresponding to the corners of the Jones zone. We have already seen that

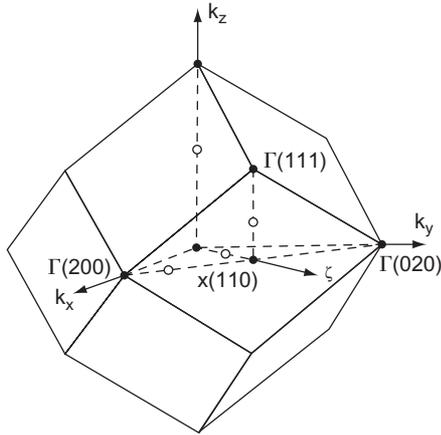


FIG. 1.3. The Jones zone for face-centred cubic crystals containing eight electrons per cell includes the first four Brillouin zones.

$\mathbf{q} = \mathbf{K}_1$ at one. Along the cube-edge direction $\mathbf{q} = \mathbf{K}_2(q = 4\pi/a)$, which has a greater magnitude. Thus the energy of the latter is greater in the free-electron model. The periodic potential with its cubic symmetry (in the case under discussion) alters that picture to one in which these energies are identical. This common energy is the maximum attained at the surface of the Jones zone, which is a free-electron-like result since the corners are indeed furthest from the zone centre. The minimum energy for bands 3 and 4 appearing at $\mathbf{k} = \mathbf{K}_2/2$, corresponding to the middle of the face which is the nearest point of the surface to the zone centre, is equally free-electron like. The valence band has the general form depicted in Fig. 1.4.

Pushing the free-electron model beyond the Jones zone in order to describe the conduction band is simplest, as in the case of the valence band, when \mathbf{k} and \mathbf{K} are in the same direction. In bands 3 and 4 and in bands 5 and 6 this will correspond to directions perpendicular to the surface of the Jones zone, and in these directions the mass of the electron will be near that for the free electron and the bands will be parabolic. However, in the principal directions $\langle 100 \rangle$ and $\langle 111 \rangle$, for which \mathbf{k} lies more or less parallel to the surface of the Jones zone, bands 5 and 6 will be flat like bands 3 and 4, and for the same reason. They might also be expected to be separated from the valence bands by an energy gap which remains constant in these directions of \mathbf{k} , and this is approximately true. However, unlike the valence bands, bands 5 and 6

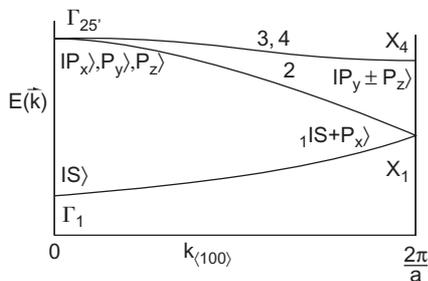


FIG. 1.4. The general form of the valence band. The symmetry symbols and orbitals are appropriate for diamond and silicon.

are not degenerate. Moreover, their form, particularly near the corners and edges of the Jones zone, is much more complex, and minima at $\mathbf{k} = 0(\Gamma)$, $\mathbf{k} = \mathbf{K}_2/2(\text{X})$, and $\mathbf{k} = \mathbf{K}_1/2(\text{L})$ are general features. Because the minimum energy of bands 3 and 4 is at the X point and because the energy gap is roughly constant it is expected that the lowest minimum of the conduction band also lies at X, but this is only roughly correct and holds only for semiconductors involving elements in the early rows of the periodic table.

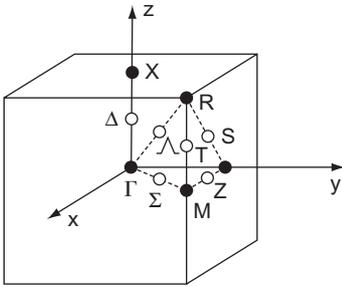


Fig. 1.5. Special points in the Brillouin zone of a simple cubic lattice.

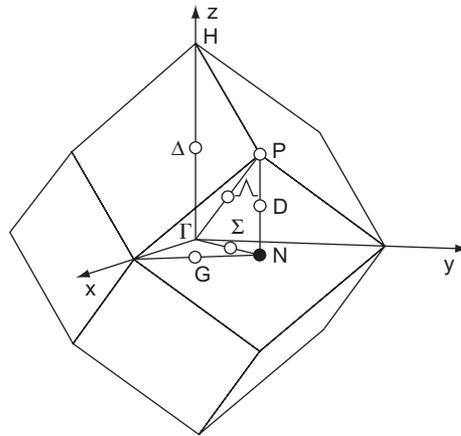


Fig. 1.6. Special points in the Brillouin zone of a body-centred cubic lattice.

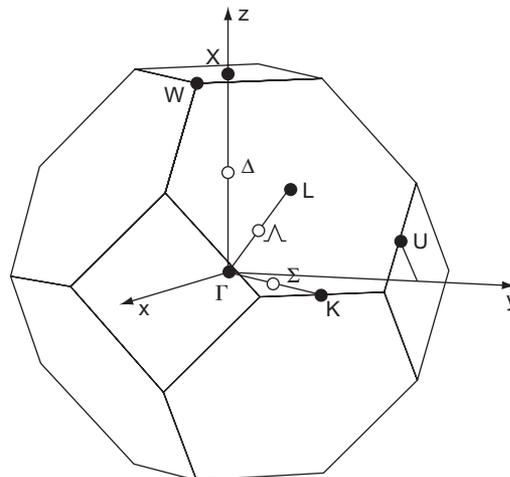


Fig. 1.7. Special points in the Brillouin zone of a face-centred cubic lattice.

1.6.1. Group theory notation

Many features of band structure depend upon symmetry, and in particular the symmetry of the cubic lattice. The special points in the Brillouin zone of a simple lattice are shown in Fig. 1.5 (see also Figs. 1.6 and 1.7). The symmetry types as illustrated by simple basis functions are given in Table 1.1. The representations for the Γ group are given in Table 1.2.

TABLE 1.1. Symmetry types and basis functions of a cubic lattice.

Points	Notation	Basis functions		
Γ, R	Γ_1^a	or A_{1g}^b	1	
	Γ_2	A_{2g}	$x^4(y^2 - z^2) + y^4(z^2 - x^2) + z^4(x^2 - y^2)$	
	Γ_{12}	E_g	$z^2 - \frac{1}{2}(x^2 + y^2), x^2 - y^2$	
	Γ_{15}'	T_{1g}	$xy(x^2 - y^2), yz(y^2 - z^2), zx(z^2 - x^2)$	
	Γ_{25}'	T_{2g}	xy, yz, zx	
	Γ_1'	A_{1u}	$xyz \{x^4(y^2 - z^2) + y^4(z^2 - x^2) + z^4(x^2 - y^2)\}$	
	Γ_2'	A_{20}	xyz	
	Γ_{12}'	E_u	$xyz \{z^2 - \frac{1}{2}(x^2 + y^2)\}, xyz(x^2 - y^2)$	
	Γ_{15}	T_{1u}	x, y, z	
	Γ_{25}	T_{2u}	$z(x^2 - y^2), x(y^2 - z^2), y(z^2 - x^2)$	
	X, M	X_1		1
		X_2		$x^2 - y^2$
X_3			xy	
X_4			$xy(x^2 - y^2)$	
X_5			yz, zx	
$X_{1'}$			$xyz(x^2 - y^2)$	
$X_{2'}$			xyz	
$X_{3'}$			$z(x^2 - y^2)$	
$X_{4'}$			z	
$X_{5'}$			x, y	
Δ, T	Δ_1		1	
	Δ_2		$x^2 - y^2$	
	Δ_3		xy	
	Δ_4		$xy(x^2 - y^2)$	
	Δ_5		x, y	
Σ, S	Σ_1		1	
	Σ_2		$z(x - y)$	
	Σ_3		z	
	Σ_4		$x - y$	
Z	z_1		1	
	z_2		yz	
	z_3		y	
	z_4		z	
Λ	Λ_1		1	
	Λ_2		$xy(x - y) + yz(y - z) + zx(z - x)$	
	Λ_3		$x - z, y - z$	

From Kittel 1963.

^aNotation of Bouckaert, Smoluchowski, and Wigner 1936.

^bChemical notation: g stands for even function, u for odd function, under inversion.

TABLE 1.2. Γ Group representation with spin.

Without spin	$\Gamma_1 \Gamma_2 \Gamma_{12}$	Γ_{15}'	Γ_{25}'	$\Gamma_1' \Gamma_2' \Gamma_{12}'$	Γ_{15}	Γ_{25}
With spin	$\Gamma_6 \Gamma_7 \Gamma_8$	$\Gamma_6 + \Gamma_8$	$\Gamma_7 + \Gamma_8$	$\Gamma_6' \Gamma_7' \Gamma_8'$	$\Gamma_6' + \Gamma_8'$	$\Gamma_7' + \Gamma_8'$

 Γ_6 and Γ_7 are doublets; Γ_8 is a quartet.

1.7. Energy gaps

The periodic potential of the crystal (eqn (1.20)) can be expanded in a Fourier series:

$$\sum_l V(\mathbf{r} - \mathbf{R}_l) = \sum_{\mathbf{K}'} V_{\mathbf{K}'} \exp(i\mathbf{K}' \cdot \mathbf{r}) \quad (1.38)$$

and a constant potential can be added to make $V_{\mathbf{K}'} = 0$ when $\mathbf{K}' = 0$. If the Bloch function is taken to be of the form (eqn (1.30))

$$\psi_{\mathbf{k}} = \sum_{\mathbf{K}} c_{\mathbf{k}-\mathbf{K}} \exp[i(\mathbf{k} - \mathbf{K}) \cdot \mathbf{r}] \quad (1.39)$$

the Schrödinger equation yields[†]

$$\begin{aligned} & \sum_{\mathbf{K}, \mathbf{K}'} V_{\mathbf{K}'} c_{\mathbf{k}-\mathbf{K}} \exp[i(\mathbf{k} - \mathbf{K} + \mathbf{K}') \cdot \mathbf{r}] \\ & = \sum_{\mathbf{K}, \mathbf{K}'} V_{\mathbf{K}'-\mathbf{K}} c_{\mathbf{k}-\mathbf{K}'} \exp[i(\mathbf{k} - \mathbf{K}) \cdot \mathbf{r}] \end{aligned} \quad \left\{ \frac{\hbar^2}{2m} (\mathbf{k} - \mathbf{K})^2 - E(\mathbf{k}) \right\} c_{\mathbf{k}-\mathbf{K}} + \sum_{\mathbf{K}'} V_{\mathbf{K}'-\mathbf{K}} c_{\mathbf{k}-\mathbf{K}'} = 0 \quad (1.40)$$

which will hold for each \mathbf{k} and for each \mathbf{K} . When the periodic potential is weak the solution of eqn (1.40) is approximately given by

$$E(\mathbf{k}) \approx \frac{\hbar^2}{2m} (\mathbf{k} - \mathbf{K})^2 \quad c_{\mathbf{k}-\mathbf{K}} \neq 0 \quad (1.41)$$

or

$$E(\mathbf{k}) \neq \frac{\hbar^2}{2m} (\mathbf{k} - \mathbf{K})^2 \quad c_{\mathbf{k}-\mathbf{K}} \approx 0. \quad (1.42)$$

In the non-degenerate case, when eqn (1.41) is satisfied for only one \mathbf{K} , say $\mathbf{K} = \mathbf{K}_a$, a more accurate solution is obtained by noting that for $\mathbf{K} = \mathbf{K}_b$, where $\mathbf{K}_b \neq \mathbf{K}_a$,

$$c_{\mathbf{k}-\mathbf{K}_b} = \frac{\sum_{\mathbf{K}'} V_{\mathbf{K}'-\mathbf{K}_b} c_{\mathbf{k}-\mathbf{K}'}}{E_a - (\hbar^2/2m)(\mathbf{k} - \mathbf{K}_b)^2} \approx \frac{V_{\mathbf{K}_a-\mathbf{K}_b} c_{\mathbf{k}-\mathbf{K}_a}}{E_a - E_b} \quad (1.43)$$

where $E_a = \hbar^2(\mathbf{k} - \mathbf{K}_a)^2/2m$ and $E_b = \hbar^2(\mathbf{k} - \mathbf{K}_b)^2/2m$. Putting this back in eqn (1.40) gives

$$E(\mathbf{k}) = E_a + \sum_{\mathbf{K}'} \frac{V_{\mathbf{K}'-\mathbf{K}_a} V_{\mathbf{K}_a-\mathbf{K}'}}{E_a - E'} = E_a + \sum_{\mathbf{K}'} \frac{|V_{\mathbf{K}'-\mathbf{K}_a}|^2}{E_a - E'}. \quad (1.44)$$

The effect of energy states lying above E_a is to depress $E(\mathbf{k})$, and the effect of those lying below E_a is to raise $E(\mathbf{k})$, i.e. energy levels tend to repel one another.

In the degenerate case, when eqn (1.41) is satisfied for a given eigenvalue $E(\mathbf{k})$ by, say, $\mathbf{K} = \mathbf{K}_a = \mathbf{K}_b$, neglecting all coefficients except those for \mathbf{K}_a and \mathbf{K}_b gives

$$\begin{aligned} \{E_a - E(\mathbf{k})\} c_{\mathbf{k}-\mathbf{K}_a} + V_{\mathbf{K}_b-\mathbf{K}_a} c_{\mathbf{k}-\mathbf{K}_b} &= 0 \\ \{E_b - E(\mathbf{k})\} c_{\mathbf{k}-\mathbf{K}_b} + V_{\mathbf{K}_a-\mathbf{K}_b} c_{\mathbf{k}-\mathbf{K}_a} &= 0, \end{aligned} \quad (1.45)$$

whence, with $K = |\mathbf{K}_b - \mathbf{K}_a|$,

$$\{E_a - E(\mathbf{k})\}\{E_b - E(\mathbf{k})\} - |V_K|^2 = 0.$$

Therefore

$$E(\mathbf{k}) = \frac{1}{2}(E_a + E_b) \pm \frac{1}{2} \left\{ (E_a + E_b)^2 - 4(E_a E_b - |V_K|^2) \right\}^{1/2} \quad (1.46)$$

which for $E_a = E_b = E_0$ becomes

$$E(\mathbf{k}) = E_0 \pm |V_K| \quad (1.47)$$

corresponding to an energy gap of magnitude $2|V_K|$.

When the unit cell consists of two atoms, which are not necessarily identical, the potential can be denoted

$$V(\mathbf{r} - \mathbf{R}_{j0}) = V_1(\mathbf{r} - \mathbf{R}_{j1}) + V_2(\mathbf{r} - \mathbf{R}_{j2}) \quad (1.48)$$

where \mathbf{R}_{j0} , \mathbf{R}_{j1} , and \mathbf{R}_{j2} are the position vectors of the centre of the cell, atom 1, and atom 2 respectively. If $\mathbf{R}_{j1} = \mathbf{R}_{j0} + \mathbf{r}_c$ and $\mathbf{R}_{j2} = \mathbf{R}_{j0} - \mathbf{r}_c$, where \mathbf{r}_c is the covalent bond length (half the interatomic distance), eqn (1.48) becomes

$$V(\mathbf{r} - \mathbf{R}_{j0}) = V_1(\mathbf{r} - \mathbf{r}_c - \mathbf{R}_{j0}) + V_2(\mathbf{r} + \mathbf{r}_c - \mathbf{R}_{j0}) \quad (1.49)$$

and so, instead of eqn (1.38), we can put

$$\begin{aligned} \sum_l V(\mathbf{r} - \mathbf{R}_{j0}) &= \sum_{\mathbf{K}'} \{V_{1\mathbf{K}'} \exp(-i\mathbf{K}' \cdot \mathbf{r}_c) + V_{2\mathbf{K}'} \exp(i\mathbf{K}' \cdot \mathbf{r}_c)\} \exp(i\mathbf{K}' \cdot \mathbf{r}) \\ &= \sum_{\mathbf{K}'} \left(V_{\mathbf{K}'}^S \cos \mathbf{K}' \cdot \mathbf{r}_c + iV_{\mathbf{K}'}^A \sin \mathbf{K}' \cdot \mathbf{r}_c \right) \exp(i\mathbf{K}' \cdot \mathbf{r}) \end{aligned} \quad (1.50)$$

where a division has been made into a symmetric part $V_{\mathbf{K}'}^S = (V_{1\mathbf{K}'} + V_{2\mathbf{K}'})/2$ and an antisymmetric part $V_{\mathbf{K}'}^A = (V_{2\mathbf{K}'} - V_{1\mathbf{K}'})/2$. The latter will be zero if the atoms are identical. The energy gap is therefore given by

$$E_g^2 = 4 \left(\left| V_K^S \cos Kr_c \right|^2 + \left| V_K^A \sin Kr_c \right|^2 \right). \quad (1.51)$$

Thus the energy gap can be regarded as having a symmetric or homopolar component and an antisymmetric or polar component. A division in this way is useful for understanding how the covalent and polar aspects of binding influence the electronic structure of semiconductors, as we shall see.

The eigenfunctions for the two bands (eqn (1.39)), at the band-edges, are of the form

$$\psi_{a,b\mathbf{k}} = \{c_{a\mathbf{k}} \exp(-i\mathbf{K}_a \cdot \mathbf{r}) \pm c_{b\mathbf{k}} \exp(-i\mathbf{K}_b \cdot \mathbf{r})\} \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (1.52)$$

with $|c_{a\mathbf{k}}| = |c_{b\mathbf{k}}|$. In the case of the direct gap between valence and conduction bands we can take $\mathbf{K}_a = 0, \mathbf{k} \rightarrow \mathbf{k} + \mathbf{k}_F$, where \mathbf{k} is now restricted to the first Brillouin zone, \mathbf{k}_F is the Fermi-level vector for the free-electron valence band, and $\mathbf{K}_b \approx 2\mathbf{k}_F$, which in effect assumes the

Jones zone to be a sphere of radius k_F . In this approximation (isotropic model)

$$\psi_{a,bk} = \{c_{ak}\exp(i\mathbf{k}_F \cdot \mathbf{r}) \pm c_{bk}\exp(-i\mathbf{k}_F \cdot \mathbf{r})\}\exp(i\mathbf{k} \cdot \mathbf{r}). \quad (1.53)$$

1.8. Spin-orbit coupling and orbital characteristics

In the crystal Hamiltonian of eqn (1.1) we left out magnetic energy. Even in non-magnetic semiconductors the contribution of magnetic energy, although small compared with purely electrostatic components, is not negligible in relation to the energy gaps. The source of this energy in non-magnetic materials is the interaction between spin and orbit, as described by the one-electron Hamiltonian

$$H_{so} = \frac{\hbar^2}{4m^2c^2}[\nabla V(\mathbf{r} - \mathbf{R}_{l0}) \times \mathbf{p}] \cdot \boldsymbol{\sigma} \quad (1.54)$$

where \mathbf{p} is the momentum operator and $\boldsymbol{\sigma}$ is the spin operator. As in the case of a free atom the effect of spin-orbit coupling is to remove orbital degeneracies.

The degeneracies occurring in the band structure of semiconductors near the principle energy gap are those associated with bands 2, 3, and 4 at $\mathbf{k} = 0$, and between bands 3 and 4 elsewhere in the zone. In order to discover how spin-orbit coupling splits these degeneracies we need to know the orbital characteristics associated with these bands. For diamond, sphalerite, and wurzite type crystals the constituent atoms in their free state possess valence electrons in the atomic $|s\rangle$ and $|p\rangle$ states, and we expect the valence and indeed the lower conduction bands to have $|s\rangle$ - and $|p\rangle$ -like orbital characteristics. (In the case of heavier atoms $|d\rangle$ states are also involved, but we ignore this complication for the present.) The chemical picture of bonding in these lattices involves the production of sp^3 hybridized bonds directed towards the corners of a regular tetrahedron. The two atoms in the unit cell contribute their $|s\rangle$ and $|p\rangle$ orbitals in bonding combinations to give the valence band in which the electron density is high between the atoms, or in anti-bonding combinations to give the conduction band in which the electron density tends to be high at, rather than between, the atoms. We can therefore associate bonding $|s\rangle$ orbital characteristics with the lowest valence band and bonding $|p\rangle$ orbital characteristics with bands 2, 3, and 4 at $\mathbf{k} = 0$, and also anti-bonding $|s\rangle$ orbital characteristics with the conduction band, again at $\mathbf{k} = 0$. Orbital characteristics can be assigned to high symmetry points in the zone on the basis of symmetry (Table 1.3). (Note that the Bloch function in eqn (1.53), which is that for the two-band isotropic model, evidently has a bonding $|s\rangle$ character or an anti-bonding $|s\rangle$ character.) At other points of the zone the orbitals become a mixture of $|s\rangle$ and $|p\rangle$ characteristics.

Returning to the question of spin-orbit splitting, we can write eqn (1.54) as follows for a spherically symmetric field:

$$H_{so} = \frac{\hbar^2}{4m^2c^2} \frac{1}{r} \frac{dV}{dr} (\mathbf{r} \times \mathbf{p}) \cdot \boldsymbol{\sigma} = \lambda \mathbf{L} \cdot \mathbf{S} \quad (1.55)$$

TABLE 1.3. Orbital character of Bloch functions.

Symmetry point	Symmetry symbol	Orbital character
Γ	Γ_1	$ s\rangle$
$(\mathbf{k} = 0)$	Γ_{15}	$ p_x\rangle, p_y\rangle, p_z\rangle$
L, Δ	L_1	$ s\rangle$
$(\mathbf{k} = \mathbf{k}_{(111)})$	L_1	$\frac{1}{\sqrt{3}} p_x + p_y + p_z\rangle$
	L_3	$\begin{cases} \frac{1}{\sqrt{6}} p_x + p_y - 2p_z\rangle \\ \frac{1}{\sqrt{2}} p_x - p_y\rangle \end{cases}$
X, Δ	X_1	$\frac{1}{\sqrt{2}} (s\rangle_A + p_x\rangle_B) \begin{pmatrix} \text{Atom A} \\ \text{Atom B} \end{pmatrix}$
$(\mathbf{k} = \mathbf{k}_{(100)})$	X_3	$\frac{1}{\sqrt{2}} (s\rangle_B + p_x\rangle_A)$
	X_5	$\frac{1}{\sqrt{2}} p_y - p_z\rangle$
		$\frac{1}{\sqrt{2}} p_y + p_z\rangle$

From Bassani 1966.

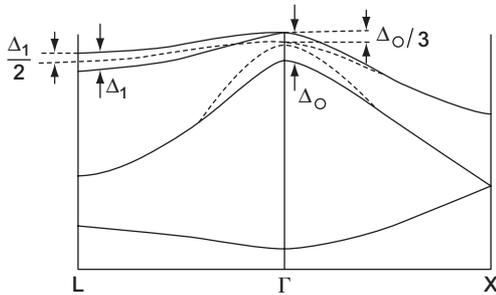
where \mathbf{L} is the orbital angular momentum operator, \mathbf{S} is the spin angular momentum operator, and $\lambda = (dV/dr)/2m^2c^2r$. Since \mathbf{L} and \mathbf{S} combine vectorially to give a total angular momentum \mathbf{J} we can write

$$\mathbf{J}^2 = \mathbf{L}^2 + \mathbf{S}^2 + 2\mathbf{L} \cdot \mathbf{S} \quad (1.56)$$

whence

$$\begin{aligned} \langle \mathbf{L} \cdot \mathbf{S} \rangle &= \frac{1}{2} \langle \mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2 \rangle \\ &= \frac{\hbar^2}{2} \{j(j+1) - l(l+1) - s(s+1)\}. \end{aligned} \quad (1.57)$$

At the top of the valence band are three degenerate $|p\rangle$ -like bands. Thus $l = 1$ and $j = \frac{3}{2}$ or $\frac{1}{2}$, since $j = l \pm s$, and $s = \frac{1}{2}$ for one electron. For $j = \frac{1}{2}$, $\langle \mathbf{L} \cdot \mathbf{S} \rangle = -\hbar^2$, and for $j = \frac{3}{2}$, $\langle \mathbf{L} \cdot \mathbf{S} \rangle = +\hbar^2/2$. Thus the states are split by an amount Δ_0 proportional to $\frac{3}{2}\hbar^2$, the double degenerate $j = \frac{3}{2}$ state moving up $\Delta_0/3$ and the single $j = \frac{1}{2}$ state moving down by $2\Delta_0/3$ (Fig. 1.8). The upper pair remain degenerate

**FIG. 1.8.** Valence band with spin-orbit splitting.

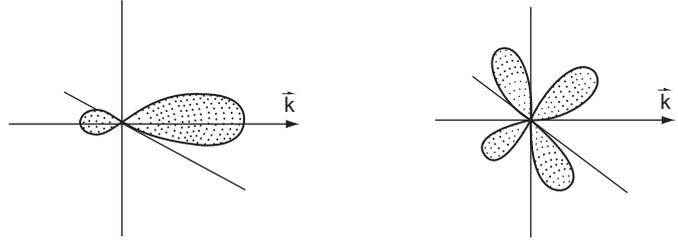


Fig. 1.9.
Orbital characters of valence bands.

at $\mathbf{k} = 0$ and each member retains an orbital characteristic which is $|p\rangle$ like with, for $\mathbf{k} \neq 0$, lobes directed along mutually perpendicular directions which are themselves perpendicular to the \mathbf{k} direction. However, the split-off band assumes an orbital which is $|p\rangle$ like with a lobe along the \mathbf{k} direction when $\mathbf{k} \neq 0$ (Fig. 1.9). When $\mathbf{k} \neq 0$ the degeneracy of the upper bands is removed by spin-orbit splitting, except for \mathbf{k} along the cube edge. The splitting is maximum along the $\langle 111 \rangle$ (Δ) directions and at L it is denoted by Δ_1 . If we take $|p\rangle$ orbitals at the L point we obtain an identical splitting to that at Γ , but since we have only two $|p\rangle$ states and not three we must weight that splitting by a factor of $\frac{2}{3}$. Thus $\Delta_1 = 2\Delta_0/3$, and one band moves up $\Delta_1/2$ and the other down by $\Delta_1/2$ (Fig. 1.8). (In directions other than $\langle 100 \rangle$ and $\langle 111 \rangle$ the degeneracy is removed not only by spin-orbit splitting but by the much larger splitting of the cubic field.)

In silicon the lowest conduction band at Γ is triply degenerate like the top of the valence band and is split only weakly by the spin-orbit interaction. In other semiconductors the conduction band in the vicinity of the gap exhibits non-degenerate valleys at Γ , L, and Δ or X, all of which have an orbital character which is $|s\rangle$ like.

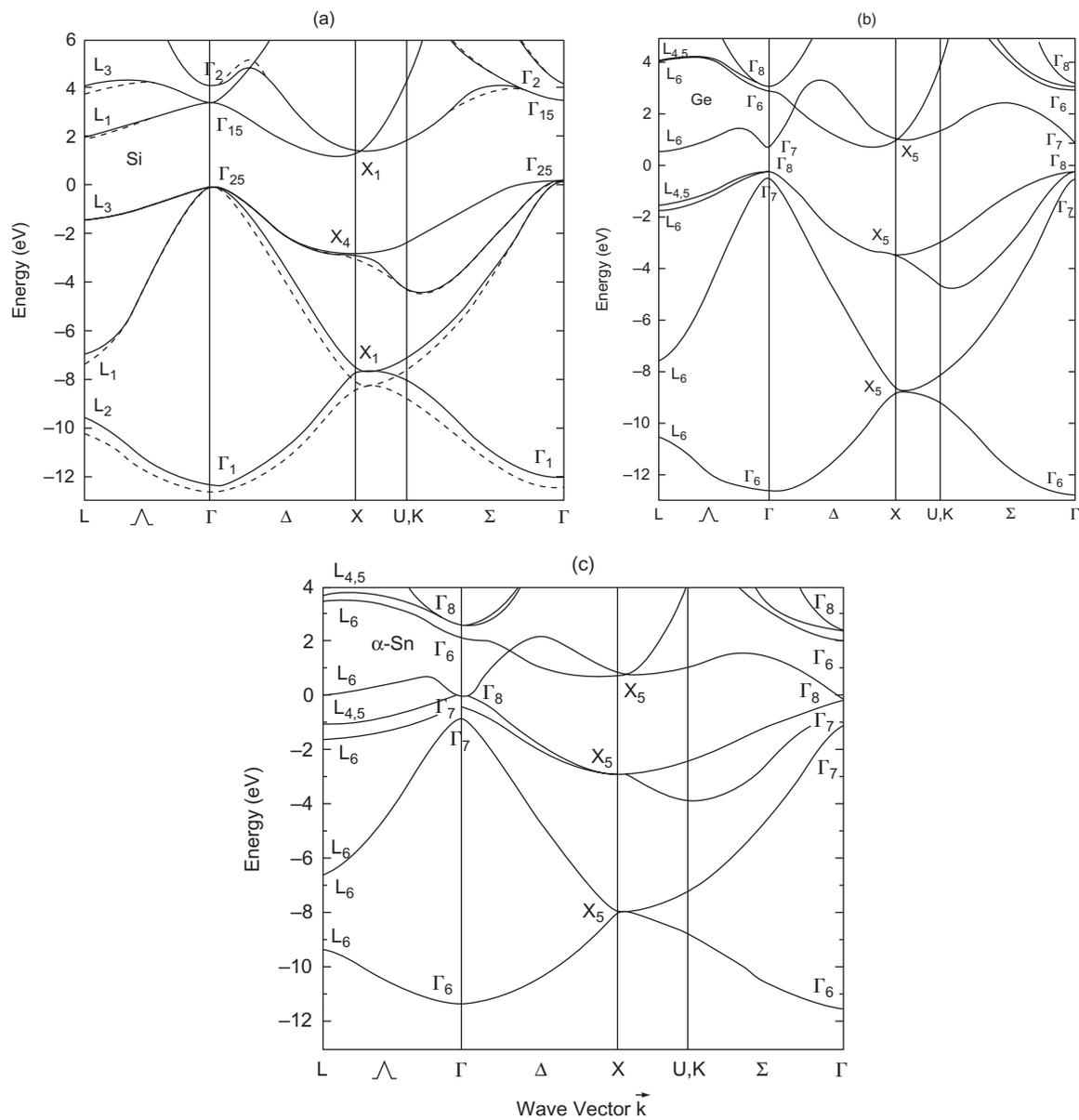
1.9. Band structures

Electronic band structures have been calculated using various models:

1. linear combination of atomic orbitals (LCAO);
2. linear combination of molecular orbitals (LCMO);
3. free-electron approximations;
4. cellular methods;
5. muffin-tin potential (a) augmented plane wave (APW) and (b) Green's function;
6. orthogonalized plane-wave (OPW);
7. pseudopotential.

Models (1) and (2) are tight-binding approximations.

These methods are well reviewed in standard texts. In Figs. 1.10–1.12 we give the results of pseudopotential calculations for Group IV elemental, III–V compound, and II–VI compound semiconductors. Observed energy gaps at 300 K between the top of the valence band and the Γ , L, and X valleys in the conduction band for several Group IV and III–V compound semiconductors are depicted in Fig. 1.13.

**FIG. 1.10.**

Band structures for (a) silicon, (b) germanium, and (c) α -tin. In the case of silicon two results are presented: the non-local pseudopotential (solid curve) and the local pseudopotential (dotted curve). (From Chelikowsky and Cohen 1976.)

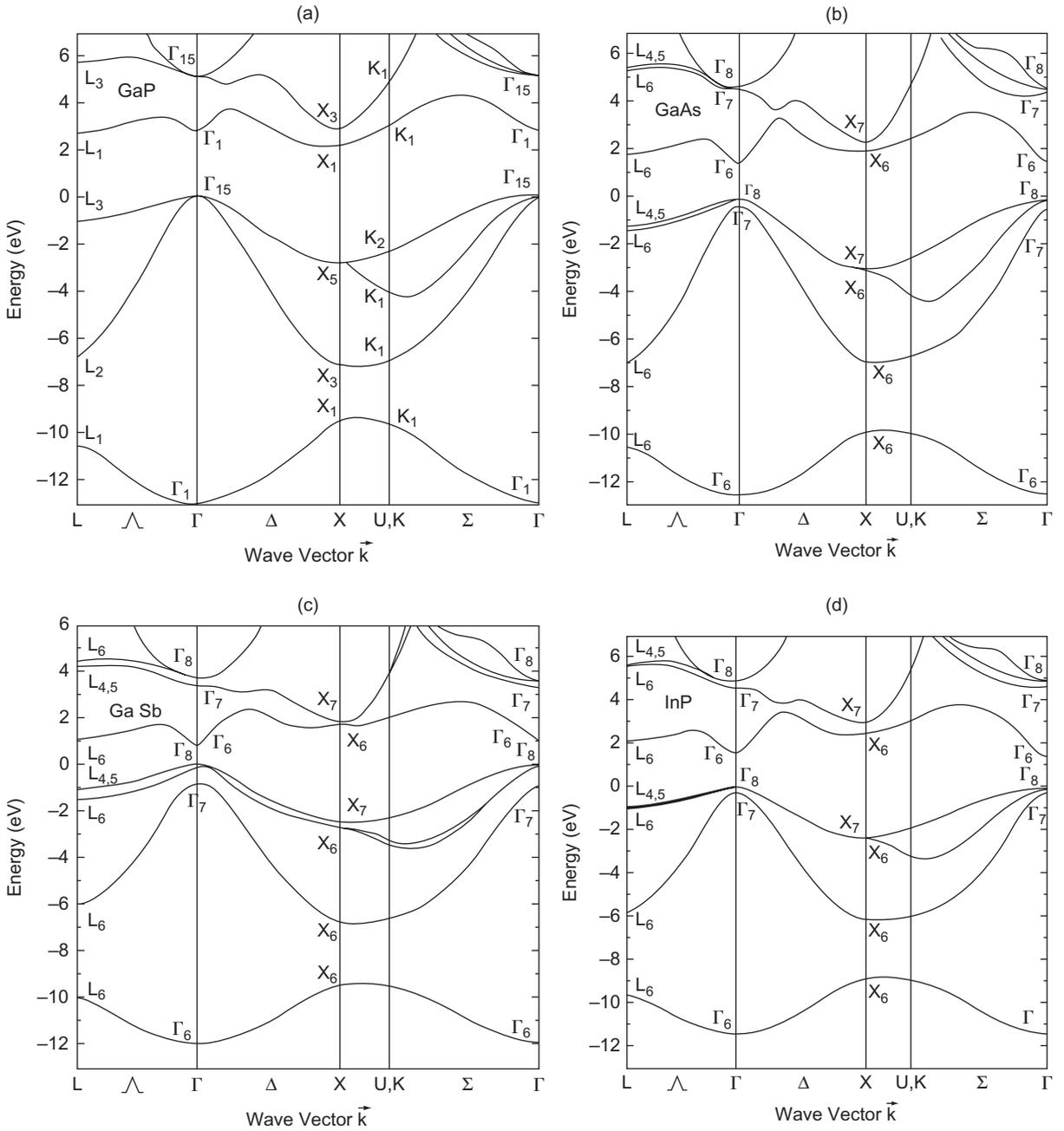


Fig. 1.11. Band structures of III–IV compounds: (a) GaP; (b) GaAs; (c) GaSb; (d) InP; (e) InAs; (f) InSb. (From Chelikowsky and Cohen 1976.)

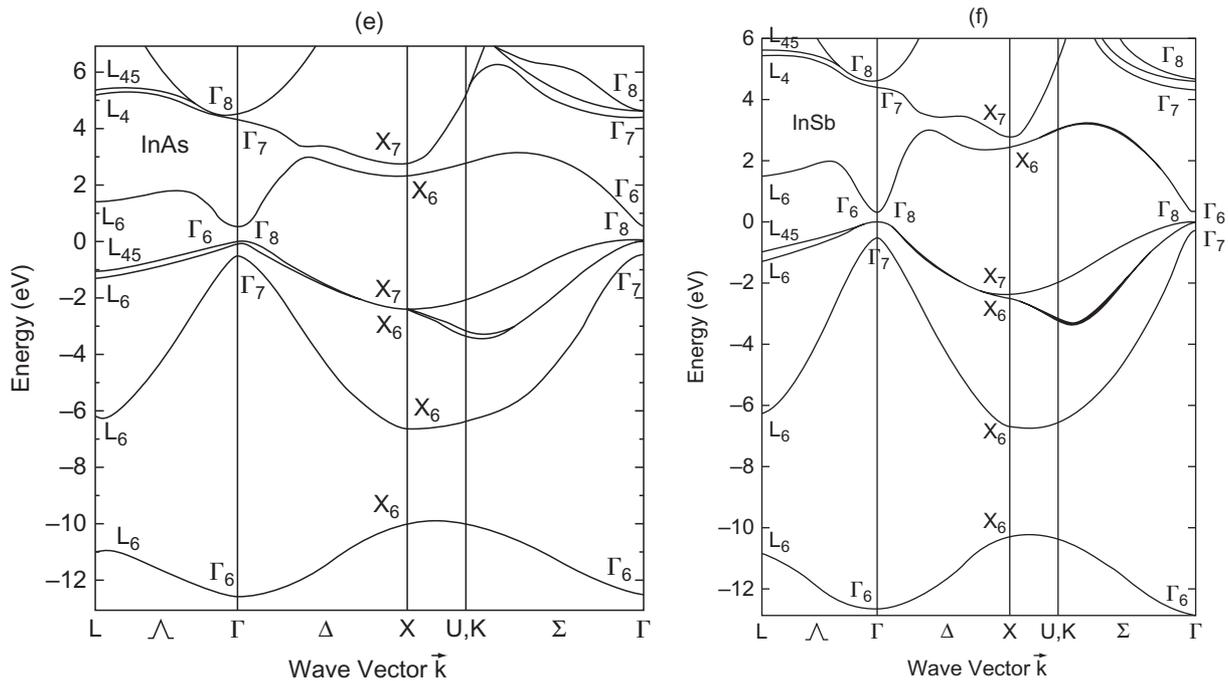


FIG. 1.11.
(continued).

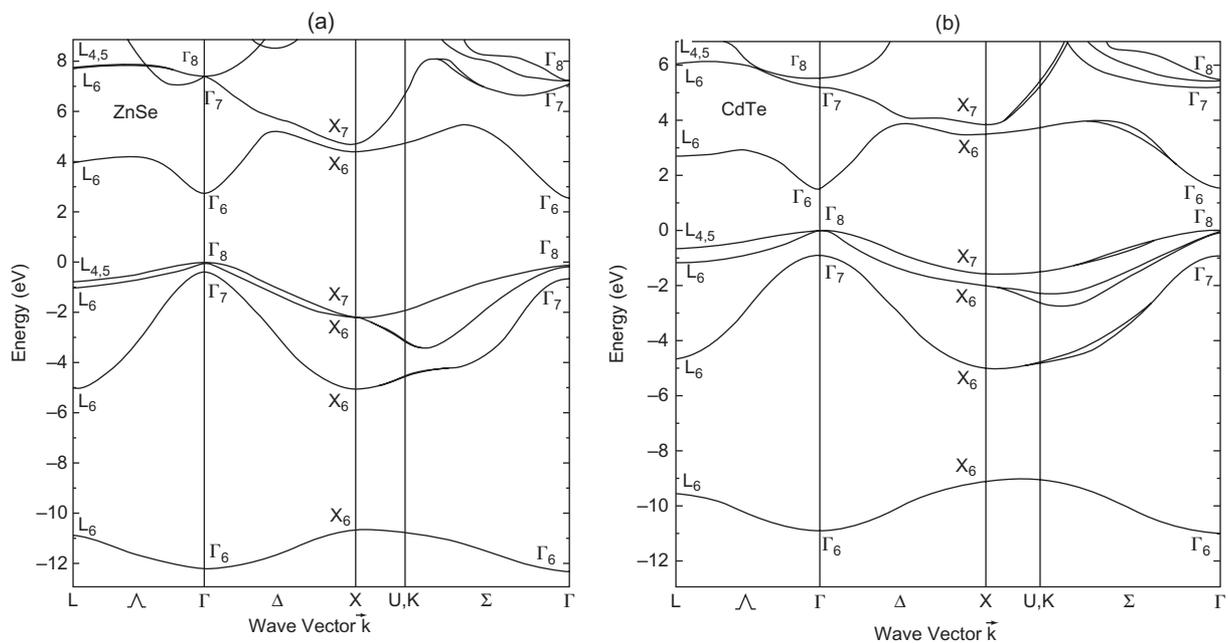


FIG. 1.12.
Band structures for (a) ZnSe and (b) CdTe. (From Chelikowsky and Cohen 1976.)

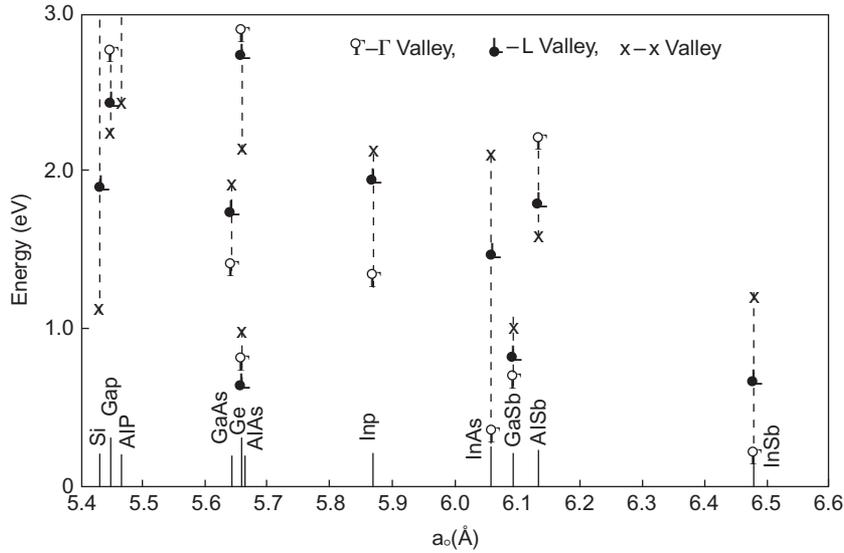


FIG. 1.13.

Energies of Γ , L, and X valleys above the valence band for some Group IV elements and III–IV compounds at 300 K: \circ , Γ valley; \bullet , L valley; \times , X valley.

The major points to be noted are as follows:

1. light atoms tend to have the X valley lowest in agreement with the nearly-free-electron model;
2. heavy atoms tend to have small energy gaps;
3. polar materials tend to have larger energy gaps than non-polar materials;
4. the energy gaps of polar materials tend to be direct gaps, i.e. at the Γ point.
5. germanium is peculiar in having the lowest conduction valley at L.

1.10. Chemical trends

A semi-quantitative account of chemical trends in band structure can be given on the basis of a model developed by Phillips (1968, 1973) and Van Vechten (1969) which is very much in the spirit of Pauling's discussion of the nature of the chemical bond (Pauling 1960).

The energy gaps which are of most importance in semiconductors are those between the valence band and the valleys in the conduction band at the centre of the zone (Γ) and at the zone edge in the $\langle 111 \rangle$ direction (L) and in the $\langle 100 \rangle$ direction (X). These gaps are labelled E_0 , E_1 , and E_2 respectively (Fig. 1.14).

The absolute levels below the vacuum level are fixed by the valence band energy at X, i.e. E_{X4} . This point is distinctive as being at the centre of a face in the Jones zone, and is an obvious candidate to be the central strut about which the band structure is supported.

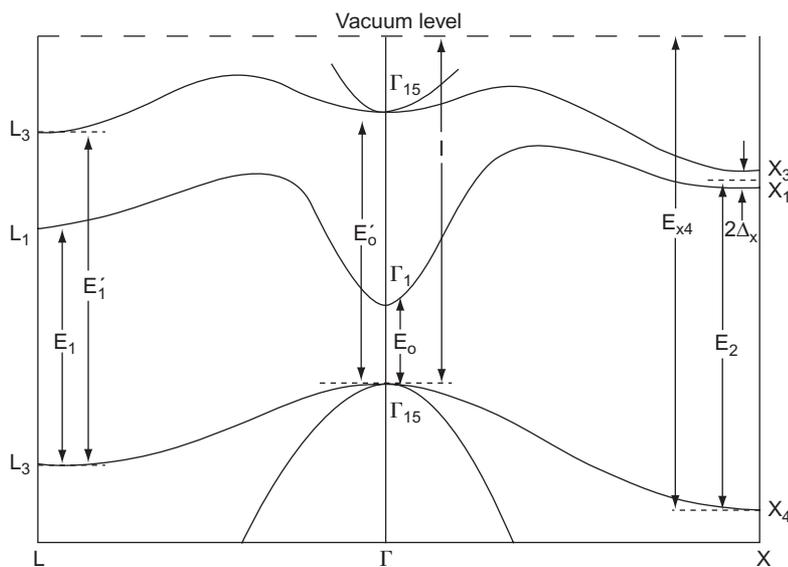


FIG. 1.14.
Principal energy gaps (ignoring spin-orbit splitting).

Increasing polarity causes bands to become narrower, and in the limit to go to the free-atomic level. This level is therefore one which is independent of polarity and about which the valence band broadens. It turns out that E_{X4} is within 5 per cent of the ionization potential for the free atom in the cases of silicon and germanium, and so is a good choice to be the polarity-independent pivot for the band structure. It is therefore assumed that

$$E_{X4} = E_{X4}(\text{Si}) \left\{ \frac{(d_A d_B)^{1/2}}{d_{\text{Si}}} \right\}^{-S} \quad (1.58)$$

where $E_{X4}(\text{Si})$ is the silicon gap and d_{Si} is its covalent interatomic distance (2.35 Å). Since E_{X4} is observed to be close to the free-atom ionization energy for silicon and germanium, it is taken to be a property of the row of the periodic table to which the constituent atoms belong. Thus d_A and d_B are the interatomic distances of the Group IV elements in rows A and B, to which the constituent atoms belong, and these distances are referred to as 'normal' covalent interatomic distances to distinguish them from the actual ones found in compounds AB.

The other basic strut is the ionization energy I . The structure of energy gaps implied by the nearly-free-electron picture (eqn (1.51)) suggests that an energy gap can be considered to be formed of two components, a symmetric or homopolar part and an antisymmetric or polar part. Thus

$$I^2 = I_h^2 + C^2 \quad (1.59)$$

where I_h is the homopolar ionization energy and C is a polar contribution which is dependent on the difference between the electron affinities of the two atoms in the unit cell. Values for C have been given by

Van Vechten (1969). The purely covalent part is taken to depend upon the observed nearest neighbour d as follows:

$$I_h = I(\text{Si}) (d/d_{\text{Si}})^{-S}. \quad (1.60)$$

The absolute level of the valence band at L is simply assumed to be the arithmetic mean of the energies at Γ and X, i.e.

$$E_{L1} = (I + E_{X4})/2 \quad (1.61)$$

All energy gaps are then taken to be of the form of eqn (1.59) provided that no d-band intrudes. A filled d-band affects E_0 and E_1 because the conduction band states with which these gaps are associated have |s> orbitals which penetrate the atomic core, while the |p> orbitals of the valence states do not, and thus the gaps are sensitive to the presence of a full d-shell. However, E'_0 and E'_1 , which involve only |p> orbitals which do not penetrate the core, are not affected. Although the orbital character is |s> like for the X_1 and X_3 bands, it is only half as strongly so as Γ_1 or L_1 , and the effect of the d-band mixing is deemed negligible for E_2 . Thus for the unaffected gaps we have equations of the form

$$E_2^2 = E_{2h}^2 + C^2 \quad (1.62)$$

but for E_0 and E_1 ,

$$E_i^2 = (E_{ih}^2 + C^2) \{1 - (D_{\text{av}} - 1) (\Delta E_i/E_{ih})\}^2 \quad i = 0, 1 \quad (1.63)$$

$$D_{\text{av}} = (\nu_A D_A + \nu_B D_B)/8 \quad (1.64)$$

where ν_A and ν_B are the valencies of the two atoms, $D_A = 1$ for rows up to and including silicon, $D_A = 1.25$ for the row containing germanium, and $D_A = 1.46$ for the row containing tin. Like the E_{ih} , the ΔE_i depend upon the nearest-neighbour distance in the usual way:

$$E_{ih} = E_{ih}(\text{Si}) \left(\frac{d}{d_{\text{Si}}}\right)^{-S} \quad \Delta E_i = \Delta E_i(\text{Si}) \left(\frac{d}{d_{\text{Si}}}\right)^{-S}. \quad (1.65)$$

Finally, the splitting of the conduction band at X in polar compounds is taken to be simply related to the polar component of energy C as follows (Fig. 1.14):

$$\Delta_X = 0.071 C \text{ eV}. \quad (1.66)$$

The lower minima X_1 correspond to Bloch functions with anti-nodes at the anion site. Since the anion is more electronegative (Table 1.4) it is more attractive for electrons. The higher minima X_3 correspond to Bloch functions with anti-nodes at the cation.

The foregoing equations provide a reasonably successful basis for understanding the sources of variation of band structure and for predicting roughly the principal energies. Table 1.5 gives the various quantities which appear, and these are plotted in Fig. 1.15. Table 1.6 gives the observed and calculated parameters for semiconductors with diamond or sphalerite structure.

TABLE 1.4. Phillips' electronegativities.

I	II	III	IV	V	VI	VII
Li 1.00	Be 1.50	B 2.00	C 2.50	N 3.00	O 3.50	F 4.00
Na 0.72	Mg 0.95	Al 1.18	Si 1.41	P, 1.64	S 1.87	Cl 2.10
Cu 0.79	Zn 0.91	Ga 1.13	Ge 1.35	As 1.57	Se 1.79	Br 2.01
Ag 0.57	Cd 0.83	In 0.99	Sn 1.15	Sb 1.31	Te 1.47	I 1.63
Au 0.64	Hg 0.79	Tl 0.94	Pb 1.09	Bi 1.24		

From Phillips 1973.

Figure 1.16 shows the homopolar energy gaps between the top of the valence band and each of the conduction-band valleys. These are distinguished notationally from the gaps described above by the subscript *g* (although $E_g = E_0$). These gaps apply directly only to silicon, germanium, and α -tin. For silicon $D = 1$ and the smallest gap is the indirect one to X. For germanium $D = 1.25$ and d-shell mixing has reduced the gaps at Γ and L just enough to make the indirect gap to L the smallest. For α -tin $D = 1.46$ and the direct gap undercuts the gap at L. Figure 1.17 shows a similar set of curves for a polar material with $C = 3$. The polar interaction enhances the gaps appreciably even in the case of strong d-shell mixing.

TABLE 1.5. Basic energies for predicting chemical trends.

Parameter	Value for Si (eV)	Exponent <i>S</i>
I_h	5.17	1.308
E_{X4}	8.63	1.43
E_{0h}	4.10	2.75
E_{1h}	3.60	2.22
E_{2h}	4.50	2.382
E'_{0h}	3.40	1.92
E'_{1h}	5.90	1.67
ΔE_0	12.80	5.07
ΔE_1	4.98	4.97

From Van Vechten 1969.

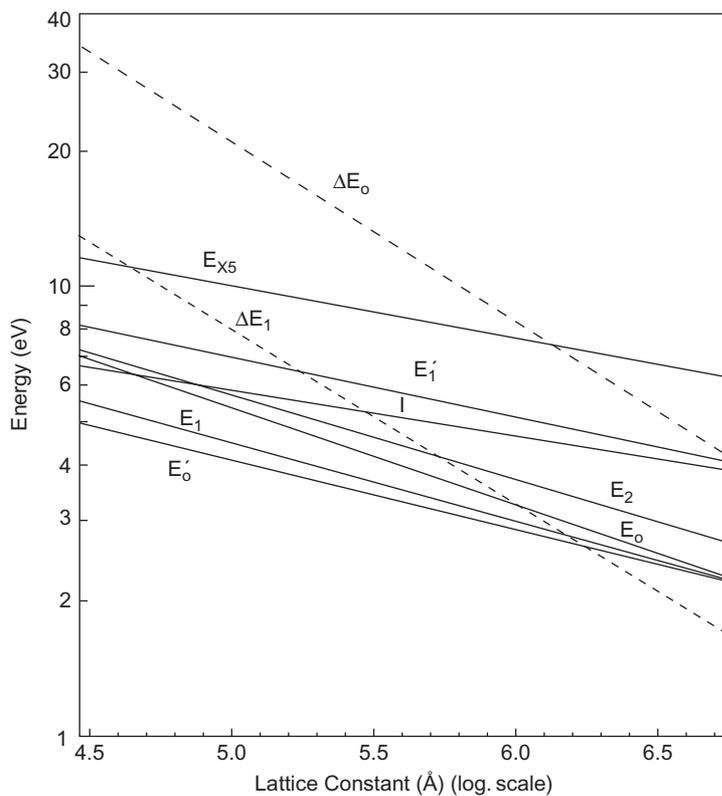
**FIG. 1.15.** Homopolar energies. (Data from Van Vechten 1969.)

TABLE 1.6. Band-structure parameters of diamond and sphalerite structures.

Semiconductor	$a_0(\text{\AA})$	$C(\text{eV})$	D	Band gaps without spin-orbit splitting						$\Delta_0(\text{eV})$
				$E_{g\Gamma}(\text{eV})$		$E_{gL}(\text{eV})$		$E_{gX}(\text{eV})$		
				Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	
C	3.567	0	1.00	13.04		5.77		5.48	5.48	0.006
Si	5.431	0	1.00	4.10	4.08	1.87		1.04	1.13	0.044
Ge	5.657	0	1.25	0.96	0.89	0.61	0.76	0.84	0.96	0.29
Sn	6.489	0	1.46	0.13	0.1	0.18	0.3	0.35		0.48
BP	4.538	0.68	1.00	6.76		2.88		1.81	2.0	
BA _s	4.777	0.38	1.11							
AlP	5.4625	3.14	1.00		4.6		3.7		2.4	0.06
AlAs	5.6611	2.67	1.11		3.05		2.9		2.26	0.29
AlSb	6.1355	3.10	1.19	2.67	2.5	2.39	2.0	2.15	1.87	0.75
GaP	5.4495	3.30	1.11	2.85	2.77	2.75	2.5	3.05	2.38	0.127
GaAs	5.6419	2.90	1.235	1.55	1.55	1.89	1.86	2.37	2.03	0.34
GaSb	6.094	2.10	1.325	1.00	0.99	1.17	1.07	1.36	1.30	0.80
InP	5.868	3.34	1.19	1.45	1.37	2.25	2.0	2.93	2.1	0.11
InAs	6.058	2.74	1.325	0.56	0.5	1.45	1.77	2.14	2.1	0.38
InSb	6.478	2.10	1.425	0.39	0.5	1.01	0.91	1.40	1.47	0.82
ZnS	5.4093	6.20	1.08	4.37	3.80	5.72		6.96		0.07
ZnSe	5.6676	5.60	1.175	3.37	2.9	4.26		5.82		0.43
ZnTe	6.101	4.48	1.235	2.72	2.56	3.64		4.26		0.93
CdTe	6.477	4.90	1.303	1.89	1.80	3.40		4.32		0.92

Spin-orbit splitting is readily incorporated. Following the discussion in Section 1.8 we merely subtract $\Delta_0/3$ from the ‘unprimed’ gaps (E_0, E_1, I) at Γ and L, and $2\Delta_0/3$ from the ‘primed’ gaps (E'_0, E'_1).

1.11. $\mathbf{k} \cdot \mathbf{p}$ perturbation and effective mass

In practice the important regions of the band structure are those which are most commonly populated by the mobile excitations of the crystal: electrons in the lowest conduction-band valley and holes at the top of the valence band. If we know the solution of the one-electron Schrödinger equation at these points in the Brillouin zone it is possible to obtain solutions in the immediate neighbourhood by regarding the scalar product $\mathbf{k} \cdot \mathbf{p}$, where \mathbf{k} is the wavevector measured from the Brillouin zone point, as a perturbation.

Suppose that the eigenvalues and Bloch functions are known for all bands at the centre of the zone (Γ). The Schrödinger equation

$$\left\{ \frac{\mathbf{p}^2}{2m} + \sum_l V(\mathbf{r} - \mathbf{R}_{l0}) \right\} u_{n\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) = E_{n\mathbf{k}} u_{n\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (1.67)$$

can be transformed to an equation containing only the periodic part of the Bloch function by using $\mathbf{p} = -i\hbar\nabla$:

$$\left\{ \frac{1}{2m} + (\mathbf{p} + \hbar\mathbf{k})^2 + V(\mathbf{r}) \right\} u_{n\mathbf{k}} = E_{n\mathbf{k}} u_{n\mathbf{k}} \quad (1.68)$$

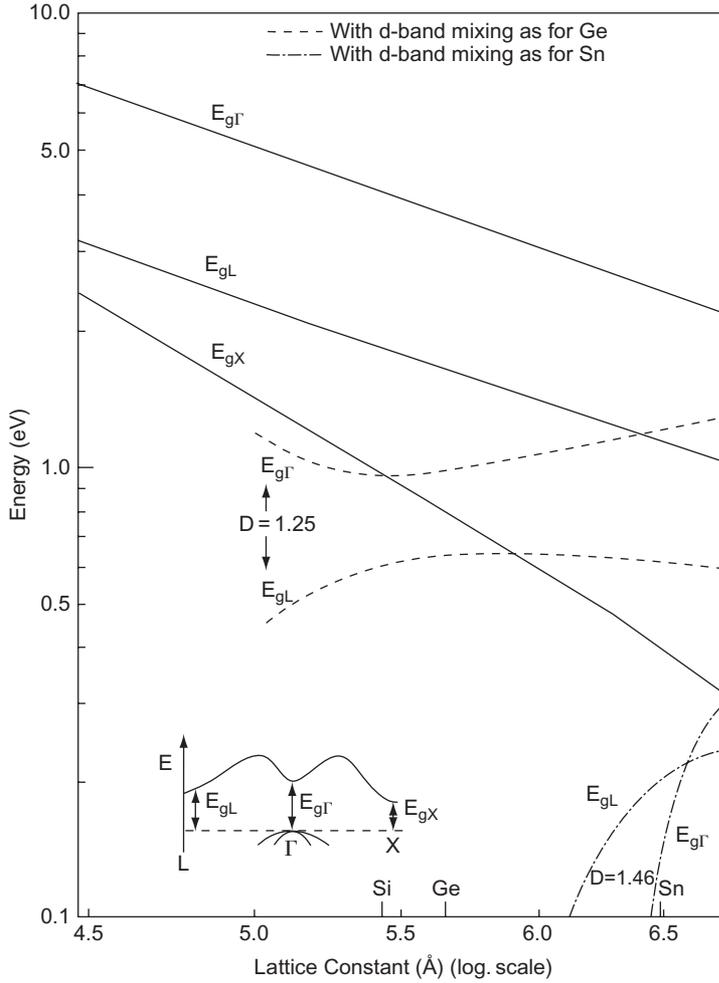


FIG. 1.16. Homopolar energy gaps: broken curves, with d-band mixing as for germanium; chain curves, with d-band mixing as for tin.

where $V(\mathbf{r})$ is merely a simpler notation for the periodic potential. Provided that $\hbar\mathbf{k} \ll \mathbf{p}$ we can write eqn (1.68) as

$$(H_0 + H_1 + H_2)u_{n\mathbf{k}} = E_{n\mathbf{k}}u_{n\mathbf{k}} \quad (1.69)$$

$$H_1 = \frac{\hbar}{m} \mathbf{k} \cdot \mathbf{p} \quad (1.70)$$

$$H_2 = \frac{\hbar^2 \mathbf{k}^2}{2m} \quad (1.71)$$

and regard H_1 as a first-order and H_2 as a second-order perturbation.

To zero order

$$\begin{aligned} u_{n\mathbf{k}} &= u_{n0} \\ E_{n\mathbf{k}} &= E_{n0}. \end{aligned} \quad (1.72)$$

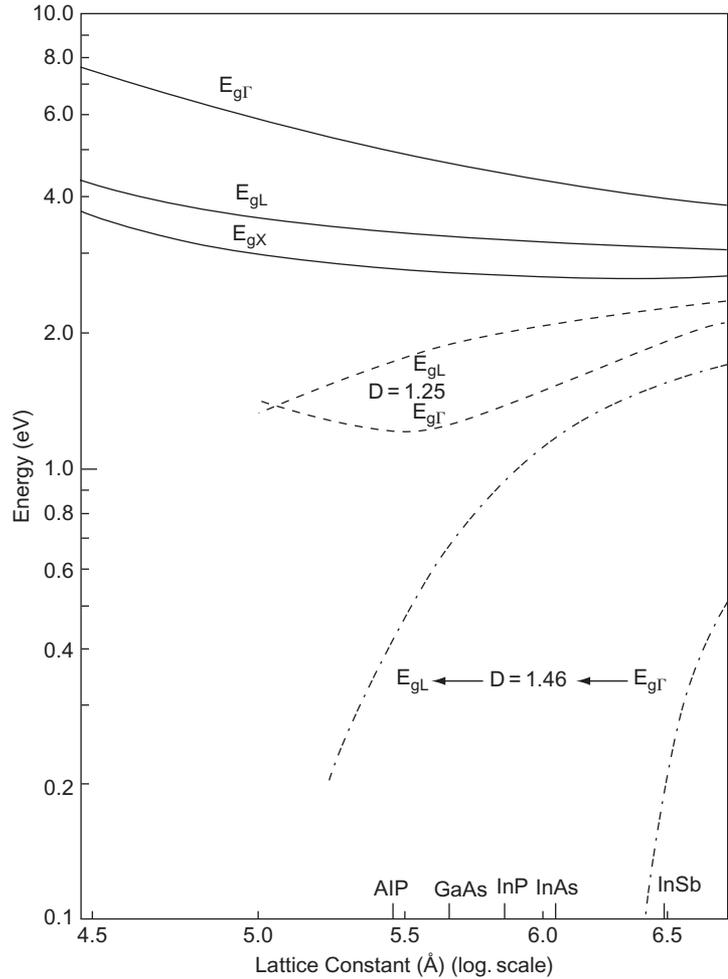


Fig. 1.17.
Energy gaps in polar material ($C = 3$).

To first order

$$u_{n\mathbf{k}} = u_{n0} + \frac{\hbar}{m} \sum_{m \neq n} \frac{\mathbf{k} \cdot \langle m0 | \mathbf{p} | n0 \rangle u_{m0}}{E_{n0} - E_{m0}} \quad (1.73)$$

$$E_{n\mathbf{k}} = E_{n0} + \frac{\hbar}{m} \mathbf{k} \cdot \langle n0 | \mathbf{p} | n0 \rangle.$$

Crystals with inversion symmetry, like silicon and germanium, have similarly symmetrical Bloch functions. Since \mathbf{p} is antisymmetric there will be no first-order correction to the energy. Where inversion symmetry is lacking, as in the sphalerite and wurtzite structures, the correction may be non-zero and a term proportional to \mathbf{k} appears. This indeed occurs at the top of the valence band and at the X point in the conduction band in these materials, and results in a shift of the extremum in \mathbf{k} (Fig. 1.18).

However, there is always a first-order correction to the Bloch function. For example, the Bloch functions of the conduction band and

valence band are respectively $|s\rangle$ like and $|p\rangle$ like and have a strong momentum matrix element connecting them. The second-order correction is not as urgent for the Bloch function as it is for the energy.

To second-order the energy becomes, neglecting the first-order correction,

$$E_{n\mathbf{k}} = E_{n0} + \frac{\hbar^2 \mathbf{k}^2}{2m} + \frac{\hbar^2}{m^2} \sum_{m \neq n} \frac{|\mathbf{k} \cdot \langle m0 | \mathbf{p} | n0 \rangle|^2}{E_{n0} - E_{m0}}. \quad (1.74)$$

We can express this result in terms of an effective mass m^* as follows:

$$E_{n\mathbf{k}} = E_{n0} + \sum_{i,j} \frac{\hbar^2}{2m_{ij}^*} k_i k_j \quad (1.75)$$

where

$$\frac{m}{m_{ij}^*} = \delta_{ij} + \frac{2}{m} \sum_{m \neq n} \frac{\langle n0 | p_i | m0 \rangle \langle m0 | p_j | n0 \rangle}{E_{n0} - E_{m0}} \quad (1.76)$$

and clearly m^* is a second-order tensor. Once again we have an example in eqn (1.76) of the repulsion of bands. A narrow gap between the conduction and valence bands produces small effective masses.

Degenerate perturbation theory is necessary for the valence band. The result near $\mathbf{k} = 0$ at the Γ point, considering only the conduction and valence bands, and neglecting terms linear in \mathbf{k} is (Kane 1957)

$$E_{c\mathbf{k}} = E_{g\Gamma} + \frac{\hbar^2 k^2}{2m_c^*} \quad (1.77)$$

$$\frac{m}{m_c^*} = 1 + \frac{2p_{cv}^2}{m} \frac{1}{3} \left(\frac{2}{E_{g\Gamma}} + \frac{1}{E_{g\Gamma} + \Delta_0} \right)$$

heavy hole

$$E_{v_1\mathbf{k}} = \frac{\hbar^2 k^2}{2m_h^*} \quad \frac{m}{m_h^*} = 1 \quad (1.78)$$

light hole

$$E_{v_2\mathbf{k}} = -\frac{\hbar^2 k^2}{2m_1^*} \quad \frac{m}{m_1^*} = \frac{4p_{cv}^2}{3mE_{g\Gamma}} - 1 \quad (1.79)$$

split-off

$$E_{v_3\mathbf{k}} = -\Delta_0 - \frac{\hbar^2 k^2}{2m_{so}^*} \quad (1.80)$$

$$\frac{m}{m_{so}^*} = \frac{2p_{cv}^2}{3m(E_{g\Gamma} + \Delta_0)} - 1$$

where p_{cv} is the momentum matrix element between the conduction band $|s\rangle$ and the valence band $|p\rangle$. For germanium $2p_{cv}^2/m = 22.5$ eV and for GaAs $2p_{cv}^2/m = 21.5$ eV (Phillips 1973). When the effect of

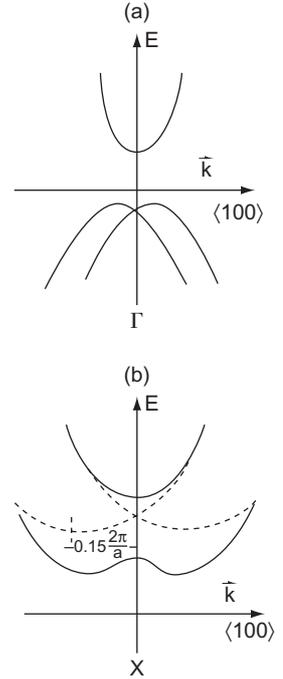
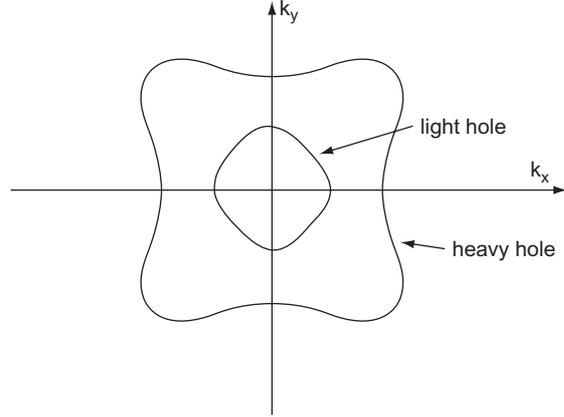


FIG. 1.18. Shifts in extrema caused by lack of inversion symmetry: (a) top of valence band; (b) camel's back structure in the conduction band near X (the broken curves are for silicon).

**Fig. 1.19.**

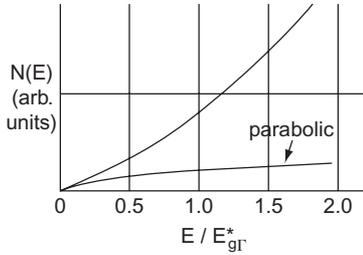
Constant-energy surfaces for valence bands in the (x, y) plane.

more remote bands is included, the effective mass of the heavy-hole band becomes negative, as in the case of the other two. The two valence bands, which are degenerate at $\mathbf{k} = 0$, have the form of warped spheres (Fig. 1.19):

$$E_{v_{1,2}\mathbf{k}} = -\frac{\hbar^2}{2m} \left\{ Ak^2 \pm \left(B^2 k^4 + C^2 \left[k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2 \right] \right)^{1/2} \right\}. \quad (1.81)$$

For silicon $A = 4.0$, $B = 1.1$, and $C = 4.1$ which corresponds to a heavy-hole mass of about $0.49m$ and a light-hole mass of about $0.16m$. For germanium $A = 13.1$, $B = 8.3$, and $C = 12.5$ which corresponds to a heavy-hole mass of about $0.28m$ and a light-hole mass of about $0.044m$.

The split-off band is spherical, as is the $|s\rangle$ -like conduction band at Γ . However, away from $\mathbf{k} = 0$ the bands become non-parabolic and take the form (Fig. 1.20) (for Δ_0 not too close to E_g)

**Fig. 1.20.**

The non-parabolicity of the conduction band of InSb. $N(E)$ is the density of states and $E_{g\Gamma}^*$ is the effective energy gap equal to 0.23 eV at 0 K. (From Ehrenreich 1957.)

$$\frac{\hbar^2 k^2}{2m^*} = \gamma(E_{\mathbf{k}}) \approx E_{\mathbf{k}} \left(1 + \alpha E_{\mathbf{k}} + \beta E_{\mathbf{k}}^2 \right) \quad (1.82)$$

$$\alpha = \frac{1}{E_{g\Gamma}} \left(1 - \frac{m^*}{m} \right)^2 \quad (1.83)$$

$$\beta = -\frac{2}{E_{g\Gamma}^2} \frac{m^*}{m} \left(1 - \frac{m^*}{m} \right)$$

The conduction-band valleys at L and X are prolate spheroidal, each with the general form near the extremum

$$\frac{\hbar^2 k_{x'}^2}{2m_1^*} + \frac{\hbar^2 (k_{y'}^2 + k_{z'}^2)}{2m_t^*} \approx E_{\mathbf{k}} \quad (1.84)$$

where \mathbf{k} is measured from the minimum, x' is along the principal direction of the prolate spheroid, and y' and z' are at right-angles. Silicon has six equivalent valleys at points along the $\langle 100 \rangle$ directions distant from the X points by $0.15(2\pi/a)$. For these $m_1^* = 0.98m$ and $m_t^* = 0.19m$.