

THE PHYSICS OF SOLAR CELLS

Jenny Nelson

Imperial College Press

HANWHA 1018

THE PHYSICS OF SOLAR CELLS



THE PHYSICS OF SOLAR CELLS

Jenny Nelson
Imperial College, UK



ICP
Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

2005271257

THE PHYSICS OF SOLAR CELLS

Copyright © 2003 by Imperial College Press

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 1-86094-340-3
ISBN 1-86094-349-7 (pbk)

Printed in Singapore by Mainland Press

LIBRARY OF CONGRESS
DEC 0 6 2004
COPY
COPYRIGHT OFFICE

Preface

Over the last ten years, photovoltaics has emerged to become an application of recognised potential and has attracted the interest of increasing numbers of students and researchers. The purpose of this book is to provide an introduction to, and overview of, the physics of the photovoltaic cell. It should be suitable for undergraduate physicists and engineers who are interested in this application of semiconductor physics, and to non-specialist graduates and others who require a background in the physical principles of solar cells. The focus is on the basic semiconductor physics relevant to photovoltaics, physical models of photovoltaic devices and how these relate to the design and function of practical devices. It should enable the reader to understand how solar cells work, to understand the concepts and models of solar cell device physics, and to formulate and solve relevant physical problems. Although practical materials and device designs are used as examples, the book is not intended as a comprehensive review of photovoltaic materials and devices, nor of the latest developments in photovoltaics research.

Chapter 1 introduces the solar cell as a simple current generator and defines the performance characteristics which are used to describe and compare solar cells. Chapter 2 describes in general terms how light energy is converted into electricity, comparing the photo*voltaic* converter with other systems and evaluating the limits to efficiency. Chapters 3 and 4 cover the basic physics of the semiconductor, the semiconductor transport equations and the processes of light absorption and carrier recombination. Chapter 5 focuses on the concept of the asymmetric junction, and details the different types of junction which are exploited in photovoltaics. Chapter 6 applies the theory of earlier chapters to a p–n junction, the classical model of a solar cell. Chapters 7 and 8 are concerned with the range of photovoltaic materials and device designs. Chapter 7 deals with monocrystalline p–n junction devices, relating the model of Chapter 6 to practical devices,

using crystalline silicon and gallium arsenide cells as examples. Chapter 8 deals with thin film photovoltaic materials, discussing physical processes and design issues relevant to thin films and focusing on the ways in which the standard model must be adapted for thin film devices. Chapter 9 deals with various techniques for managing light in order to maximise performance, and Chapter 10 covers a range of approaches, mainly theoretical, to increasing the efficiency of solar cells above the limit for a single band gap photoconverter.

I am grateful to all of the people who have helped me prepare this book. In particular, to Keith Barnham for passing the original proposal from Imperial College Press in my direction; to Leon Freris and David Infield for giving me the opportunity to teach the physics of solar cells to MSc students at Loughborough, and so establish the basic course from which this book developed; to all the research students in photovoltaics at Imperial College for raising so many interesting questions, especially Jenny Barnes, James Connolly and Benjamin Kluftinger; to Ralph Gottshalg, Tom Markvart and Peter Wuerfel for help with questions related to material in this book; to Ned Ekins-Daukes and Jane Nelson for their helpful comments on the text; to Clare Nelson for the cover illustration and to all other colleagues who have helped in my endeavours to understand how these things work, in particular to Richard Corkish, James Durrant, Michael Grätzel, Martin Green, Christiana Honsberg, Stefan Kettemann and Ellen Moons. I am grateful to the Greenpeace Environmental Trust for funding me to study solar cells before they were popular, and to the UK Engineering and Physical Sciences Research Council and for an Advanced Research Fellowship which allowed me to spend my Saturday afternoons writing chapters instead of lectures. Finally I am grateful to John Navas for his encouragement to start on this project and to Laurent Chaminade and his staff at IC Press and to Lakshmi Narayan and colleagues at World Scientific, for their help in seeing it through.

This book is dedicated to the memory of Stephen Robinson and M.V. McCaughan.

Jenny Nelson
London, April 2002

Contents

Preface v

Chapter 1 Introduction 1
1.1. Photons In, Electrons Out: The Photovoltaic Effect 1
1.2. Brief History of the Solar Cell 2
1.3. Photovoltaic Cells and Power Generation 4
1.3.1. Photovoltaic cells, modules and systems 4
1.3.2. Some important definitions 6
1.4. Characteristics of the Photovoltaic Cell: A Summary 7
1.4.1. Photocurrent and quantum efficiency 7
1.4.2. Dark current and open circuit voltage 9
1.4.3. Efficiency 11
1.4.4. Parasitic resistances 13
1.4.5. Non-ideal diode behaviour 15
1.5. Summary 15
References 16

Chapter 2 Photons In, Electrons Out: Basic Principles of PV 17
2.1. Introduction 17
2.2. The Solar Resource 17
2.3. Types of Solar Energy Converter 22
2.4. Detailed Balance 24
2.4.1. In equilibrium 24
2.4.2. Under illumination 26
2.5. Work Available from a Photovoltaic Device 28
2.5.1. Photocurrent 28

2.5.2. Dark current . . . 30
2.5.3. Limiting efficiency . . . 31
2.5.4. Effect of band gap . . . 33
2.5.5. Effect of spectrum on efficiency . . . 34
2.6. Requirements for the Ideal Photoconverter . . . 35
2.7. Summary . . . 38
References . . . 39

Chapter 3 Electrons and Holes in Semiconductors 41
3.1. Introduction . . . 41
3.2. Basic Concepts . . . 42
3.2.1. Bonds and bands in crystals . . . 42
3.2.2. Electrons, holes and conductivity . . . 44
3.3. Electron States in Semiconductors . . . 46
3.3.1. Band structure . . . 46
3.3.2. Conduction band . . . 48
3.3.3. Valence band . . . 49
3.3.4. Direct and indirect band gaps . . . 50
3.3.5. Density of states . . . 51
3.3.6. Electron distribution function . . . 54
3.3.7. Electron and hole currents . . . 55
3.4. Semiconductor in Equilibrium . . . 56
3.4.1. Fermi Dirac statistics . . . 56
3.4.2. Electron and hole densities in equilibrium . . . 57
3.4.3. Boltzmann approximation . . . 58
3.4.4. Electron and hole currents in equilibrium . . . 60
3.5. Impurities and Doping . . . 61
3.5.1. Intrinsic semiconductors . . . 61
3.5.2. n type doping . . . 62
3.5.3. p type doping . . . 63
3.5.4. Effects of heavy doping . . . 65
3.5.5. Imperfect and amorphous crystals . . . 65
3.6. Semiconductor under Bias . . . 66
3.6.1. Quasi thermal equilibrium . . . 66
3.6.2. Electron and hole densities under bias . . . 68
3.6.3. Current densities under bias . . . 69
3.7. Drift and Diffusion . . . 72
3.7.1. Current equations in terms of drift and diffusion . . . 72
3.7.2. Validity of the drift-diffusion equations . . . 75

3.7.3. Current equations for non-crystalline solids 76
3.8. Summary 77

Chapter 4 Generation and Recombination **79**
4.1. Introduction: Semiconductor Transport Equations 79
4.2. Generation and Recombination 81
4.3. Quantum Mechanical Description of Transition Rates 83
4.3.1. Fermi's Golden Rule 83
4.3.2. Optical processes in a two level system 85
4.4. Photogeneration 87
4.4.1. Photogeneration rate 88
4.4.2. Thermalisation 89
4.4.3. Microscopic description of absorption 90
4.4.4. Direct gap semiconductors 93
4.4.5. Indirect gap semiconductors 94
4.4.6. Other types of behaviour 96
4.4.7. Examples and data 98
4.5. Recombination 99
4.5.1. Types of recombination 99
4.5.2. Radiative recombination 99
4.5.3. Simplified expressions for radiative recombination . . 102
4.5.4. Auger recombination 105
4.5.5. Shockley Read Hall recombination 106
4.5.6. Surface and grain boundary recombination 110
4.5.7. Traps versus recombination centres 111
4.6. Formulation of the Transport Problem 112
4.6.1. Comments on the transport problem 113
4.6.2. Transport equations in a crystal 114
4.7. Summary 115
References 117

Chapter 5 Junctions **119**
5.1. Introduction 119
5.2. Origin of Photovoltaic Action 120
5.3. Work Function and Types of Junction 124
5.4. Metal–Semiconductor Junction 125
5.4.1. Establishing a field 125
5.4.2. Behaviour in the light 126
5.4.3. Behaviour in the dark 127

5.4.4. Ohmic contacts 129
5.4.5. Limitations of the Schottky barrier junction 130
5.5. Semiconductor–Semiconductor Junctions 131
5.5.1. p–n junction 131
5.5.2. p–i–n junction 132
5.5.3. p–n heterojunction 133
5.6. Electrochemical Junction 133
5.7. Junctions in Organic Materials 137
5.8. Surface and Interface States 139
5.8.1. Surface states on free surfaces 139
5.8.2. Effect of interface states on junctions 141
5.9. Summary 143
References 144

Chapter 6 Analysis of the p–n Junction **145**
6.1. Introduction 145
6.2. The p–n Junction 146
6.2.1. Formation of p–n junction 146
6.2.2. Outline of approach 147
6.3. Depletion Approximation 149
6.3.1. Calculation of depletion width 150
6.4. Calculation of Carrier and Current Densities 152
6.4.1. Currents and carrier densities in the neutral regions 152
6.4.2. Currents and carrier densities in the space charge region 154
6.4.3. Total current density 156
6.5. General Solution for $J(V)$ 156
6.6. p–n Junction in the Dark 160
6.6.1. At equilibrium 160
6.6.2. Under applied bias 160
6.7. p–n Junction under Illumination 165
6.7.1. Short circuit 165
6.7.2. Photocurrent and QE in special cases 167
6.7.3. p–n junction as a photovoltaic cell 169
6.8. Effects on p–n Junction Characteristics 172
6.8.1. Effects of parasitic resistances 172
6.8.2. Effect of irradiation 172
6.8.3. Effect of temperature 173

6.8.4. Other device structures 174
6.8.5. Validity of the approximations 174
6.9. Summary 175
References 176

Chapter 7 Monocrystalline Solar Cells 177
7.1. Introduction: Principles of Cell Design 177
7.2. Material and Design Issues 178
7.2.1. Material dependent factors 178
7.2.2. Design factors 179
7.2.3. General design features of p–n junction cells 180
7.3. Silicon Material Properties 180
7.3.1. Band structure and optical absorption 180
7.3.2. Doping 181
7.3.3. Recombination 182
7.3.4. Carrier transport 185
7.4. Silicon Solar Cell Design 186
7.4.1. Basic silicon solar cell 186
7.4.2. Cell fabrication 186
7.4.3. Optimisation of silicon solar cell design 188
7.4.4. Strategies to enhance absorption 190
7.4.5. Strategies to reduce surface recombination 191
7.4.6. Strategies to reduce series resistance 194
7.4.7. Evolution of silicon solar cell design 194
7.4.8. Future directions in silicon cell design 197
7.4.9. Alternatives to silicon 198
7.5. III–V Semiconductor Material Properties 198
7.5.1. III–V semiconductor band structure and optical absorption 198
7.5.2. Gallium arsenide 200
7.5.3. Doping 201
7.5.4. Recombination 202
7.5.5. Carrier transport 203
7.5.6. Reflectivity 203
7.6. GaAs Solar Cell Design 204
7.6.1. Basic GaAs solar cell 204
7.6.2. Optimisation of GaAs solar cell design 204
7.6.3. Strategies to reduce front surface recombination . . . 205
7.6.4. Strategies to reduce series resistance 207

7.6.5. Strategies to reduce substrate cost 208
7.7. Summary 208
References 210

Chapter 8 Thin Film Solar Cells 211
8.1. Introduction 211
8.2. Thin Film Photovoltaic Materials 213
8.2.1. Requirements for suitable materials 213
8.3. Amorphous Silicon 213
8.3.1. Materials properties 213
8.3.2. Defects in amorphous material 215
8.3.3. Absorption 217
8.3.4. Doping 217
8.3.5. Transport 219
8.3.6. Stability 220
8.3.7. Related alloys 221
8.4. Amorphous Silicon Solar Cell Design 221
8.4.1. Amorphous silicon p–i–n structures 221
8.4.2. p–i–n solar cell device physics 222
8.4.3. Fabrication of a-Si solar cells 227
8.4.4. Strategies to improve a-Si cell performance 227
8.5. Defects in Polycrystalline Thin Film Materials 229
8.5.1. Grain boundaries 230
8.5.2. Effects of grain boundaries on transport 233
8.5.3. Depletion approximation model for grain boundary 234
8.5.4. Majority carrier transport 236
8.5.5. Effect of illumination 239
8.5.6. Minority carrier transport 240
8.5.7. Effects of grain boundary recombination on solar cell performance 242
8.6. $CuInSe_2$ Thin Film Solar Cells 243
8.6.1. Materials properties 243
8.6.2. Heterojunctions in thin film solar cell design 244
8.6.3. $CuInGaSe_2$ solar cell design 245
8.7. CdTe Thin Film Solar Cells 246
8.7.1. Materials properties 246
8.7.2. CdTe solar cell design 247
8.8. Thin Film Silicon Solar Cells 248

8.8.1. Materials properties 248
8.8.2. Microcrystalline silicon solar cell design 248
8.9. Summary 249
References 251

Chapter 9 Managing Light **253**
9.1. Introduction 253
9.2. Photon Flux: A Review and Overview of Light Management 255
9.2.1. Routes to higher photon flux 257
9.3. Minimising Reflection 258
9.3.1. Optical properties of semiconductors 258
9.3.2. Antireflection coatings 260
9.4. Concentration 263
9.4.1. Limits to concentration 263
9.4.2. Practical concentrators 264
9.5. Effects of Concentration on Device Physics 266
9.5.1. Low injection 266
9.5.2. High injection 267
9.5.3. Limits to efficiency under concentration 269
9.5.4. Temperature 270
9.5.5. Series resistance 270
9.5.6. Concentrator cell design 270
9.5.7. Concentrator cell materials 271
9.6. Light Confinement 272
9.6.1. Light paths and ray tracing 272
9.6.2. Mirrors 274
9.6.3. Randomising surfaces 275
9.6.4. Textured surfaces 276
9.6.5. Practical schemes 278
9.6.6. Light confining structures: restricted acceptance areas and external cavities 280
9.6.7. Effects of light trapping on device physics 281
9.7. Photon Recycling 282
9.7.1. Theory of photon recycling 282
9.7.2. Practical schemes 285
9.8. Summary 286
References 288

Chapter 10 Over the Limit: Strategies for High Efficiency 289
10.1. Introduction 289
10.2. How Much is Out There? Thermodynamic Limits to Efficiency 291
10.3. Detailed Balance Limit to Efficiency, Reviewed 292
10.4. Multiple Band Gaps 297
10.5. Tandem Cells 298
10.5.1. Principles of tandem cells 298
10.5.2. Analysis 300
10.5.3. Practical tandem systems 301
10.6. Intermediate Band and Multiple Band Cells 302
10.6.1. Principles of intermediate and multiple band cells . . 302
10.6.2. Conditions 303
10.6.3. Practical strategies 306
10.7. Increasing the Work Per Photon using 'Hot' Carriers 309
10.7.1. Principles of cooling and 'hot' carriers 309
10.7.2. Analysis of the hot carrier solar cell 311
10.7.3. Practical strategies 316
10.8. Impact Ionisation Solar Cells 318
10.8.1. Analysis of impact ionisation solar cell 320
10.9. Summary 323
References 324

Exercises 327

Solutions to the Exercises 337

Index 355

Fundamental constants

h	Planck's constant
$\hbar$	Planck's constant/2π
ε_0	dielectric permittivity of free space
σ_{B}	Stefan's constant
c	speed of light in vacuum
k_{B}	Boltzmann's constant
m_0	free electron mass
q	charge on the electron

Symbols used in the text

α	absorption coefficient
β	spectral photon flux density per unit solid angle
χ	electron affinity
$\Delta\mu$	quasi Fermi level separation or chemical potential of light
ε	emissivity *i.e.* probability of photon emission
ε_{s}	dielectric permittivity of semiconductor
$\boldsymbol{\epsilon}$	polarisation vector of light
ϕ	electrostatic potential
Φ	work function
ϕ_0	neutrality level
η	power conversion efficiency
κ_{s}	imaginary part of refractive index of semiconductor
λ	wavelength of light
μ	chemical potential
μ_{n}; μ_{p}	electron mobility; hole mobility
ν	frequency of light
θ_{c}	critical angle at optical interface
θ_{sun}	angular width of the sun
ρ	charge density or resistivity
σ	conductivity
τ	lifetime
τ_{n}; τ_{p}	electron lifetime; hole lifetime
ω	angular frequency of light
Ω	solid angle
ψ	wavefunction
∇, ∇_{r}	grad operator with respect to position
∇_{k}	grad operator with respect to wavevector

a	probability of photon absorption
A	cell area
b	spectral photon flux density normal to surface
B_{n}; B_{p}	coefficient for bimolecular capture by trap of electrons; holes
B_{rad}	coefficient of bimolecular radiative recombination
D	diffusion coefficient
D_{n}; D_{p}	diffusion coefficient of electron; diffusion coefficient of hole
E	energy
E_0	electromagnetic field strength
E_{F}	Fermi energy or Fermi level
$E_{\mathrm{F_n}}$; $E_{\mathrm{F_p}}$	electron quasi-Fermi energy level; hole quasi-Fermi level
E_{g}	band gap
E_{t}	energy of trap state
E_{v}	energy of valence band edge
E_{c}	energy of conduction band edge
E_{i}	intrinsic energy level
E_{vac}	vacuum energy level
F	electrostatic field
$\mathbf{F}$	force
$f(\mathbf{k}, \mathbf{r})$	probability occupation function for electronic state at $\mathbf{k}$, $\mathbf{r}$
f_0	Fermi Dirac probability occupation function
F_{a}; F_{e}; F_{s}; F_{X}	geometrical factor relating normal to angular photon flux density for emission from: ambient; cell; sun; concentrated light source
FF	fill factor
g	spectral photogeneration rate per unit volume
G	generation rate per unit volume
G_{n}	electron generation rate per unit volume
G_{p}	hole generation rate per unit volume
$g(E)$	density of electronic states per unit energy per unit crystal volume
$\mathbf{H}$	Hamiltonian operator
I	current
I_0	incident light intensity
J	current density
J_{sc}	short circuit current density
J_{dark}	dark current density
J_{n}; J_{p}	electron current density; hole current density

j_n; j_p	spectral electron current density; spectral hole current density
$\mathbf{k}$	crystal wavevector
L	diffusion length or length of crystal sample
L_n; L_p	diffusion length of electron; diffusion length of hole
m	diode ideality factor
m_c^*	conduction band effective mass
m_v^*	valence band effective mass
$\mathbf{M}$	dipole matrix element
n	density of electrons per unit volume
n_0	equilibrium electron density
n_i	intrinsic carrier density
n_s	refractive index of semiconductor
N_a	density of acceptor impurity atoms
N_d	density of donor impurity atoms
N_c	effective conduction band density of states
N_v	effective valence band density of states
N_i	charged background doping in intrinsic layer
N_t	density of trap states
N_s	density of interface states per unit area
p	density of holes per unit volume
$\mathbf{p}$	momemtum
P	power density
p_0	equilibrium hole density
Q	charge
QE	quantum efficiency
$\mathbf{r}$	position
r	transition rate
R	reflectivity
R_s	series resistance
R_{sh}	shunt or parallel resistance
$\mathbf{s}$	vector defining a point on surface
S_n	electron surface recombination velocity
S_p	hole surface recombination velocity
t	time
T	temperature
T_s	temperature of sun
U_E	energy density of radiation per unit volume
U	recombination rate per unit volume
U_n	electron recombination rate per unit volume

U_{p}	hole recombination rate per unit volume
U_{Aug}	Auger recombination rate
U_{rad}	radiative recombination rate
U_{SRH}	Shockley Read Hall recombination rate
$\mathcal{V}$	volume
v	velocity
V	voltage or bias
V_{bi}	built in bias
V_{oc}	open circuit voltage
V_{n}	donor ionisation energy
V_{p}	acceptor ionisation energy
$w_{\mathrm{n}}; w_{\mathrm{p}}$	thickness of depletion region in n layer; in p layer
X	concentration factor

Acronyms

e.m.f.	electromotive force
ac	alternating current
dc	direct current
AM	air mass
AM1.5	air mass 1.5 spectrum — standard for solar cell calibration
STC	standard test conditions
CB	conduction band
VB	valence band
FGR	Fermi's golden rule
PV	photovoltaics
a-Si	amorphous silicon
c-Si	crystalline silicon
μ-Si	microcrystalline silicon
a-Si:H	hydrogenated amorphous silicon
a-SiC	amorphous silicon-carbon alloy
a-SiGe	amorphous silicon-germanium alloy
DOS	density of states
JDOS	joint density of states
i region	intrinsic or undoped region of a p–i–n junction
SCR	space charge region
QE	quantum efficiency
SRH	Shockley–Read–Hall
GaAs	gallium arsenide
CdTe	cadmium telluride

CIGS	copper indium gallium diselenide
AR	antireflection
PERL	passivated emitter, rear locally diffused solar cell
PESC	passivated emitter solar cell
TCO	transparent conducting oxide
PR	photon recycling
QD	quantum dot
QW	quantum well
LO	longitudinal optical (of phonons)

Chapter 1

Introduction

1.1. Photons In, Electrons Out: The Photovoltaic Effect

Solar photovoltaic energy conversion is a one-step conversion process which generates electrical energy from light energy. The explanation relies on ideas from quantum theory. Light is made up of packets of energy, called *photons*, whose energy depends only upon the frequency, or colour, of the light. The energy of visible photons is sufficient to excite electrons, bound into solids, up to higher energy levels where they are more free to move. An extreme example of this is the photoelectric effect, the celebrated experiment which was explained by Einstein in 1905, where blue or ultraviolet light provides enough energy for electrons to escape completely from the surface of a metal. Normally, when light is absorbed by matter, photons are given up to excite electrons to higher energy states within the material, but the excited electrons quickly relax back to their ground state. In a photovoltaic device,

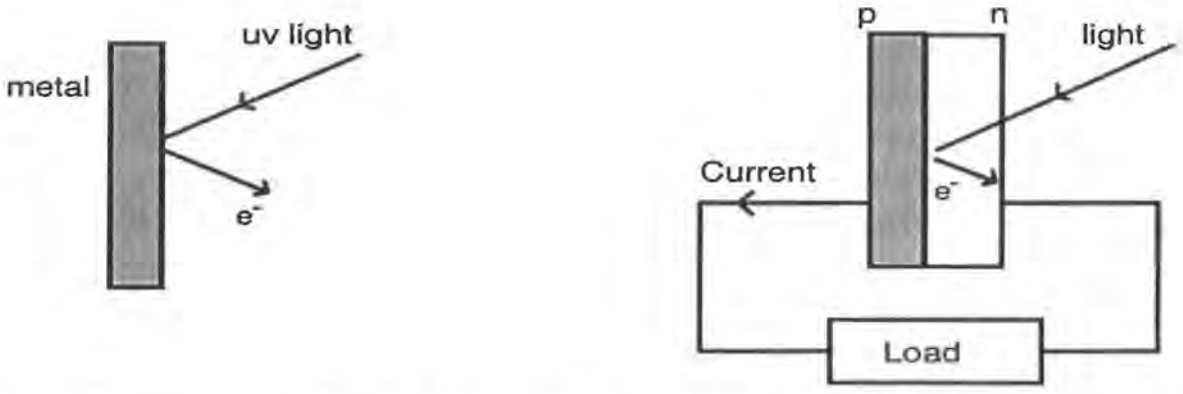


Fig. 1.1. Comparison of the photoelectric effect (left), where uv light liberates electrons from the surface of a metal, with the photovoltaic effect in a solar cell (right). The photovoltaic cell needs to have some spatial asymmetry, such as contacts with different electronic properties, to drive the excited electrons through the external circuit.

however, there is some built-in asymmetry which pulls the excited electrons away before they can relax, and feeds them to an external circuit. The extra energy of the excited electrons generates a *potential difference*, or electromotive force (e.m.f.). This force drives the electrons through a load in the external circuit to do electrical work.

The effectiveness of a photovoltaic device depends upon the choice of light absorbing materials and the way in which they are connected to the external circuit. The following chapters will deal with the underlying physical ideas, the device physics of solar cells, the properties of photovoltaic materials and solar cell design. In this chapter we will summarise the main characteristics of a photovoltaic cell without discussing its physical function in detail.

1.2. Brief History of the Solar Cell

The photovoltaic effect was first reported by Edmund Bequerel in 1839 when he observed that the action of light on a silver coated platinum electrode immersed in electrolyte produced an electric current. Forty years later the first solid state photovoltaic devices were constructed by workers investigating the recently discovered photoconductivity of selenium. In 1876 William Adams and Richard Day found that a photocurrent could be produced in a sample of selenium when contacted by two heated platinum contacts. The photovoltaic action of the selenium differed from its photoconductive action in that a current was produced spontaneously by the action of light. No external power supply was needed. In this early photovoltaic device, a rectifying junction had been formed between the semiconductor and the metal contact. In 1894, Charles Fritts prepared what was probably the first large area solar cell by pressing a layer of selenium between gold and another metal. In the following years photovoltaic effects were observed in copper–copper oxide thin film structures, in lead sulphide and thallium sulphide. These early cells were thin film Schottky barrier devices, where a semitransparent layer of metal deposited on top of the semiconductor provided both the asymmetric electronic junction, which is necessary for photovoltaic action, and access to the junction for the incident light. The photovoltaic effect of structures like this was related to the existence of a barrier to current flow at one of the semiconductor–metal interfaces (*i.e.*, rectifying action) by Goldman and Brodsky in 1914. Later, during the 1930s, the theory of metal–semiconductor barrier layers was developed by Walter Schottky, Neville Mott and others.

However, it was not the photovoltaic properties of materials like selenium which excited researchers, but the photoconductivity. The fact that the current produced was proportional to the intensity of the incident light, and related to the wavelength in a definite way meant that photoconductive materials were ideal for photographic light meters. The photovoltaic effect in barrier structures was an added benefit, meaning that the light meter could operate without a power supply. It was not until the 1950s, with the development of good quality silicon wafers for applications in the new solid state electronics, that potentially useful quantities of power were produced by photovoltaic devices in crystalline silicon.

In the 1950s, the development of silicon electronics followed the discovery of a way to manufacture p–n junctions in silicon. Naturally n type silicon wafers developed a p type skin when exposed to the gas boron trichloride. Part of the skin could be etched away to give access to the n type layer beneath. These p–n junction structures produced much better rectifying action than Schottky barriers, and better photovoltaic behaviour. The first silicon solar cell was reported by Chapin, Fuller and Pearson in 1954 and converted sunlight with an efficiency of 6%, six times higher than the best previous attempt. That figure was to rise significantly over the following years and decades but, at an estimated production cost of some $200 per Watt, these cells were not seriously considered for power generation for several decades. Nevertheless, the early silicon solar cell did introduce the possibility of power generation in remote locations where fuel could not easily be delivered. The obvious application was to satellites where the requirement of reliability and low weight made the cost of the cells unimportant and during the 1950s and 60s, silicon solar cells were widely developed for applications in space.

Also in 1954, a cadmium sulphide p–n junction was produced with an efficiency of 6%, and in the following years studies of p–n junction photovoltaic devices in gallium arsenide, indium phosphide and cadmium telluride were stimulated by theoretical work indicating that these materials would offer a higher efficiency. However, silicon remained and remains the foremost photovoltaic material, benefiting from the advances of silicon technology for the microelectronics industry. Short histories of the solar cell are given elsewhere [Shive, 1959; Wolf, 1972; Green, 1990].

In the 1970s the crisis in energy supply experienced by the oil-dependent western world led to a sudden growth of interest in alternative sources of energy, and funding for research and development in those areas. Photovoltaics was a subject of intense interest during this period, and a range of

strategies for producing photovoltaic devices and materials more cheaply and for improving device efficiency were explored. Routes to lower cost included photoelectrochemical junctions, and alternative materials such as polycrystalline silicon, amorphous silicon, other 'thin film' materials and organic conductors. Strategies for higher efficiency included tandem and other multiple band gap designs. Although none of these led to widespread commercial development, our understanding of the science of photovoltaics is mainly rooted in this period.

During the 1990s, interest in photovoltaics expanded, along with growing awareness of the need to secure sources of electricity alternative to fossil fuels. The trend coincides with the widespread deregulation of the electricity markets and growing recognition of the viability of decentralised power. During this period, the economics of photovoltaics improved primarily through economies of scale. In the late 1990s the photovoltaic production expanded at a rate of 15–25% per annum, driving a reduction in cost. Photovoltaics first became competitive in contexts where conventional electricity supply is most expensive, for instance, for remote low power applications such as navigation, telecommunications, and rural electrification and for enhancement of supply in grid-connected loads at peak use [Anderson, 2001]. As prices fall, new markets are opened up. An important example is building integrated photovoltaic applications, where the cost of the photovoltaic system is offset by the savings in building materials.

1.3. Photovoltaic Cells and Power Generation

1.3.1. *Photovoltaic cells, modules and systems*

The solar *cell* is the basic building block of solar photovoltaics. The cell can be considered as a two terminal device which conducts like a diode in the dark and generates a photovoltage when charged by the sun. Usually it is a thin slice of semiconductor material of around 100 cm^2 in area. The surface is treated to reflect as little visible light as possible and appears dark blue or black. A pattern of metal contacts is imprinted on the surface to make electrical contact (Fig. 1.2(a)).

When charged by the sun, this basic unit generates a dc photovoltage of 0.5 to 1 volt and, in short circuit, a photocurrent of some tens of milliamps per cm^2. Although the current is reasonable, the voltage is too small for most applications. To produce useful dc voltages, the cells are connected together in series and encapsulated into *modules*. A module typically contains

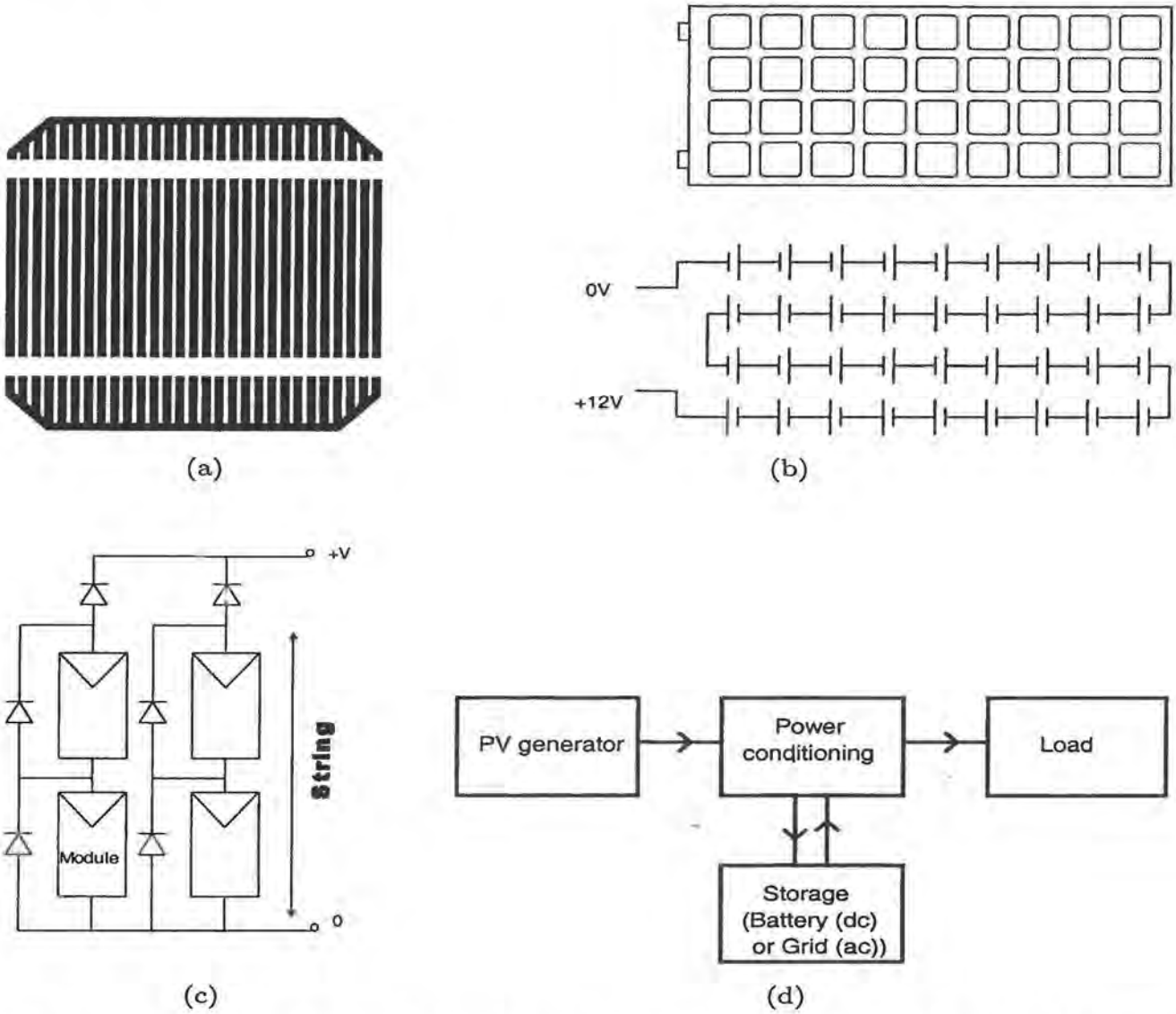


Fig. 1.2. (a) Photovoltaic cell showing surface contact patterns (b) In a module, cells are usually connected in series to give a standard dc voltage of 12 V (c) For any application, modules are connected in series into strings and then in parallel into an array, which produces sufficient current and voltage to meet the demand. (d) In most cases the photovoltaic array should be integrated with components for charge regulation and storage.

28 to 36 cells in series, to generate a dc output voltage of 12 V in standard illumination conditions (Fig. 1.2(b)). The 12 V modules can be used singly, or connected in parallel and series into an array with a larger current and voltage output, according to the power demanded by the application (Fig. 1.2(c)). Cells within a module are integrated with bypass and blocking diodes in order to avoid the complete loss of power which would result if one

cell in the series failed. Modules within arrays are similarly protected. The array, which is also called a photovoltaic *generator*, is designed to generate power at a certain current and a voltage which is some multiple of 12 V, under standard illumination. For almost all applications, the illumination is too variable for efficient operation all the time and the photovoltaic generator must be integrated with a charge storage system (a battery) and with components for power regulation (Fig. 1.2(d)). The battery is used to store charge generated during sunny periods and the power conditioning ensures that the power supply is regular and less sensitive to the solar irradiation. For ac electrical power, to power ac designed appliances and for integration with an electricity grid, the dc current supplied by the photovoltaic modules is converted to ac power of appropriate frequency using an *inverter*.

The design and engineering of photovoltaic systems is beyond the scope of this book. A more detailed introduction is given by Markvart [Markvart, 2000] and Lorenzo [Lorenzo, 1994]. Photovoltaic systems engineering depends to a large degree upon the electrical characteristics of the individual cells.

1.3.2. *Some important definitions*

The solar cell can take the place of a battery in a simple electric circuit (Fig. 1.3). In the dark the cell in circuit A does nothing. When it is switched on by light it develops a voltage, or e.m.f., analogous to the e.m.f. of the battery in circuit B. The voltage developed when the terminals are isolated (infinite load resistance) is called the *open circuit voltage* V_{oc}. The current drawn when the terminals are connected together is the *short circuit current* I_{sc}. For any intermediate load resistance R_L the cell develops a voltage V between 0 and V_{oc} and delivers a current I such that $V = IR_L$ and $I(V)$ is determined by the *current–voltage characteristic* of the cell under

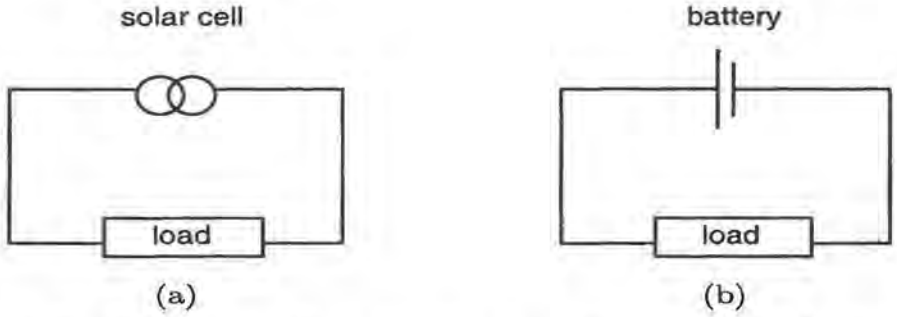


Fig. 1.3. The solar cell may replace a battery in a simple circuit.

that illumination. Thus both I and V are determined by the illumination as well as the load. Since the current is roughly proportional to the illuminated area, the *short circuit current density* J_{sc} is the useful quantity for comparison. These quantities are defined for a simple, ideal diode model of a solar cell in Sec. 1.4 below.

Box 1.1. Solar cell compared with conventional battery

The photovoltaic cell differs from a simple dc battery in these respects: the e.m.f. of the battery is due to the permanent electrochemical potential difference between two phases in the cell, while the solar cell derives its e.m.f. from a temporary change in electrochemical potential caused by light. The power delivered by the battery to a constant load resistance is relatively constant, while the power delivered by the solar cell depends on the incident light intensity, and not primarily on the load (Fig. 1.4). The battery is completely discharged when it reaches the end of its life, while the solar cell, although its output varies with intensity, is in principle never exhausted, since it can be continually recharged with light.

The battery is modelled electrically as a *voltage generator* and is characterised by its e.m.f. (which, in practice, depends upon the degree of discharge), its charge capacity, and by a polarisation curve which describes how the e.m.f. varies with current [Vincent 1997]. The solar cell, in contrast, is better modelled as a *current generator*, since for all but the largest loads the current drawn is independent of load. But its characteristics depend entirely on the nature of the illuminating source, and so I_{sc} and V_{oc} must be quoted for a known spectrum, usually for standard test conditions (defined below).

1.4. Characteristics of the Photovoltaic Cell: A Summary

1.4.1. *Photocurrent and quantum efficiency*

The photocurrent generated by a solar cell under illumination at short circuit is dependent on the incident light. To relate the photocurrent density, J_{sc}, to the incident spectrum we need the cell's *quantum efficiency*, (QE). $QE(E)$ is the probability that an incident photon of energy E will deliver one electron to the external circuit. Then

$$J_{sc} = q \int b_s(E) QE(E) dE \tag{1.1}$$

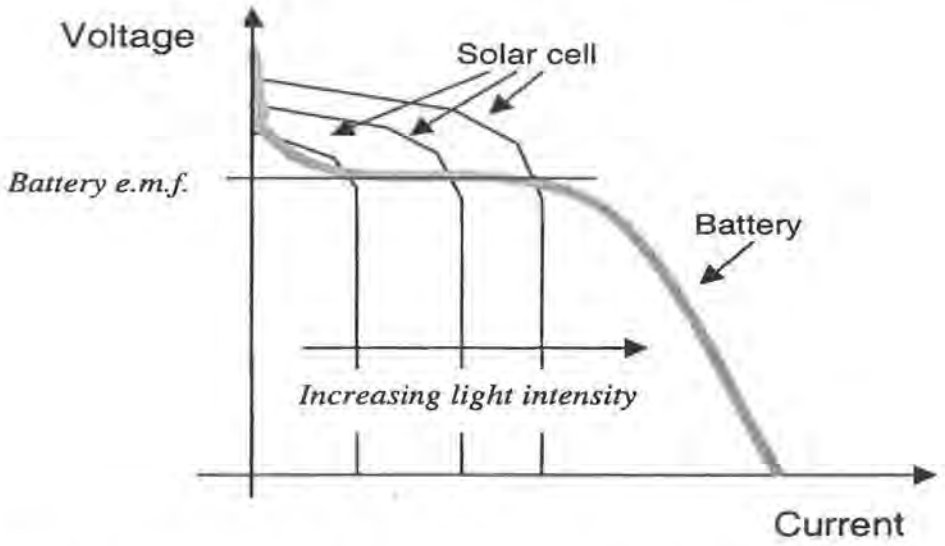


Fig. 1.4. Voltage–current curves of a conventional battery (grey) and a solar cell under different levels of illumination. A battery normally delivers a constant e.m.f. at different levels of current drain except for very low resistance loads, when the e.m.f. begins to fall. The battery e.m.f. will also deteriorate when the battery is heavily discharged. The solar cell delivers a constant current for any given illumination level while the voltage is determined largely by the resistance of the load. For photovoltaic cells it is usual to plot the data in the opposite sense, with current on the vertical axis and voltage on the horizontal axis. This is because the photovoltaic cell is essentially a current source, while the battery is a voltage source.

where $b_s(E)$ is the incident spectral photon flux density, the number of photons of energy in the range E to $E+dE$ which are incident on unit area in unit time and q is the electronic charge. QE depends upon the absorption coefficient of the solar cell material, the efficiency of charge separation and the efficiency of charge collection in the device but does not depend on the incident spectrum. It is therefore a key quantity in describing solar cell performance under different conditions. Figure 1.5 shows a typical QE spectrum in comparison with the spectrum of solar photons.

QE and spectrum can be given as functions of either photon energy or wavelength, λ. Energy is a more convenient parameter for the physics of solar cells and it will be used in this book. The relationship between E and λ is defined by

$$E = \frac{hc}{\lambda} \tag{1.2}$$

where h is Planck's constant and c the speed of light in vacuum. A convenient rule for converting between photon energies, in electron-Volts, and wavelengths, in nm, is $E/\mathrm{eV} = 1240/(\lambda/\mathrm{nm})$.

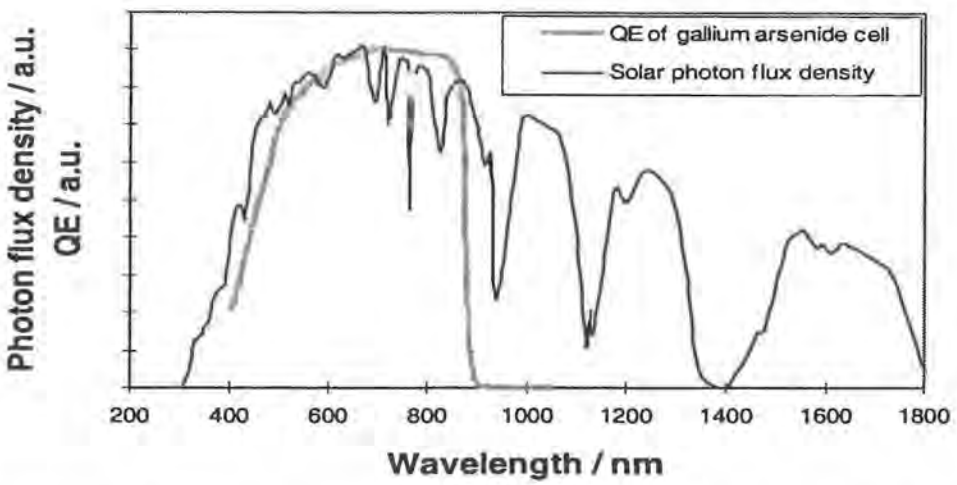


Fig. 1.5. Quantum efficiency of GaAs cell compared to the solar spectrum. The vertical scale is in arbitrary units, for comparison. The short circuit photocurrent is obtained by integrating the product of the photon flux density and QE over photon energy. It is desirable to have a high QE at wavelengths where the solar flux density is high.

1.4.2. *Dark current and open circuit voltage*

When a load is present, a potential difference develops between the terminals of the cell. This potential difference generates a current which acts in the opposite direction to the photocurrent, and the net current is reduced from its short circuit value. This reverse current is usually called the *dark current* in analogy with the current $I_{\mathrm{dark}}(V)$ which flows across the device under an applied voltage, or *bias*, V in the dark. Most solar cells behave like a diode in the dark, admitting a much larger current under forward bias ($V > 0$) than under reverse bias ($V < 0$). This rectifying behaviour is a feature of photovoltaic devices, since an asymmetric junction is needed to achieve charge separation. For an ideal diode the *dark current density* $J_{\mathrm{dark}}(V)$ varies like

$$J_{\mathrm{dark}}(V) = J_{\mathrm{o}}(e^{\mathrm{qV}/\mathrm{k_B T}} - 1) \tag{1.3}$$

where J_{o} is a constant, k_{B} is Boltzmann's constant and T is temperature in degrees Kelvin.

The overall current voltage response of the cell, its *current–voltage characteristic*, can be approximated as the sum of the short circuit photocurrent and the dark current (Fig. 1.6). This step is known as the *superposition* approximation. Although the reverse current which flows in reponse to voltage in an illuminated cell is not formally equal to the current which flows in the dark, the approximation is reasonable for many photovoltaic materials and

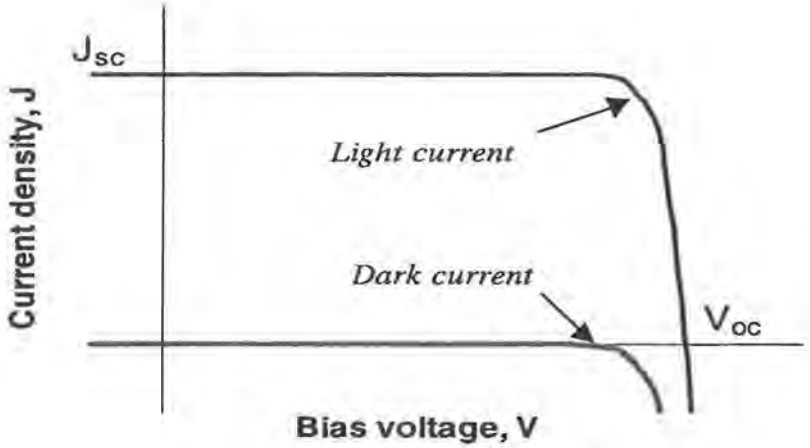


Fig. 1.6. Current–voltage characteristic of ideal diode in the light and the dark. To a first approximation, the net current is obtained by shifting the bias dependent dark current up by a constant amount, equal to the short circuit photocurrent. The sign convention is such that the short circuit photocurrent is positive.

will be used for the present discussion. The sign convention for current and voltage in photovoltaics is such that the photocurrent is positive. This is the opposite to the usual convention for electronic devices. With this sign convention the net current density in the cell is

$$J(V) = J_{\rm sc} - J_{\rm dark}(V)\,, \tag{1.4}$$

which becomes, for an ideal diode,

$$J = J_{\rm sc} - J_{\rm o}(e^{\rm qV/k_B T} - 1)\,. \tag{1.5}$$

This result is derived in Chapter 6.

When the contacts are isolated, the potential difference has its maximum value, the open circuit voltage $V_{\rm oc}$. This is equivalent to the condition when the dark current and short circuit photocurrent exactly cancel out. For the ideal diode, from Eq. 1.5,

$$V_{\rm oc} = \frac{kT}{q} \ln\left(\frac{J_{\rm sc}}{J_0} + 1\right). \tag{1.6}$$

Equation 1.6 shows that $V_{\rm oc}$ increases logarithmically with light intensity. Note that voltage is defined so that the photovoltage occurs in forward bias, where $V > 0$.

Figure 1.6 shows that the current–voltage product is positive, and the cell generates power, when the voltage is between 0 and $V_{\rm oc}$. At $V < 0$, the illuminated device acts as a photodetector, consuming power to generate a

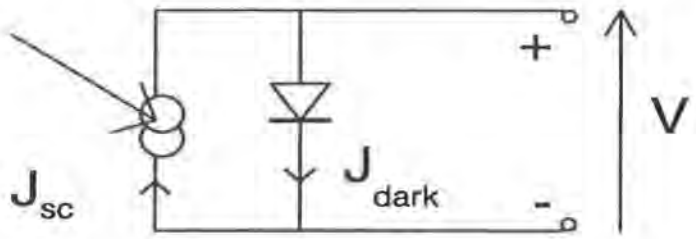


Fig. 1.7. Equivalent circuit of ideal solar cell.

photocurrent which is light dependent but bias independent. At $V > V_{oc}$, the device again consumes power. This is the regime where light emitting diodes operate. We will see later that in some materials the dark current is accompanied by the emission of light.

Electrically, the solar cell is equivalent to a current generator in parallel with an asymmetric, non linear resistive element, *i.e.*, a diode (Fig. 1.7). When illuminated, the ideal cell produces a photocurrent proportional to the light intensity. That photocurrent is divided between the variable resistance of the diode and the load, in a ratio which depends on the resistance of the load and the level of illumination. For higher resistances, more of the photocurrent flows through the diode, resulting in a higher potential difference between the cell terminals but a smaller current though the load. The diode thus provides the photovoltage. Without the diode, there is nothing to drive the photocurrent through the load.

1.4.3. *Efficiency*

The operating regime of the solar cell is the range of bias, from 0 to V_{oc}, in which the cell delivers power. The cell *power density* is given by

$$P = JV\,. \tag{1.7}$$

P reaches a maximum at the cell's operating point or *maximum power point*. This occurs at some voltage V_m with a corresponding current density J_m, shown in Fig. 1.8. The optimum load thus has sheet resistance given by V_m/J_m. The *fill factor* is defined as the ratio

$$FF = \frac{J_m V_m}{J_{sc} V_{oc}} \tag{1.8}$$

and describes the 'squareness' of the J–V curve.

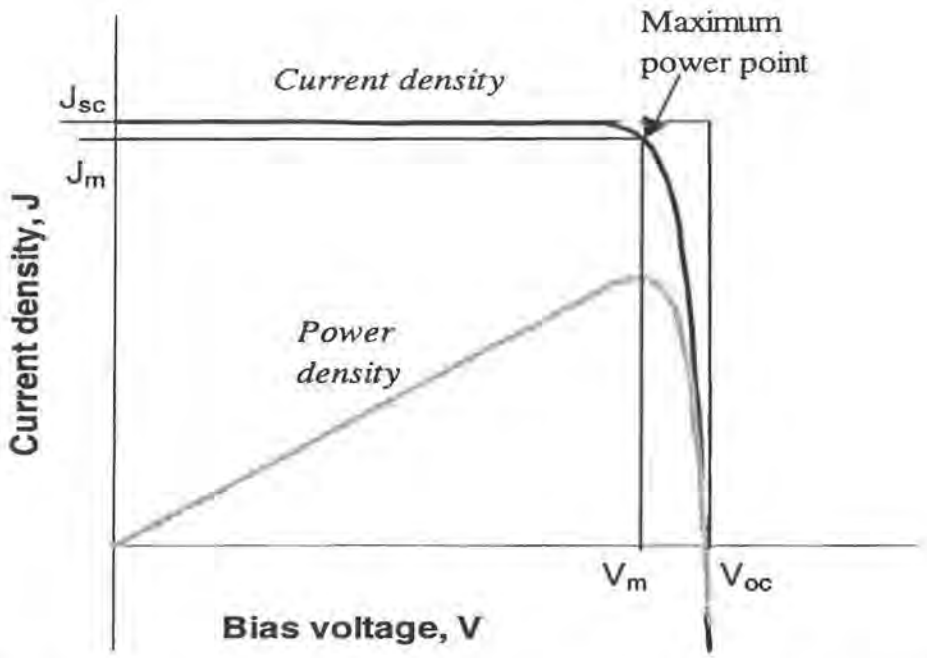


Fig. 1.8. The current voltage (black) and power–voltage (grey) characteristics of an ideal cell. Power density reaches a maximum at a bias $V_{\rm m}$, close to $V_{\rm oc}$. The maximum power density $J_{\rm m} \times V_{\rm m}$ is given by the area of the inner rectangle. The outer rectangle has area $J_{\rm sc} \times V_{\rm oc}$. If the fill factor were equal to 1, the current voltage curve would follow the outer rectangle.

The *efficiency* η of the cell is the power density delivered at operating point as a fraction of the incident light power density, $P_{\rm s}$,

$$\eta = \frac{J_{\rm m} V_{\rm m}}{P_{\rm s}} \,. \tag{1.9}$$

Efficiency is related to $J_{\rm sc}$ and $V_{\rm oc}$ using FF,

$$\eta = \frac{J_{\rm sc} V_{\rm oc} FF}{P_{\rm s}} \,. \tag{1.10}$$

These four quantities: $J_{\rm sc}$, $V_{\rm oc}$, FF and η are the key performance characteristics of a solar cell. All of these should be defined for particular illumination conditions. The Standard Test Condition (STC) for solar cells is the Air Mass 1.5 spectrum, an incident power density of 1000 W $\rm m^{-2}$, and a temperature of 25°C. (Standard and other solar spectra are discussed in Chapter 2.) The performance characteristics for the most common solar cell materials are listed in Table 1.1.

Table 1.1 shows that solar cell materials with higher $\rm J_{sc}$ tend to have lower $\rm V_{oc}$. This is a consequence of the material used, and particularly of

Table 1.1. Performance of some types of PV cell [Green *et al.*, 2001].

Cell Type	Area (cm^2)	V_{oc} (V)	J_{sc} (mA/cm^2)	FF	Efficiency (%)
crystalline Si	4.0	0.706	42.2	82.8	24.7
crystalline GaAs	3.9	1.022	28.2	87.1	25.1
poly-Si	1.1	0.654	38.1	79.5	19.8
a-Si	1.0	0.887	19.4	74.1	12.7
$CuInGaSe_2$	1.0	0.669	35.7	77.0	18.4
CdTe	1.1	0.848	25.9	74.5	16.4

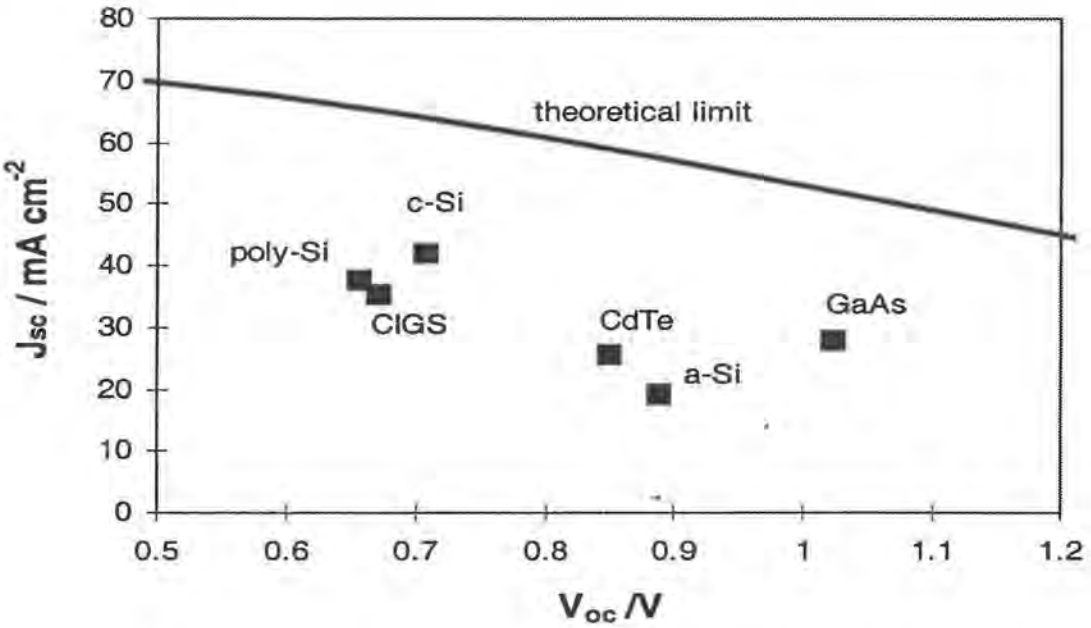


Fig. 1.9. Plot of J_{sc} against V_{oc} for the cells listed in Table 1.1. Materials with high V_{oc} tend to have lower J_{sc}. This is due to the band gap of the semiconductor material. The grey line shows the relationship expected in the theoretical limit.

the band gap of the semiconductor. In Chapter 2 we will see that there is a fundamental compromise between photocurrent and voltage in photovoltaic energy conversion. Figure 1.9 illustrates the correlation between J_{sc} and V_{oc} for the cells in Table 1.1, together with the relationship for a cell of maximum efficiency.

1.4.4. *Parasitic resistances*

In real cells power is dissipated through the resistance of the contacts and through leakage currents around the sides of the device. These effects are

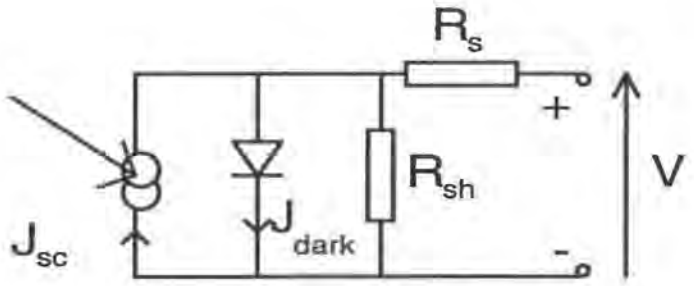


Fig. 1.10. Equivalent circuit including series and shunt resistances.

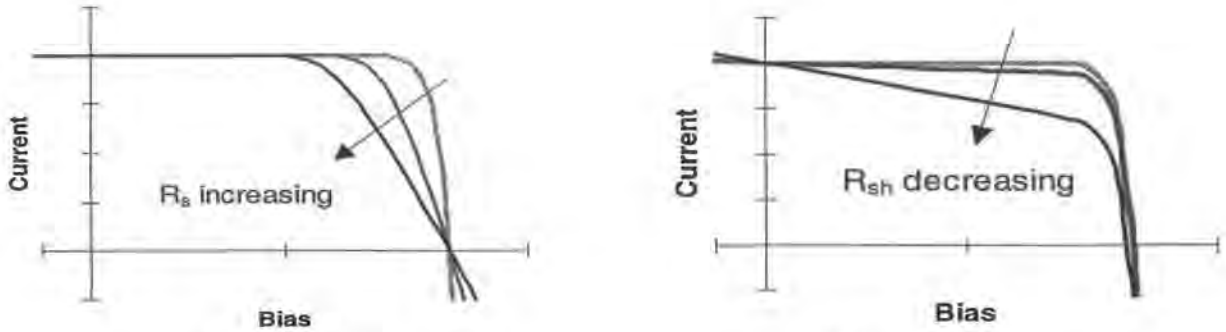


Fig. 1.11. Effect of (a) increasing series and (b) reducing parallel resistances. In each case the outer curve has $R_s = 0$ and $R_{sh} = \infty$. In each case the effect of the resistances is to reduce the area of the maximum power rectangle compared to $J_{sc} \times V_{oc}$.

equivalent electrically to two parasitic resistances in series (R_s) and in parallel (R_{sh}) with the cell (Fig. 1.10).

The series resistance arises from the resistance of the cell material to current flow, particularly through the front surface to the contacts, and from resistive contacts. Series resistance is a particular problem at high current densities, for instance under concentrated light. The parallel or shunt resistance arises from leakage of current through the cell, around the edges of the device and between contacts of different polarity. It is a problem in poorly rectifying devices.

Series and parallel resistances reduce the fill factor as shown in Fig. 1.11. For an efficient cell we want R_s to be as small and R_{sh} to be as large as possible.

When parasitic resistances are included, the diode equation becomes

$$J = J_{sc} - J_o(e^{q(V+JAR_s)/kT} - 1) - \frac{V + JAR_s}{R_{sh}}. \qquad (1.11)$$

1.4.5. *Non-ideal diode behaviour*

The 'ideal' diode behaviour of Eq. 1.5 is seldom seen. It is common for the dark current to depend more weakly on bias. The actual dependence on V is quantified by an ideality factor, m and the current–voltage characteristic given by the non-ideal diode equation,

$$J = J_{sc} - J_o(e^{qV/mk_BT} - 1) \tag{1.12}$$

m typically lies between 1 and 2. The reasons for non-ideal behaviour will be discussed in later chapters.

1.5. Summary

A photovoltaic cell consists of a light absorbing material which is connected to an external circuit in an asymmetric manner. Charge carriers are generated in the material by the absorption of photons of light, and are driven towards one or other of the contacts by the built-in spatial asymmetry. This light driven charge separation establishes a photovoltage at open circuit, and generates a photocurrent at short circuit. When a load is connected to the external circuit, the cell produces both current and voltage and can do electrical work.

The size of the current generated by the cell in short circuit depends upon the intensity and the energy spectrum of the incident light. Photocurrent is related to incident spectrum by the quantum efficiency of the cell, which is the probability of generating an electron per incident photon as a function of photon energy. When a load is present, a potential difference is created between the terminals of the cell and this drives a current, usually called the dark current, in the opposite direction to the photocurrent. As the load resistance is increased, the potential difference increases and the net current decreases until the photocurrent and dark current exactly cancel out. The potential difference at this point is called the open circuit voltage. At some point before V_{oc} is reached, the current–voltage product is maximum. This is the maximum power point and the cell should be operated with a load resistance which corresponds to this point.

The solar cell can be modelled as a current generator in parallel with an ideal diode, and the current–voltage characteristic given by the ideal diode equation, Eq. 1.5. In real cells, the behaviour is degraded by the presence of series and parallel resistances.

References

D. Anderson, *Clean Electricity from Photovoltaics*, eds. M.D. Archer and R.D. Hill (London: Imperial College Press, 2001).

M.A. Green, "Photovoltaics: Coming of age", *Conf. Record 21st IEEE Photovoltaic Specialists Conf.*, 1–7 (1990).

E. Lorenzo, *Solar Electricity: Engineering of Photovoltaic Systems* (Progensa, 1994).

T. Markvart, *Solar Electricity* (Wiley, 2000).

J.N. Shive, *Semiconductor Devices* (Van Nostrand, 1959).

C.A. Vincent, *Modern Batteries* (Arnold, 1997).

M. Wolf, "Historical development of solar cells", *Proc. 25th Power Sources Symposium*, 1972. In *Solar Cells*, ed. C.E. Backus (IEEE Press, 1976).

Chapter 3

Electrons and Holes in Semiconductors

3.1. Introduction

In Chapter 2 we considered the requirements for photovoltaic energy conversion. We concluded that a suitable photovoltaic material should absorb visible light, possess a band gap between the initial, occupied states and the final, unoccupied states which are involved in photon absorption and be able to transport charges efficiently. The gap is necessary in order to make the extra potential energy which electrons gain from photon absorption available as electrical energy. All semiconducting and insulating solids possess an energy gap but only semiconductors are suitable for photovoltaics, because the band gap of insulators is too large to permit absorption of visible light. Most molecular solids possess an energy gap, but with the exception of some conjugated molecular materials, the charge transport is too inefficient to be useful for solar cells.

The band gap is important because it enables excited electrons to remain in higher energy levels for long enough to be exploited. If electrons were simply promoted through a continuum of energy levels as in a metal, for example, they would very quickly decay back down to their ground state through a series of intermediate levels. The abundance of empty levels at intermediate energy means that the probability of an excited electron being scattered to a lower energy state within the thermal energy of the original level, is high. At room temperature this 'thermalisation' of carriers to the band edge occurs in femtoseconds. When an electron is excited across a band gap, it quickly decays to the lowest available energy state in the conduction band (the conduction band edge) but the next stage — decay across the band gap to a vacant site in the valence band — is slow, as shown in Fig. 3.1.

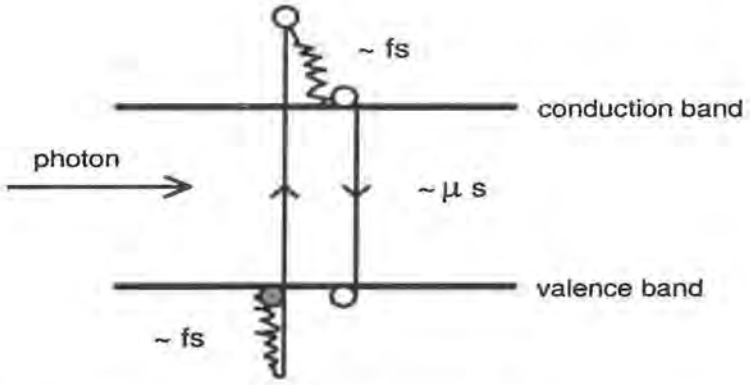


Fig. 3.1. Promotion of an electron from valence band to conduction band by a photon, thermalisation to the band edges, and recombination. Although thermalisation is very fast and occurs in femtoseconds, relaxation across the band gap is many orders of magnitude slower.

In practice, the great majority of experience with photovoltaic materials is based on a small number of semiconductor materials. In this chapter and the next we will examine some of the basic physical principles of semiconductors. We will focus on the electronic and optical properties of crystalline materials. We will show how the optical and electronic properties result from the crystal structure. In this chapter we introduce the concepts of density of states, electron distribution function, doping, quasi thermal equilibrium and the definition of electron and hole currents. In Chapter 4 we will treat the processes of charge carrier generation and recombination in semiconductors and show how to set up the semiconductor transport equations, which are key to the physics of photovoltaic devices.

3.2. Basic Concepts

3.2.1. *Bonds and bands in crystals*

When a pair of atoms are brought together into a molecule, their atomic orbitals combine to form pairs of molecular orbitals arranged slightly higher and slightly lower in energy than each original level. We say that the energy levels have *split*. When a very large number of atoms come together in a solid, each atomic orbital splits into a very large number of levels, so close together in energy that they effectively form a continuum, or *band*, of allowed levels. The bands due to different molecular orbitals may or may not overlap. The energy distribution of the bands depends upon the electronic properties of the atoms and the strength of the bonding between them.

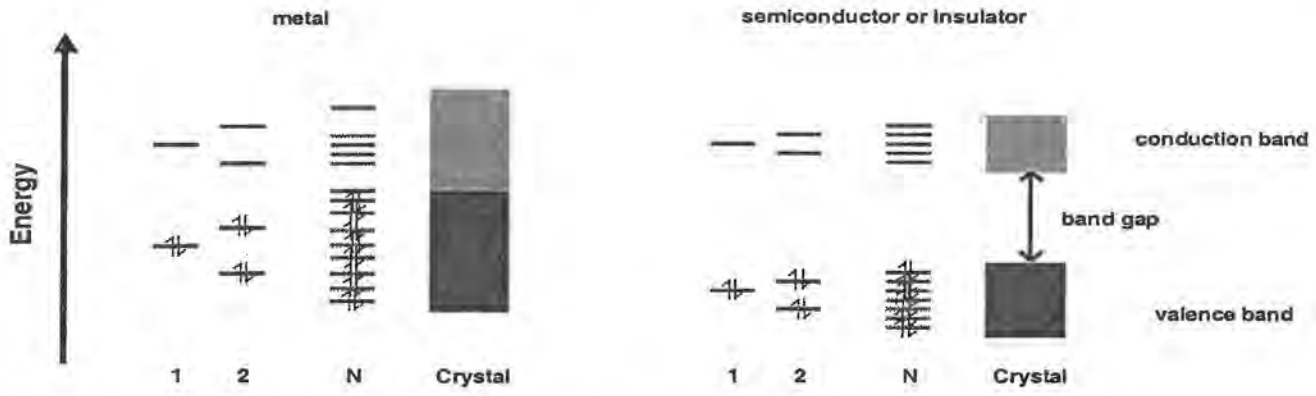


Fig. 3.2. As the number of atoms in a molecule or cluster increases, the atomic orbitals split into multiple levels, eventually coming together to form bands in the limit of many atoms. Overlapping (left) and non overlapping (right) bands represent a metal and a semiconductor, respectively.

Bands are occupied or not depending upon whether the original molecular orbitals were occupied. The highest occupied band, which contains the valence electrons, is normally called the *valence band* (VB). The lowest unoccupied band is called the *conduction band* (CB). If the valence band is partly full, or if it overlaps in energy with the lowest unoccupied band, the solid is a metal. In a metal, the availability of empty states at similar energies makes it easy for a valence electron to be excited, or *scattered*, into a neighbouring state. These electrons can readily act as transporters of heat or charge, and so the solid conducts heat and electric current.

If the valence band is completely full and separated from the next band by an energy gap, then the solid is a semiconductor or an insulator. The electrons in the valence band are all completely involved in bonding and cannot be easily removed. They require an energy equivalent to the band gap to be removed to the nearest available unoccupied level. These materials therefore do not conduct heat or electricity easily.

Semiconductors are distinguished, roughly, as the group of materials with a band gap in the range 0.5 to 3 eV. Semiconductors have a small conductivity in the dark because only a small number of valence electrons will have enough kinetic energy at room temperature to be excited across the band gap at room temperature. This *intrinsic* conductivity decreases with increasing band gap. *Insulators* are wider band gap materials whose conductivity is negligible at room temperature. Materials of band gap < 0.5 eV have a reasonably high conductivity and are usually known as *semimetals.*

When the solid forms a regular crystal, then the energies of the bands, or the *band structure* can be predicted exactly. Exactly which crystal structure

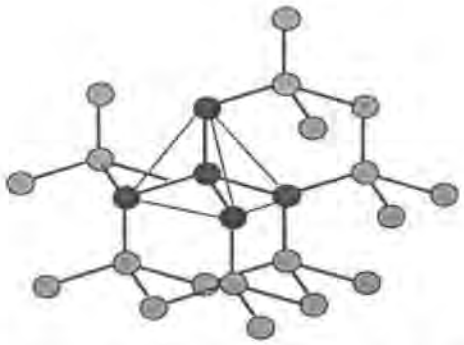


Fig. 3.3. Structure of crystalline silicon. Each silicon atom is bonded to four others in a tetrahedral arrangement.

a solid will adopt depends upon the number of valence electrons and other factors. It will prefer a configuration that minimises the total energy. A band gap is likely to arise in a crystal structure where all valence electrons are used in bonding. For example, the silicon atom possesses four valence electrons in its outermost 3s and 3p atomic orbitals. If the atom could form bonds with four neighbours, each contributing one electron, then all valence electrons would be occupied in bonding. In crystalline silicon this is achieved by the hybridisation of the 3s and 3p orbitals into a set of four degenerate sp^3 orbitals, which are directed in space with tetrahedral symmetry, and allow the formation of four identical silicon-silicon bonds with neighbouring atoms (Fig. 3.3). When the crystalline solid is formed, the sp^3 orbitals split to form a pair of bands. The lower, bonding band is completely filled by the valence electrons and the upper, antibonding or conduction band, is completely empty in a perfect crystal at absolute zero. Some solids can exist in different phases. For instance, carbon can form either the highly insulating, wide band gap, diamond crystal structure where all four valence electrons are tied up in covalent bonds with neighbouring carbon atoms, or the semimetallic graphite structure where only three valence electrons are involved in directed bonds with neighbouring atoms while the remaining electron is loosely involved in bonding with another plane of carbon atoms and is relatively mobile.

3.2.2. *Electrons, holes and conductivity*

At absolute zero temperature, a pure semiconductor is unable to conduct heat or electricity since all of its electrons are involved in bonding. As the temperature is raised, the electrons gain some kinetic energy from vibrations

of the lattice and some are able to break free. The freed electrons have been excited into the conduction band and are able to travel and transport charge or energy. Meanwhile, the vacancies which they have left behind are able to move, and can also conduct (see Box 3.1). The higher the temperature, the greater the number of electrons and holes which are mobilised, and the higher the conductivity.

Box 3.1. Electrons and holes

When an electron is removed from a bond between atoms, a positively charged vacancy remains. This vacancy can be filled by another electron, most easily by electrons which are involved in neighbouring bonds. If this

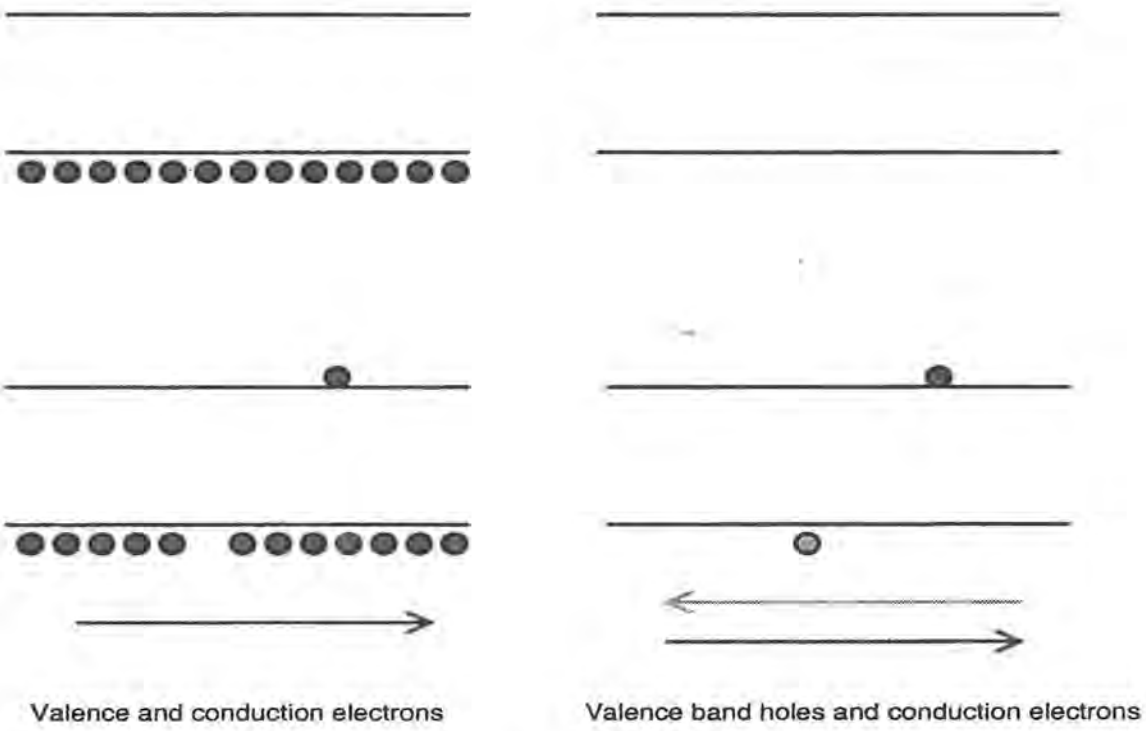


Fig. 3.4. The valence band is represented, very crudely, by marbles in a tray. When the valence band is filled (top) no net electron motion can occur, despite an applied force, because there are no vacant orbitals for the electrons to be scattered into. Motion becomes possible when an electron is removed (bottom), freeing up some vacant orbitals. The nearly filled valence band is most easily characterised in terms of those orbitals which are not filled, called 'holes'. The dynamics of the hole are equivalent to those of a particle with the opposite charge, wavevector and energy to the missing electron. It responds to an applied field by moving in the opposite sense to an electron. The two situations depicted on the right are equivalent to those on the left, using holes to represent missing electrons.

happens, the vacancy moves to the neighbouring bond. The creation of holes in the valence band creates a means whereby charge can be transferred. In the presence of an electric field, a bonding, or *valence*, electron can respond to the field by moving into the hole. The vacancy which it leaves behind can be filled by another valence electron, and so on. The net movement of the valence electrons against the field is equivalent to the movement of a small number of positive *holes* in the direction of the field. Since there are many fewer valence holes than electrons it is much more convenient to think in terms of the motion of holes through the valence band. These positive holes can be characterised with a mobility and an effective mass, just like conduction electrons.

A semiconductor can be made to conduct in other ways: if the material is exposed to light of energy greater than the band gap, a photon can be absorbed by a valence electron to free it from the lattice and promote it into the conduction band. This creates a vacancy in the valence band. The free electron and hole created in this way are available to conduct electricity. This is called *photoconductivity*, the phenomenon which first drew attention to photovoltaic materials (see Chapter 1).

Alternatively, impurities with different numbers of valence electrons or different bond strengths can be added to the material. Electrons or holes may be freed more easily from these impurities than from the native atoms, thereby increasing the number of carriers normally available for conduction compared to the pure semiconductor. Deliberate addition of impurities to increase conductivity in this way is called *doping*. Both of these processes — photoconductivity and doping — will be discussed later.

In the next section we will consider what controls the energetic arrangement and the dynamics of the electrons and holes in a semiconductor.

3.3. Electron States in Semiconductors

3.3.1. *Band structure*

To locate the energy levels of an atom or a molecule, we need to solve Schrödinger's equation. In a solid we do the same thing, except that now we need to take account of an infinite array of atomic potentials, and not just the few which make up our molecule. In a crystalline material we are able to take advantage of the fact that the atomic potentials are arranged on an infinite periodic lattice. The *periodicity* of the lattice means that

the probability distribution of the electrons must also be periodic. There is no physical reason why an electron should occupy a particular site within one unit cell rather than within any other. Because the lattice is *infinite*, the electrons should form delocalised states which extend throughout the crystal, just like an electron in free space. The sort of wavefunction which satisfies these conditions is the *Bloch* wavefunction. It is the product of a periodic part $u_{\mathbf{ik}}(\mathbf{r})$, which possesses the periodicity of the lattice, and a plane wave part. The periodic part is related to one of the atomic orbitals involved in the crystal structure and inherits its symmetry. The plane wave modulates the wavefunction and is analogous to the wavefunction of an electron in free space. Each Bloch state is thus characterised by a crystal band i and a wavevector $\mathbf{k}$,

$$\psi(\mathbf{k},\mathbf{r}) = u_{\mathbf{ik}}(\mathbf{r})e^{\mathbf{ik}.\mathbf{r}} \,. \tag{3.1}$$

In quantum mechanical terms, $\mathbf{k}$ is a 'good quantum number'. For each original atomic orbital i there is a continuous set of solutions with different $\mathbf{k}$ which make up the wavefunctions in the ith crystal band. The $u_{\mathbf{ik}}(\mathbf{r})$ and eigenenergies $E(\mathbf{k})$ are, in general, found by solving the Schrödinger equation for each band i, for each $\mathbf{k}$. The map of energies E against wavevector $\mathbf{k}$ is called the crystal *band structure* (see Box 3.2). The $u_{\mathbf{ik}}(\mathbf{r})$ are usually rapidly varying with distance and weakly dependent on $\mathbf{k}$. In a commonly used approximation, called the *effective mass approximation*, the $\mathbf{k}$ dependence of the $u_{\mathbf{ik}}(\mathbf{r})$ is neglected. Then all of the dynamic information is contained in the plane wave part of the wavefunction, and the $u_{\mathrm{i}}(\mathbf{r})$ come in through effective parameters.

Box 3.2. Band structure diagrams

$E(\mathbf{k})$ is usually plotted against $|\mathbf{k}|$ for the two or three most important directions in the crystal. Directions are labelled with Miller indices ⟨hkl⟩ where (h,k,l) is a vector parallel to $\mathbf{k}$, measured in unit cell widths along the natural axes of the crystal. For instance, the direction ⟨100⟩ represents the set of $\mathbf{k}$ parallel to the first of the natural axes, say the x axis. In a cubic system, ⟨100⟩ also represents the directions parallel to the y and z axes, because these three directions are indistinguishable in cubic symmetry.

k is varied from 0 to π/a along the chosen direction, where a is the spacing of the planes of atoms in that direction. For $k > \pi/a$, the energy spectrum repeats itself. This is because in a periodic structure with period a, wavevectors which differ by multiples of $2\pi/a$ cannot be distinguished.

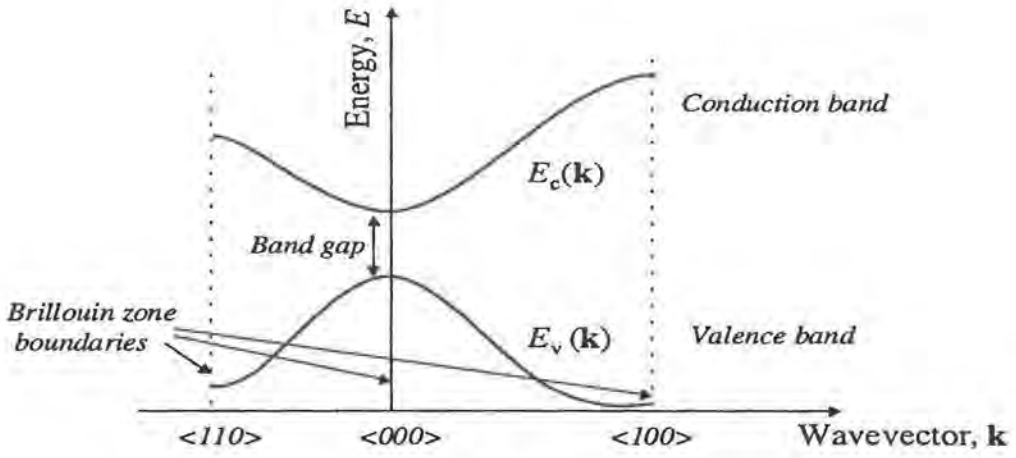


Fig. 3.5. Schematic band structure diagram of a direct gap semiconductor.

For most crystals, for $k < 0$ the function $E(\mathbf{k})$ is identical to $E(-\mathbf{k})$ since states with positive and negative k are degenerate. Therefore, all the information about $E(\mathbf{k})$ is contained in the range $0 < k < \pi/a$, and only this section needs to be plotted. The point $k = \pi/a$ is called the *Brillouin zone boundary*. At these points, the wave function is a standing wave and the gradient of $E(\mathbf{k})$ vanishes. Whether the energy reaches a maximum or a minimum depends on the symmetry of the band.

We are interested in the minimum and maximum energies of the crystal bands, since these determine the band gap, and in the form of $E(\mathbf{k})$ near to these stationary points.

3.3.2. *Conduction band*

For the conduction band, the minimum of the energy $E(\mathbf{k})$ may occur at $\mathbf{k} = 0$ or another value of $\mathbf{k}$ corresponding to an important direction in the crystal. Near to a minimum of energy at $\mathbf{k} = \mathbf{k}_{0c}$, it is often convenient to expand the $E(\mathbf{k})$ in powers of $(\mathbf{k} - \mathbf{k}_{0c})$ and approximate $E(\mathbf{k})$ by

$$E(\mathbf{k}) = E_{c0} + \frac{\hbar^2 |\mathbf{k} - \mathbf{k}_{0c}|^2}{2m_c^*} \tag{3.2}$$

where $E_{c0} = E_c(\mathbf{k}_0)$, m_c^* is a parameter with the dimensions of mass and is defined from the band structure by

$$\frac{1}{m_c^*} = \frac{1}{\hbar^2} \frac{\partial^2 E_c(k)}{\partial k^2} . \tag{3.3}$$

This is the *parabolic band approximation.* The *effective mass* m_c^* is analogous to the mass of a free electron m_0, but differs from m_0 through the different forces experienced by an electron within a crystal lattice. m_c^* can be greater or less than m_0; large values of m_c^* imply that the conduction electrons are strongly influenced by the atomic potentials. m_c^* describes how the momentum $\mathbf{P}$ conduction electron responds to an applied force, $\mathbf{F}$

$$\mathbf{F} = m_c^* \frac{d\mathbf{p}}{dt} . \tag{3.4}$$

The value of m_c^* is determined by the atomic potentials. The effective mass is one way in which the $u_i(r)$ enter into the electron dynamics in this approximation. In general m_c^* should be a *tensor* — dependent on the direction of the applied force and the directions of the crystal axes. However, close to the band minimum the band structure is often isotropic and m_c^* can be approximated by a constant.

The two terms in Eq. 3.2 represent the potential and kinetic energy of a conduction band electron with wavevector $\mathbf{k}$. It can be used to derive the electron velocity,

$$\mathbf{v} = \nabla_{\mathbf{k}} E(\mathbf{k}) = \frac{\hbar(\mathbf{k} - \mathbf{k}_{0c})}{m_c^*} , \tag{3.5}$$

and momentum

$$\mathbf{p} = m_c^* \mathbf{v} = \hbar(\mathbf{k} - \mathbf{k}_{0c}) . \tag{3.6}$$

3.3.3. *Valence band*

In the valence band, we require an expression for the energy of the hole near to the valence band maximum. Since a hole represents the absence of an electron, and electrons prefer to have as little energy as possible, the most stable situation for a hole is where the electron energy is a maximum. Therefore holes are most likely to exist near to a valence band maximum, and have kinetic energy which increases as $E(\mathbf{k})$ is reduced. For a hole near a VB maximum at $\mathbf{k} = \mathbf{k}_{v0}$, the energy can be approximated by

$$E(\mathbf{k}) = E_{v0} - \frac{\hbar^2 |\mathbf{k} - \mathbf{k}_{v0}|^2}{2m_v^*} , \tag{3.7}$$

where $E_{v0} = E_v(\mathbf{k}_{v0})$, resulting in a hole velocity of

$$\mathbf{v} = -\frac{\hbar(\mathbf{k} - \mathbf{k}_{v0})}{m_v^*} \tag{3.8}$$

and momentum

$$\mathbf{p} = -\hbar(\mathbf{k} - \mathbf{k}_{v0}) \tag{3.9}$$

where the VB effective mass, m_v^* is defined through

$$\frac{1}{m_v^*} = -\frac{1}{\hbar^2}\frac{\partial^2 E_v(k)}{\partial k^2} \tag{3.10}$$

so that the mass is normally positive. In general, the effective masses of electrons and holes in the same semiconductor are different, because the curvature of the conduction and valence bands are different.

3.3.4. *Direct and indirect band gaps*

The minimum amount of energy which will promote an electron from the valence band into the conduction band is called the fundamental band gap, E_g. If the CB minimum and VB maximum occur at the same value of $\mathbf{k}$, then a photon of energy E_g is sufficient to create an electron-hole pair. This type of semiconductor is called a *direct band gap* material.

Now, if the CB minimum and VB maximum occur at *different* values of $\mathbf{k}$, a photon with energy E_g is *not* on its own sufficient to create an electron-hole pair. Promoting an electron from the VB maximum to the CB minimum would cause a change in its momentum, of $\eta(\mathbf{k}_{c0} - \mathbf{k}_{v0})$. Momentum has to be conserved in a crystal, but since photons possess virtually no momentum, that extra momentum must be supplied by something else. Usually it is supplied by a *phonon* — a lattice vibration — of the correct momentum. The phonon gives up its momentum to the electron at the moment of photon absorption so that both energy and momentum are conserved. This type of semiconductor is called an *indirect band gap* material. Clearly, photon absorption can only happen in indirect gap materials if there are enough phonons available with the required value of $\mathbf{k}$. This means that optical absorption is generally weaker for indirect than direct gap materials, and that it is more heavily dependent on temperature. The dependence of absorption on photon energy is also more gradual, as will be discussed in Chapter 4.

For device physics purposes, usually only the CB minimum and VB maximum energy are represented, in an energy–distance band diagram such as Fig. 3.7. This gives us no information on the $\mathbf{k}$ dependence of the band structure, and does not tell us whether the band gap is direct or indirect.

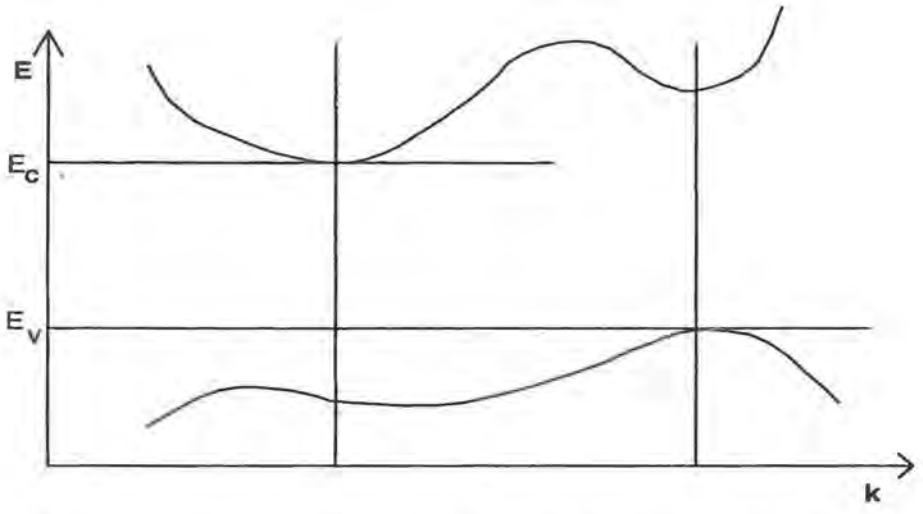


Fig. 3.6. Band structure diagram showing indirect band gap. The maximum energy of the valence band occurs at a different wavevector to the minimum of the conduction band. Excitation of an electron across the fundamental band gap requires a change in electron momentum. This may be supplied by a phonon.

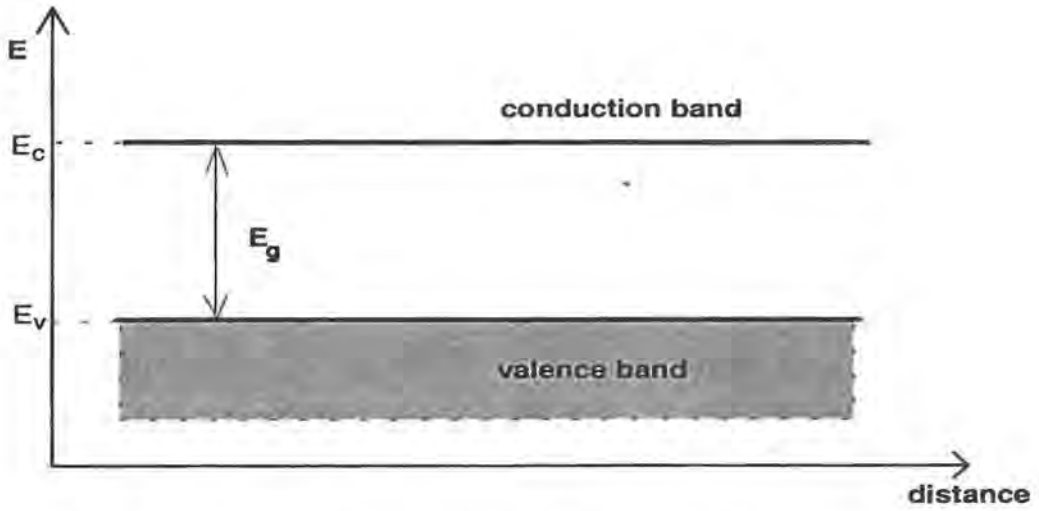


Fig. 3.7. Energy–distance band diagram of semiconductor.

3.3.5. *Density of states*

According to the Pauli exclusion principle, each quantum state can support only two electrons of different spin. Since each quantum state in a crystal is defined by a unique value of **k** (**k** is a 'good quantum number'), there should be two electrons per **k** value. A crystal of volume $L \times L \times L$ can support $(L/2\pi)^3$ different values of **k**, and hence there are $(1/2\pi)^3$ different

$\mathbf{k}$ states per unit crystal volume. This follows from the requirement that $kL/2\pi$ must be an integer for a wavefunction confined in a region of width L. Including spin degeneracy, the density of electron states per unit crystal volume, $g(\mathbf{k})$, is

$$g(\mathbf{k})d^3\mathbf{k} = \frac{2}{(2\pi)^3}d^3\mathbf{k}\,. \tag{3.11}$$

It is assumed that the electron levels are close enough for the integral over $\mathbf{k}$ to be a good approximation to the sum of $\mathbf{k}$ states.

It is generally much more useful to know the density of states (DOS) in terms of energy, rather than wavevector. We use the band structure $E(\mathbf{k})$ to convert the integral over wavevector space in Eq. 3.11 into an integral over energy. If the band structure is isotropic about the band minimum, assumed here to be at $\mathbf{k} = 0$, we can write

$$g(\mathbf{k})d^3\mathbf{k} = g(\mathbf{k}) \times 4\pi k^2 dk \tag{3.12}$$

whence

$$g(E)dE = g(\mathbf{k}) \times 4\pi k^2 \times \frac{dk}{dE}dE\,. \tag{3.13}$$

(If the band minimum is at $\mathbf{k_0} \neq 0$ we simply expand about $(\mathbf{k} - \mathbf{k_0})$. The result for $g(E)$ is the same.) Substituting for $g(\mathbf{k})$,

$$g(E) = \frac{2}{(2\pi)^3} \times 4\pi k^2 \times \frac{dk}{dE}\,. \tag{3.14}$$

In the parabolic band approximation, $E(k)$ for an electron in the CB is given by Eq. 3.2. From this we obtain k^2

$$k^2 = \frac{2m_c^*}{\hbar^2}(E - E_{c0})\,. \tag{3.15}$$

Differentiating Eq. 3.15 to find $\frac{dk}{dE}$ and substituting in 3.14 we find the DOS for the CB,

$$g_c(E) = \frac{1}{2\pi^2}\left(\frac{2m_c^*}{\hbar^2}\right)^{3/2}(E - E_{c0})^{3/2}\,. \tag{3.16}$$

In a similar way we obtain the valence band density of states

$$g_v(E) = \frac{1}{2\pi^2}\left(\frac{2m_v^*}{\hbar^2}\right)^{3/2}(E_{v0} - E)^{3/2}\,. \tag{3.17}$$

The parabolic band approximation has only limited validity for real materials. At energies far from the principal band minimum (or maximum) electron states at other symmetry points in the band structure must be included, distorting the DOS from the form in Eqs. 3.16 and 3.17. Where the parabolic band approximation is really inappropriate we should use the form of $E(\mathbf{k})$ in Eq. 3.14.

Excitons

So far we have treated the electron and hole states independently. We have not considered interactions between charge carriers which may influence the density of states. An important effect is the Coulombic interaction between electrons and holes, which gives rise to bound states called *excitons*. Excitonic states are built up from combinations of electron and hole states of the same $\mathbf{k}$. Physically they may be stationary ($\mathbf{k} = 0$) or mobile within the crystal. The exciton energies may be calculated by considering the Coulombic interaction as a perturbation of the independent particle crystal potential. A basis state of the form

$$\Psi_{\mathrm{ex}} = \psi_{\mathrm{e}}(\mathbf{k}, \mathbf{r})\psi_{\mathrm{h}}(\mathbf{k}, \mathbf{r}') \quad .$$

then satisfies a hydrogenic effective mass equation

$$-\frac{\eta^2}{2\mu^*}\nabla^2\Psi_{\mathrm{ex}} - \frac{q^2}{4\pi\varepsilon_{\mathrm{s}}|\mathbf{r}-\mathbf{r}'|} = E_{\mathrm{ex}}\Psi_{\mathrm{ex}}$$

where μ^* is the reduced effective mass of the electron-hole pair, ε_{s} the dielectric permittivity of the semiconductor and $-E_{\mathrm{ex}}$ is the binding energy of the excitonic state. This has solutions for E_{ex}

$$E_{\mathrm{ex}} = \frac{\mu^*}{m_0}\frac{\varepsilon_0^2}{\varepsilon_{\mathrm{s}}^2}\frac{Ryd}{l^2} \qquad \text{where } l = 1, 2, 3, \ldots$$

where Ryd is the Rydberg energy: The electron and hole densities of states are thus modified by the presence of a series of bound states. Although excitonic states are not relevant for electrons or holes in isolation, the excitonic levels are usually drawn on band diagrams as a series of levels below the conduction band edge. The binding energies are greatly reduced compared to the hydrogen atom on account of the screening effect of the semiconductor. In a typical semiconductor, the binding energy is a few meV, so the excitons are likely to be ionised at room temperature. Excitonic states at band

minima and maxima are important in the optical properties for semiconductors, particularly in low dimensional structures. In molecular materials, excitonic states are very important and dominate over band states, so that this perturbation approach is not appropriate.

Box 3.3. Low dimensional systems

The above treatment of density of states is based on a material system with isotropic symmetry and would be invalid for highly non-isotropic systems. One important such class are *low dimensional structures* where carriers are confined in one, two or three dimensions by the presence of quantum heterostructures such as quantum wells, quantum wires and quantum dots. In these cases the electron motion is strongly quantised in the confined directions, increasing the minimum energy above the minimum which applies in the bulk. Propagating $\mathbf{k}$ states can only exist in the unconfined directions, and so the sum over $\mathbf{k}$ states in Eq. 3.14 must be taken in two, one or zero directions rather than three. The result is that the form of $g(E)$ is changed. For two dimensional systems, the DOS, now per unit area, is the step function

$$g_{2D}(E) = \left(\frac{m_c^*}{\hbar^2}\right)\theta(E - E_c) \tag{3.18}$$

where $\theta(E)$ is the Heaviside function and E_c represents the new, shifted, minimum energy state. In one dimensional systems we have the DOS per unit length

$$g_{1D}(E) = \frac{1}{2\pi}\left(\frac{2m_c^*}{\hbar^2}\right)^{1/2}(E - E_c)^{-1/2}\,. \tag{3.19}$$

These configurations are relevant to some of the novel photovoltaic materials and device designs, which will be discussed in Chapter 10.

3.3.6. *Electron distribution function*

To calculate the densities of electrons and holes from the density of states, we need to know how the states are filled. Let's define a distribution function, $f(\mathbf{k}, \mathbf{r})$, such that f is the probability that at a point $\mathbf{r}$ the electron state of wavevector $\mathbf{k}$ is occupied. Then the density of electrons in a small

volume $d^3\mathbf{k}$ around $\mathbf{k}$ is given by

$$n(\mathbf{r})\mathbf{d}^3\mathbf{k} = g(\mathbf{k})f(\mathbf{k},\mathbf{r})\mathbf{d}^3\mathbf{k}$$

and the total electron density in the conduction band by the sum over $\mathbf{k}$ states

$$n(\mathbf{r}) = \int_{\text{conduction band } \mathbf{k}} g_{\mathrm{c}}(\mathbf{k})f(\mathbf{k},\mathbf{r})\mathbf{d}^3\mathbf{k}\,. \tag{3.20}$$

Since a hole is an unoccupied electron state, the probability of a hole state being occupied is simply $1-f(\mathbf{k},\mathbf{r})$. Thus the density of holes in the valence band is given by

$$p(\mathbf{r}) = \int_{\text{valence band } \mathbf{k}} g_{\mathrm{v}}(\mathbf{k})(1-f(\mathbf{k},\mathbf{r}))\mathbf{d}^3\mathbf{k}\,. \tag{3.21}$$

3.3.7. *Electron and hole currents*

In the effective mass approximation, used so far, electron and hole densities are represented by probability distributions of plane waves. To find the electron and hole *currents*, we need to weight each probability carrier density by its group velocity. Now, by analogy with the result for free electrons, the group velocity of an electron of effective mass m_{c}^* in state $\mathbf{k}$ is $\frac{\eta\mathbf{k}}{m_{\mathrm{c}}^*}$. The net electron current density in the conduction band should therefore be given by the sum of contributions over k states,

$$J_{\mathrm{n}}(\mathrm{r}) = -\frac{q\hbar}{m_{\mathrm{c}}^*}\int_{\text{conduction band } \mathbf{k}} \mathbf{k}g_{\mathrm{c}}(\mathbf{k})f(\mathbf{k},\mathbf{r})\mathbf{d}^3\mathbf{k} \tag{3.22}$$

and the hole current density in the conduction band by

$$J_{\mathrm{p}}(\mathbf{r}) = \frac{q\hbar}{m_{\mathrm{v}}^*}\int_{\text{valence band } \mathbf{k}} \mathbf{k}g_{\mathrm{v}}(\mathbf{k})(1-f(\mathbf{k},\mathbf{r}))\mathbf{d}^3\mathbf{k} \tag{3.23}$$

where we make use of the fact that current is parallel to hole flow, and antiparallel to electron flow.

So, in order to calculate the densities and dynamics of electrons and holes in our semiconductor we need to know the distribution function $f(\mathbf{k},\mathbf{r})$ as well as the DOS. In general, f is position dependent and relies upon (external) factors which may vary within the material. For the special case of a system in thermal equilibrium, however, this distribution is spatially invariant and described by *Fermi Dirac statistics*. This is treated in the next section.

3.4. Semiconductor in Equilibrium

3.4.1. *Fermi Dirac statistics*

At absolute zero the electrons have no kinetic energy and always occupy the lowest available levels, filling up the available states in order of increasing energy. The energy up to which states are filled is called the *Fermi energy*, E_F. At finite temperatures the electrons have some kinetic energy, and some of them are excited into states above E_F, leaving some states below E_F unoccupied.

Thermal equilibrium means that there are no *net* exchanges of particles or energy between different points in the semiconductor or between the semiconductor and its surroundings. Every point in the semiconductor is at the ambient temperature, T, so that all carriers have the same average kinetic or *internal* energy of $\frac{3}{2}k_BT$ where k_B is Boltzmann's constant. The distribution of particles has settled so that that there are no net particle flows and E_F is the same at all points. In these equilibrium conditions, Fermi Dirac statistics describe the most stable energetic distribution of electrons which is consistent with their internal energy (k_BT) and their statistical potential energy (E_F).

For this special case, f is dependent upon wavevector only implicitly through the electron energy, and is independent of position.

$$f(\mathbf{k},\mathbf{r}) = f_0(E(\mathbf{k}), E_F, T) \tag{3.24}$$

where $f_0(E, E_F, T)$ is the Fermi Dirac distribution function, giving the average probability that an electron state at energy E will be occupied at some temperature T,

$$f_0(E, E_F, T) = \frac{1}{e^{(E-E_F)/k_BT} + 1}. \tag{3.25}$$

3.4.2. *Electron and hole densities in equilibrium*

The electron and hole densities can be derived from Eqs. 3.21–3.22 using f_0. Provided that the density of states can be expressed in terms of energy, it is usually more convenient to do the calculations over electron states in terms of energy. Then the number density $n(E)$ of electrons with energy in the range E to E + dE in a system with density of states g(E) is given by

$$n(E)dE = g(E)f_0(E, E_F, T)dE \tag{3.26}$$

and the total density of electrons in a conduction band of minimum energy E_c by

$$n = \int_{E_c}^{\infty} g_c(E) f_0(E, E_F, T) dE \tag{3.27}$$

where the upper limit of integration has been extended to infinity.

The density of holes in a valence band of maximum energy E_v is given by

$$p = \int_{-\infty}^{E_v} g_v(E)(1 - f_0(E, E_F, T)) dE \,. \tag{3.28}$$

Figure 3.8 illustrates the Fermi Dirac distribution function and its effect on the energy distribution of electrons and holes in a semiconductor with roughly parabolic bands.

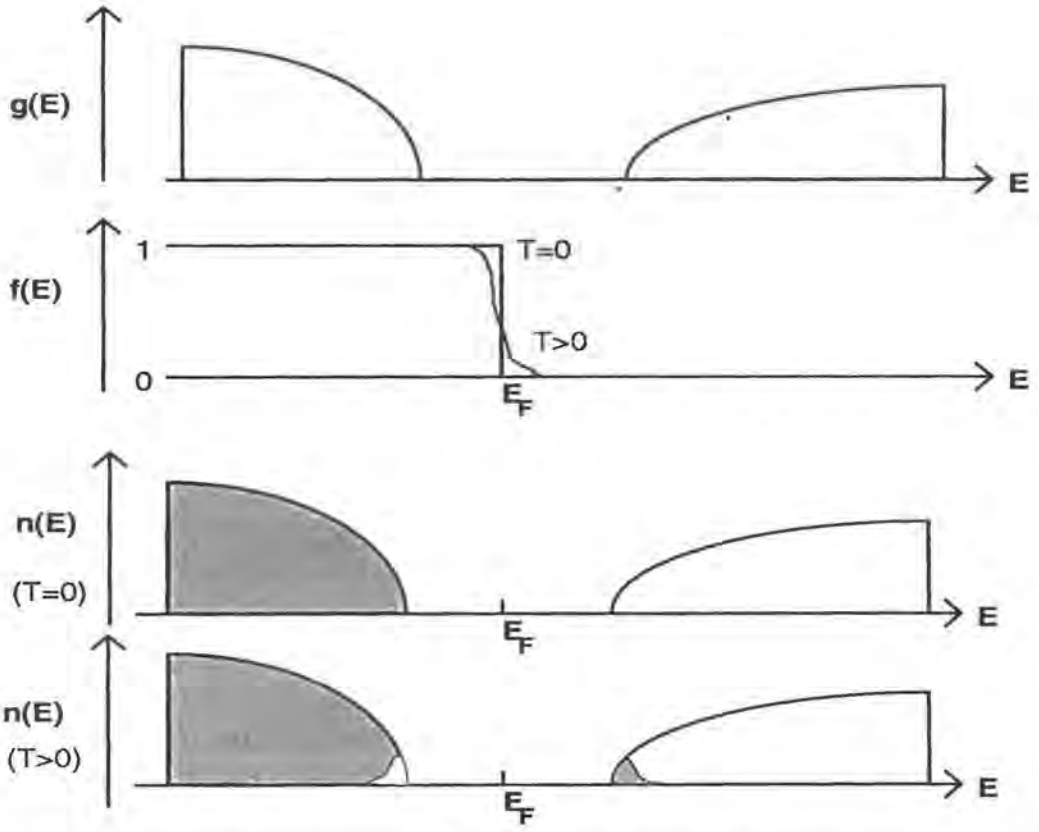


Fig. 3.8. Density of states function $g(E)$ of a semiconductor with parabolic bands, Fermi Dirac distribution function, $f(E)$, and energy distribution, $n(E)$, of electrons (shaded area) at $T = 0$ and finite temperature.

3.4.3. *Boltzmann approximation*

For a semiconductor at absolute zero, the valence band is completely filled and the conduction band completely empty. This means that the Fermi level must lie somewhere in the band gap. Now, if E_F is far enough from both band edges, f_0 can be well approximated by the Maxwell–Boltzmann form. In the conduction band when $E \gg E_F$,

$$f_0(E, E_F, T) \approx e^{(E_F - E)/k_B T} \tag{3.29}$$

and in the valence band where $E \ll E_F$,

$$1 - f_0(E, E_F, T) \approx e^{(E - E_F)/k_B T}. \qquad E \ll E_F \tag{3.30}$$

These approximations simplify the integrals in the expressions for n and p above (Eqs. 3.27 and 3.28). In the parabolic band approximation, using Eqs. 3.16 and 3.17 for $g(E)$, n and p can be evaluated exactly. We find for the electron density,

$$n = N_c \exp((E_F - E_c)/k_B T) \tag{3.31}$$

where the constant N_c is called the effective conduction band density of states and is given by

$$N_c = 2\left(\frac{m_c^* k_B T}{2\pi\hbar^2}\right)^{3/2}. \tag{3.32}$$

Similarly for the hole density,

$$p = N_v \exp((E_v - E_F)/k_B T) \tag{3.33}$$

where the effective valence band density of states is

$$N_v = 2\left(\frac{m_v^* k_B T}{2\pi\hbar^2}\right)^{3/2}. \tag{3.34}$$

Equations 3.31 and 3.33 say, very simply, that the electron and hole densities vary exponentially with the position of the Fermi level in the band gap. Moreover, they tell us that for any given material at a given temperature the product np is a constant,

$$np = N_c N_v e^{-E_g/k_B T}. \tag{3.35}$$

This relationship is used to define a constant property of the material, the *intrinsic carrier density* n_i

$$n_i^2 = np = N_c N_v e^{-E_g/k_B T} \tag{3.36}$$

n_i is equal to the density of electrons which are thermally excited into the CB of a pure semiconductor in thermal equilibrium, and is necessarily also equal to the number of holes which have been left behind.

It is sometimes convenient to write n and p in terms of n_i and an *intrinsic potential energy* E_i,

$$n = n_i \exp((E_F - E_i)/k_B T) \tag{3.37}$$

$$p = n_i \exp((E_i - E_F)/k_B T) \tag{3.38}$$

where

$$\begin{aligned} E_i &= \frac{1}{2}(E_c + E_v) - \frac{1}{2}kT \ln\left(\frac{N_c}{N_v}\right) \\ &= \frac{1}{2}(E_c + E_v) - \frac{3}{4}k_B T \ln\left(\frac{m_c^*}{m_v^*}\right). \end{aligned} \tag{3.39}$$

E_i is an energy close to the centre of the band gap, and is equal to the Fermi level of the pure semiconductor in equilibrium.

Another physically important quantity is the electron affinity of the semiconductor, χ. χ is the least amount of energy required to remove an electron from the solid. It is measured from the conduction band edge to the vacuum level E_{vac}. E_{vac} is the energy to which an electron must be raised to be free of all forces from the solid. Any spatial variations in E_{vac} are due to electrostatic fields. The conduction and valence band edges and

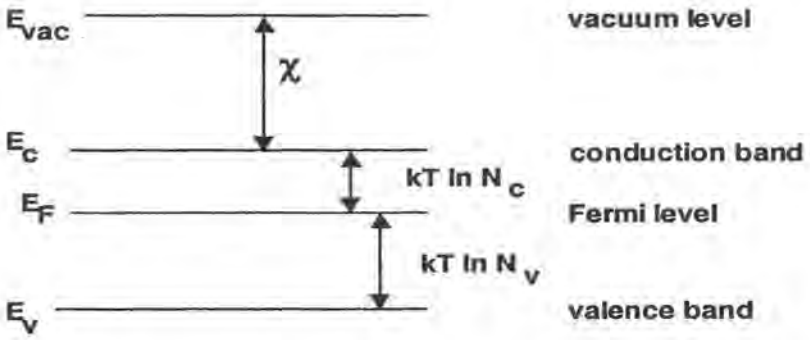


Fig. 3.9. Position of Fermi level and vacuum level on the energy band diagram.

the Fermi level can be related to E_{vac} as follows:

$$\begin{aligned} E_{\mathrm{c}} &= E_{\mathrm{vac}} - \chi \\ E_{\mathrm{v}} &= E_{\mathrm{vac}} - \chi - E_{\mathrm{g}} \\ E_{\mathrm{F}} &= E_{\mathrm{vac}} - \chi - k_{\mathrm{B}}T \ln N_{\mathrm{c}} = E_{\mathrm{vac}} - \chi - E_{\mathrm{g}} + k_{\mathrm{B}}T \ln N_{\mathrm{v}} \,. \end{aligned} \tag{3.40}$$

Validity of Boltzmann approximation

The above approximations for n and p hold so long as

$$\frac{E_{\mathrm{c}} - E_{\mathrm{F}}}{k_{\mathrm{B}}T} \gg 1 \,,$$

and

$$\frac{E_{\mathrm{F}} - E_{\mathrm{v}}}{k_{\mathrm{B}}T} \gg 1 \,,$$

conditions which are usually true for a semiconductor in equilibrium. If the parabolic band approximation is not appropriate, then the form of N_{c} and N_{v} will be different, but the energy dependence of n and p will be the same as above. n and p diverge from the simple exponential form in conditions where the Fermi level approaches the lowest CB or highest VB level. This can happen *if* the semiconductor is highly doped (see below) or if there is a high density of defect states in the band gap. Then the semiconductor is said to be *degenerate*, and the full form of the Fermi Dirac distribution, Eq. 3.24, must be used to evaluate n and p.

3.4.4. *Electron and hole currents in equilibrium*

To find J_{n} and J_{p} in equilibrium, we need to evaluate Eqs. 3.22 and 3.23 with $f(\mathbf{k},\mathbf{r}) = f_0(E(\mathbf{k}), E_{\mathrm{F}}, T)$. Here the integral must be done over $\mathbf{k}$ states since $\mathbf{k}$ occurs explicitly in the integrand. However we can take advantage of what we know about the dependence of E on $\mathbf{k}$. Since, in the parabolic band approximation, E is an *even* function of $\mathbf{k}$, $g(E(\mathbf{k}))$ and $f_0(E(\mathbf{k}), E_{\mathrm{F}}, T)$ are also even functions of k. That means that the integrand, $\mathbf{k}g(E(\mathbf{k}))f_0(E(\mathbf{k}), E_{\mathrm{F}}, T)$ will be an *odd* function of k, and so the integral when taken over all $\mathbf{k}$ states will vanish. Therefore $J_{\mathrm{n}} = J_{\mathrm{p}} = 0$; *i.e.*, no net current flows in the semiconductor in equilibrium. That is exactly what we would expect.

From the above argument we can also see that $f(\mathbf{k}, \mathbf{r})$ must be asymmetric in $\mathbf{k}$ if the electron and hole currents are not to vanish.

3.5. Impurities and Doping

3.5.1. *Intrinsic semiconductors*

The semiconductor described so far is *intrinsic* — it is a perfect crystal containing no impurities. The only energy levels permitted are the levels of well defined $\mathbf{k}$ which arise from the overlap of the atomic orbitals into crystal bands, as described above. The properties of those bands determine the position of the Fermi level in equilibrium, E_i, and the density of carriers, n_i, which have enough thermal energy to cross the band gap and conduct electricity. They also determine conductivity since the intrinsic semiconductor has n_i electrons and n_i holes available for conduction and conductivity σ is given by

$$\sigma = q\mu_\mathrm{n} n + q\mu_\mathrm{p} p \tag{3.41}$$

where μ_n, and μ_p are the mobilities of the electrons and holes, respectively, and q is the electronic charge. In equilibrium, n and p are replaced with n_i in Eq. 3.41.

At room temperature the conductivity of an intrinsic semiconductor is generally very small. For example for intrinsic silicon, $n_\mathrm{i} = 1.02 \times 10^{10}$ cm^{-3} and $\sigma = 3 \times 10^{-6}$ Ohm^{-1} cm^{-1}, at 300 K. Conductivity increases with increasing temperature, and with decreasing band gap since n_i varies as $e^{-\mathrm{E_g}/2\mathrm{k_B T}}$. For example, germanium with a band gap of 0.74 eV has a conductivity of 2×10^{-2} Ohm^{-1} cm^{-1} while gallium arsenide with a band gap of 1.42 eV has a conductivity of 10^{-8} Ohm^{-1} cm^{-1}, many orders of magnitude smaller.

Now suppose that the crystal is altered by introducing an impurity atom or a structural defect. The impurity or defect introduces bonds of different strength to those which make up the perfect crystal, and therefore changes the local distribution of electronic energy levels. The altered energy levels are localised (unless the density of defects is very high), unlike the extended $\mathbf{k}$ states which make up the intrinsic conduction and valence band states.

If the impurity energy levels occur within the band gap they can affect the electronic properties of the semiconductor. Introducing occupied impurity levels above E_i increases the Fermi level, which increases the density of electrons relative to holes in equilibrium (as is clear from Eqs. 3.37 and

3.38). In the same way, unoccupied levels below E_i reduce the Fermi level and increase the density of holes relative to electrons. The density and the nature of the carriers in the semiconductor can thus be controlled by adding definite amounts of impurities with energy levels close to the conduction or valence band edge. This is called *doping*.

3.5.2. *n type doping*

A semiconductor which has been doped to increase the density of electrons relative to holes is called n type. (The principal charge carriers are negative.) Occupied levels between E_i and E_c are introduced by replacing some of the atoms in the crystal lattice with impurity atoms which possess one too many valence electrons for the number of crystal bonds. Such impurities are called *donor* atoms because they donate an extra electron to the lattice. An example would be an atom of phosphorus, which has five valence electrons, in the tetravalent silicon lattice. The extra electron is not needed in the strong directional covalent bonds which hold each atom to its four neighbours. Therefore it is much less well bound than the other valence electrons, and is instead held rather loosely in a Coulombic bound state with the phosphorus atom. The donor can be ionised relatively easily, leaving the extra electron free to move and the fixed donor atom positively charged. For typical donors, the ionisation energy, V_n, is some meV or tens of meV. Assuming a hydrogenic bound state, the ionisation energy is approximated by

$$V_n = \frac{m_c^* \varepsilon_0^2}{m_0 \varepsilon_s^2} Ryd$$

where ε_s is the relative dielectric constant of the material and Ryd is the Rydberg energy, 13.6 eV. For typical crystalline semiconductors ε_s is of the order of 10 and m_c^* less than one, meaning V_n will be reduced by a factor of 100 compared to the Rydberg. Provided that the impurity is chosen so that V_n is small enough, virtually all the donor atoms will be ionised at room temperature.

In the band picture, the donor electrons have been promoted from a level in the band gap, V_n below E_c, into the conduction band. Since the donor states are all filled at $T = 0$, the Fermi level must now lie between the donor level and E_c.

The density of carriers can be controlled by varying the density of dopants, N_d. If $N_d \gg n_i$ and the donors are fully ionised at room

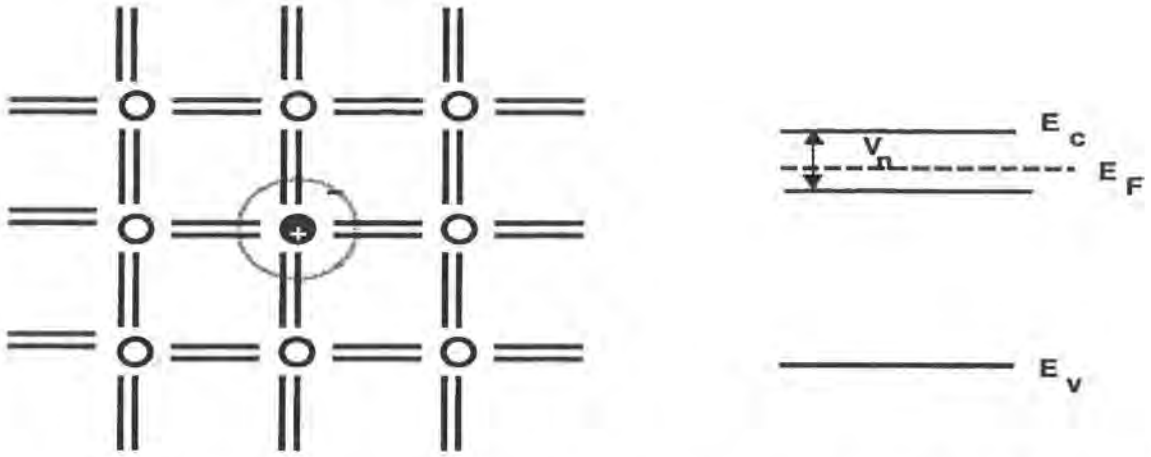


Fig. 3.10. Schematic illustration of an n type impurity in the bond (left) and band (right) picture. The extra electron is loosely bound to the donor atom. When the ionisation energy of V_n is supplied, the extra electron is freed from the donor, and the donor atom becomes positively charged. In the band picture, the electron is raised through V_n into the conduction band.

temperature, then

$$n \approx N_d \tag{3.42}$$

and

$$p = n_i^2 / N_d \,, \tag{3.43}$$

since $np = n_i^2$ at equilibrium. The electron density is now greatly increased over its equilibrium value, while the hole density is greatly reduced. The electrons are here called the *majority* carriers and the holes the *minority* carriers. Relative to the intrinsic case, the total density of carriers is increased, and so therefore is the conductivity. Conduction in n type semiconductors is mainly by electrons.

3.5.3. *p type doping*

A semiconductor which is doped to increase the density of positive charge carriers relative to negative is called p type. It is produced by replacing some of the atoms in the crystal with *acceptor* impurity atoms, which contribute too few valence electrons for the bonds they need to form. An example would be the trivalent element boron in silicon. The acceptor becomes ionised by removing a valence electron from another bond to complete the bonding between it and its four neighbours. This releases a hole into the valence

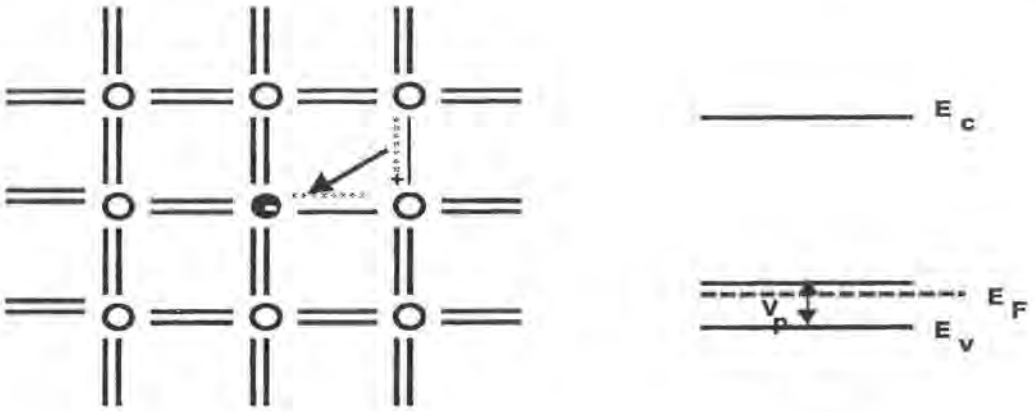


Fig. 3.11. Schematic illustration of a p type impurity in the bond (left) and band (right) picture. The acceptor atom has one too few electrons, and becomes ionised by grabbing another electron from a neighbouring bond. The grabbed electron is more tightly bound to the impurity than to a normal crystal atom. The hole created is relatively free to travel from site to site. In the band picture, the hole is lowered through V_p into the valence band.

band. The energy of this ionised state is higher, by a small amount of energy V_p, than the highest energy of a valence band electron, and so the acceptor energy level appears in the band gap close to the valence band edge. The Fermi level now lies between the acceptor level and E_v.

These p type semiconductors contain an excess of positive carriers, holes. For a doping density of $N_\mathrm{a} \gg n_\mathrm{i}$ ionisable acceptors, we find

$$p \approx N_\mathrm{a} \tag{3.44}$$

and

$$n = n_\mathrm{i}^2 / N_\mathrm{a} \,. \tag{3.45}$$

Again the total carrier density and hence the conductivity is greatly increased, but now conduction is mainly by holes, which are the majority carriers in this case.

Notice from Eqs. 3.37, 3.38 and 3.42–3.45 how the changes in n and p are consistent with the changes in the Fermi level. In an intrinsic (*i.e.*, pure) semiconductor E_F lies roughly in the middle of the band gap. The effect of doping is to shift the Fermi level away from the centre of the band gap, towards E_c in n type material, and towards E_v in p type material.

Note also that, unlike heating and illumination, doping is a way of increasing the conductivity of the semiconductor at equilibrium without requiring a constant input of energy.

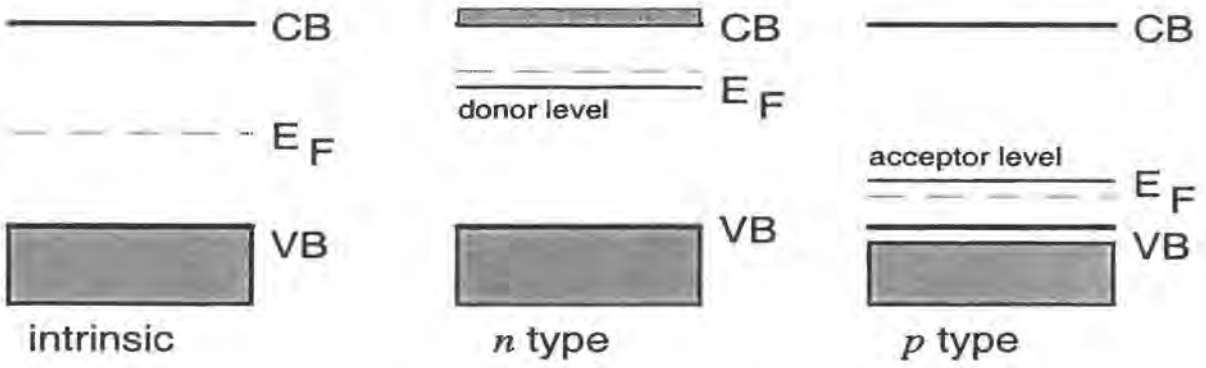


Fig. 3.12. Position of Fermi level and impurity levels relative to the conduction and valence band edges. Shaded regions indicate levels filled by electrons.

3.5.4. *Effects of heavy doping*

Heavy doping causes imperfections in the crystal structure which appear as states in the band gap close to the CB and VB edges. These states have the effect of adding a tail to the CB and VB density of states functions, and so effectively reducing the band gap. The intrinsic carrier density will increase, because thermal excitation across the band gap is easier, and this has consequences for solar cell performance. Heavy doping is also likely to increase the density of defect states which act as centres for carrier recombination or trapping.

3.5.5. *Imperfect and amorphous crystals*

As well as extrinsic impurities, semiconductors may contain intrinsic defects which introduce levels into the band gap and modify the electronic properties. Even in a perfect monocrystalline semiconductor, configurational defects arise from broken bonds and structural rearrangement at surfaces. In a polycrystalline semiconductor, made up of grains of the same crystal at various orientations, defects occur at grain boundaries. These defects may cause an otherwise intrinsic material to become slightly n or p type, by self doping. They can also be responsible for charged boundary layers near to surface or grain boundaries in deliberately doped materials, which impede carrier transport.

There is another important group of semiconductor materials, amorphous semiconductors. These have a similar structure to the crystal on the short range, but have no long range order. For instance, in amorphous silicon, while each silicon atom is bonded to four neighbours with approximate

tetrahedral symmetry (just as in crystalline silicon) the arrangement of silicon atoms four or five bond lengths away appears almost random. These small deviations in bond length and orientation give rise to a spread in energy levels so that the density of states near the conduction band minimum and valence band maximum tends to decay with energy into the band gap, in an Urbach tail, rather than cutting off abruptly. Amorphous semiconductors are also very likely to contain defects due to voids and broken bonds. Intrinsic defects and amorphous semiconductors will be discussed in Chapter 8.

A very high density of defects in the band gap is undesirable, as these tend to act as centres for charge recombination (see Chapter 4). It also gives rise to an effect called Fermi level pinning, where the density of intra band gap states is too high for all the states to be ionised at room temperature, and so the Fermi level is fixed amongst them. This makes it difficult to dope the semiconductor in the usual way. Fermi level pinning can be a problem at interfaces and is mentioned in Chapter 5.

3.6. Semiconductor under Bias

So far we have considered the semiconductor at equilibrium. One feature of the material at equilibrium is that no current flows. To understand the operation of a solar cell, we need to understand what happens to the semiconductor when it is displaced from equilibrium by exposure to light. Since current flows in the operating solar cell, by definition the device cannot be at equilibrium.

3.6.1. *Quasi thermal equilibrium*

Suppose the semiconductor is disturbed from equilibrium by some influence which changes the population of electrons and holes. This could be exposure to light of energy $E > E_g$ which increases the density of both electrons and holes above their equilibrium values, or it could be the electrical injection of electrons and holes through an applied electric bias. In either case, n and p have been disturbed so that they are no longer described by the Fermi Dirac equilibrium distribution function and $np = n_i^2$ is no longer true.

The electron distribution is now governed by the *general* distribution function $f(\mathbf{k}, \mathbf{r})$ introduced in Eq. 3.20 above. Moreover, different distribution functions, $f_c(\mathbf{k}, \mathbf{r})$ and $f_v(\mathbf{k}, \mathbf{r})$, should apply for electrons in conduction and valence bands. f_c and f_v should be position dependent, because

all points in the semiconductor are no longer in equilibrium, and also $\mathbf{k}$ dependent because the applied bias may be directional and may favour occupation of some $\mathbf{k}$ states rather than others of the same E.

Now — and here comes one of the great simplifications of semiconductor physics — if the disturbance is not too great, or not changing too quickly, the populations of electrons and holes each relax to achieve a state of *quasi thermal equilibrium*. What that means is that the population of electrons within the conduction band distribute themselves as though they were at equilibrium, with a common Fermi level and temperature, and the holes within the valence band arrange themselves as though they shared another, different Fermi level and possibly a different temperature. In quasi thermal equilibrium,

$$f_{\mathrm{c}}(\mathbf{k},\mathbf{r}) \approx f_0(E, E_{\mathrm{F_n}}, T_{\mathrm{n}})$$

and

$$f_{\mathrm{v}}(\mathbf{k},\mathbf{r}) \approx f_0(E, E_{\mathrm{Fp}}, T_{\mathrm{p}})\,.$$

The new apparent Fermi levels for electrons and holes are called the electron and hole *quasi Fermi levels*, $E_{\mathrm{F_n}}$ and E_{Fp}, respectively.

The approximation is possible because relaxation *within* each band is so much faster than relaxation *between* the bands. (Carriers relax within the bands mainly by scattering from the lattice, with the emission and absorption of phonons. This occurs on a time scale of 10^{-12}–10^{-15} s. Between bands, the carriers relax only by interacting with another carrier, with the emission of a photon or with a deep trap. Relaxation over these energies is much less likely because of the scarcity of high energy phonons and photons under normal conditions, and time scales are longer, typically 10^{-6}–10^{-9} s.) It is possible for a disturbance — such as optical generation — to be fast on the timescale of relaxation between bands, so that it continuously disturbs the system from equilibrium, but slow on the timescale of relaxation within bands.

Considering only the conduction band for the moment, $f_{\mathrm{c}}(k,r)$ cannot be *completely* described by the quasi thermal equilibrium distribution however; since, as we have seen above, f_0 is symmetric in $\mathbf{k}$, implying that no current flows. To deal with this, a small additional term, f_{A}, which is *antisymmetric* in $\mathbf{k}$, is added to f_0. The approximation is written,

$$f_{\mathrm{c}}(\mathbf{k},\mathbf{r}) = f_0(E, E_{\mathrm{F_n}}, T_{\mathrm{n}}) + f_{\mathrm{A}}(\mathbf{k},\mathbf{r}) \qquad (3.46)$$

where

$$f_0(E, E_{F_n}, T_n) = \frac{1}{e^{(E-E_{F_n})/k_B T_n} + 1} . \tag{3.47}$$

f_A can be found by setting up and solving a book-keeping equation for f called the Boltzmann Transport equation (see Box 3.4). Note that because of the well defined parity of f_0 and f_A, only f_0 contributes to the expressions for n and p, while only f_A contributes to J_n and J_p.

3.6.2. *Electron and hole densities under bias*

Assuming quasi thermal equilibrium, the electron and hole densities in a semiconductor under bias are given, using Eqs. 3.37 and 3.38, by

$$n = n_i e^{(E_{F_n}-E_i)/k_B T_n} \tag{3.48}$$

and

$$p = n_i e^{(E_i-E_{F_p})/k_B T_p} \tag{3.49}$$

or, in terms of the conduction and valence band energies using Eqs. 3.31 and 3.34, by

$$n = N_c e^{-(E_c-E_{F_n})/k_B T_n} \tag{3.50}$$

and

$$p = N_v e^{-(E_{F_p}-E_v)/k_B T_p} , \tag{3.51}$$

where T_n and T_p are the electron and hole effective temperatures. In principle, T_n and T_p may be different from the ambient temperature T. Electrons with $T_n > T$ are called 'hot' electrons and can arise in situations where the carriers gain excess kinetic energy from strong electric fields. However, the conditions giving rise to hot carriers do not usually apply in photovoltaic devices, and henceforth we will assume that $T_n = T_p = T$. (We will come back to hot electrons in Chapter 10.) It is assumed in Eqs. 3.48–3.51 that $E_c - E_{F_n} \gg k_B T$ and $E_{Fp} - E_v \gg k_B T$, so that Boltzmann statistics apply.

Equations 3.48 and 3.49 are the same as Eqs. 3.37 and 3.38 except that $E_{F_n} \neq E_{Fp} \neq E_F$. We say that the Fermi levels are *split* (see Fig. 3.13). The difference in quasi Fermi levels,

$$\Delta\mu = E_{F_n} - E_{Fp} \tag{3.52}$$

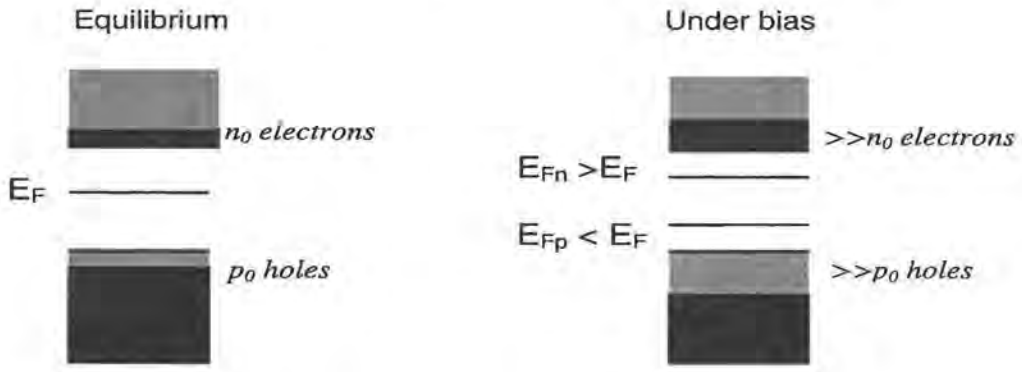


Fig. 3.13. Fermi levels at and away from equilibrium.

is the difference in chemical potentials referred to in Chapter 2, and the size of the splitting, $\Delta\mu$, depends upon the intensity of the disturbance. The electron hole product is now given by

$$np = n_i^2 e^{\Delta\mu/k_B T} . \tag{3.53}$$

In general, E_{F_n} and E_{F_p} are functions of position. At any point within the semiconductor we will suppose there is a *local* quasi thermal equilibrium and we can define *local* quasi Fermi levels.

3.6.3. *Current densities under bias*

To find J_n and J_p we need to find an approximation for the antisymmetric part of f. This is done by considering the different influences on f.

Box 3.4. Boltzmann Transport equation and the relaxation time approximation

In this box we derive a form for f_A based on a consideration of the processes by which carriers are removed from any particular $\mathbf{k}$ state.

Since f_c is a function of $\mathbf{k}$, $\mathbf{r}$ and t, the rate of change of f is given by

$$\frac{df_c}{dt} = \frac{d\mathbf{r}}{dt} \cdot \nabla_r f_c + \frac{d\mathbf{k}}{dt} \cdot \nabla_k f_c + \frac{\partial f_c}{\partial t} . \tag{3.54}$$

The terms in this equation represent the various ways in which f_c can change. In the following we will use the definitions of velocity

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} , \tag{3.55}$$

and force

$$\mathbf{F} = \hbar \frac{d\mathbf{k}}{dt} . \tag{3.56}$$

To find the derivatives of f with respect to $\mathbf{r}$ and $\mathbf{k}$ we will approximate f_c by f_0 and refer the electron energy to the conduction band edge,

$$E = E_c + E(\mathbf{k}, \mathbf{r}) . \tag{3.57}$$

(Here we have assumed, for clarity, that the conduction band minimum is at $\mathbf{k}_0 = 0$. The result is the same for $\mathbf{k}_0 \neq 0$.) Then

$$\nabla_\mathbf{r} f_c \approx -\frac{f_0}{kT} \nabla_\mathbf{r} (E_c - E_{F_n}) \tag{3.58}$$

and

$$\nabla_\mathbf{k} f_c \approx -\frac{f_0}{kT} \nabla_\mathbf{k} E = -\frac{\hbar f_0}{k_B T} \mathbf{v} \tag{3.59}$$

where we have used the definition of group velocity. (We have also assumed that there is no gradient in temperature, as discussed above. In the most general case $\nabla T \neq 0$ and the resulting expression for current contains a term driven by temperature gradients.) Substituting back into Eq. 3.54 we obtain

$$\begin{aligned} \frac{df_c}{dt} &= -\frac{f_0}{k_B T} (\mathbf{v} \cdot \nabla_\mathbf{r} (E_c - E_{F_n}) + \mathbf{F} \cdot \mathbf{v}) + \frac{\partial f_c}{\partial t} \\ &= \frac{f_0}{k_B T} \mathbf{v} \cdot \nabla_\mathbf{r} E_{F_n} + \frac{\partial f_c}{\partial t} . \end{aligned} \tag{3.60}$$

where we have related $\nabla_\mathbf{r} E_c$ to the force on the electron to cancel out the first and third terms.

To solve this for f_c we need to make two assumptions about the processes which change f_c. One, relaxation events within a band within a band are much more frequent than exchanges of carriers between bands. As mentioned above, relaxation events within a band are usually collisions with the lattice and are fast because of the abundance of phonons. Therefore

$$\frac{\partial f_c}{\partial t} \approx \left. \frac{\partial f_c}{\partial t} \right|_{\text{collisions}} .$$

Two, the distribution relaxes exponentially towards the quasi equilibrium function with a characteristic time constant τ so that

$$\left.\frac{\partial f_c}{\partial t}\right|_{\text{collisions}} = -\frac{(f_c - f_0)}{\tau} \tag{3.61}$$

This is the *relaxation time approximation.* Then we can solve Eq. 3.60 in the steady state to obtain

$$f_c = f_0\left(1 - \frac{\tau \mathbf{v}}{k_B T}\cdot \nabla_r E_{F_n}\right) \tag{3.62}$$

whence

$$f_A = -f_0 \frac{\tau \mathbf{v}}{k_B T}\cdot \nabla_r E_{F_n}\,. \tag{3.63}$$

Substituting for f_c into Eq. 3.22 for J_n we find

$$J_n(\mathbf{r}) = \left\{\frac{q}{k_B T}\int_{\text{conduction band } \mathbf{k}} \tau\left(\frac{\hbar \mathbf{k}}{m_c^*}\right)^2 g_c(\mathbf{k}) f_0(E(\mathbf{k}))\mathbf{d}^3\mathbf{k}\right\} \times \nabla_r E_{F_n}\,. \tag{3.64}$$

Finally, replacing the term in brackets by $\mu_n n$, where μ_n is the electron *mobility*, and n defined from f_0 as above, (Eq. 3.21) we have

$$\boxed{J_n(\mathbf{r}) = \mu_n n \nabla_r E_{F_n}}\,. \tag{3.65}$$

So, the electron current is proportional to the gradient in the electron quasi Fermi level.

A similar treatment for the valence band distribution function f_v produces the hole current density

$$\boxed{J_p(\mathbf{r}) = \mu_p p \nabla_r E_{F_n}}\,. \tag{3.66}$$

Henceforth we will use ∇ to mean ∇_r.

Using the Boltzmann Transport equation to solve for the non-equilibrium distribution function (see Box 3.4), we obtain the following relations for the electron and hole current densities in a semiconductor under bias:

$$J_n(\mathbf{r}) = \mu_n n \nabla E_{F_n} \qquad \text{(Eq. 3.65)}$$

$$J_p(\mathbf{r}) = \mu_p p \nabla E_{F_p}\,, \qquad \text{(Eq. 3.66)}$$

that is, the current density at any point is proportional to the relevant carrier density and the gradient of the relevant quasi Fermi level. We say the current is driven by the gradient in the quasi Fermi level. In a more general case, Eqs. 3.65 and 3.66 would contain terms in ∇T_n and ∇T_p and the currents would be driven also by temperature gradients.

At any point $\mathbf{r}$ the net current is given by the sum of the electron and hole currents at that point:

$$J(\mathbf{r}) = J_n(\mathbf{r}) + J_p(\mathbf{r}) , \tag{3.67}$$

3.7. Drift and Diffusion

3.7.1. *Current equations in terms of drift and diffusion*

Equations 3.65 and 3.66 are the most compact forms for J_n and J_p. However, the physical interpretation of a quasi Fermi level gradient may be somewhat clearer if we use the results, valid within the Boltzmann approximation, that

$$\nabla E_{F_n} = (\nabla E_c - kT\nabla \ln N_c) + \frac{k_B T}{n}\nabla n \tag{3.68}$$

and

$$\nabla E_{F_p} = (\nabla E_v + kT\nabla \ln N_v) - \frac{k_B T}{p}\nabla p . \tag{3.69}$$

(These can be obtained by differentiating Eqs. 3.31 and 3.33.) The gradient in the conduction or valence band edge is provided by the electrostatic field F and the gradients in the electron affinity χ and band gap, E_g

$$\nabla E_c = qF - \nabla\chi \tag{3.70}$$

and

$$\nabla E_v = qF - \nabla\chi - \nabla E_g \tag{3.71}$$

where we have used Eq. 3.40 and the definition of electrostatic field

$$F = \frac{1}{q}\nabla E_{vac} . \tag{3.72}$$

Substituting for ∇E_{F_n} and ∇E_{F_n} into Eqs. 3.65 and 3.66 we have

$$J_n(\mathbf{r}) = qD_n\nabla n + \mu_n n(qF - \nabla\chi - k_B T\nabla \ln N_c) . \tag{3.73}$$

and

$$J_{\rm p}(\mathbf{r}) = -qD_{\rm p}\nabla p + \mu_{\rm p}p(qF - \nabla\chi - \nabla E_{\rm g} + k_{\rm B}T\nabla \ln N_{\rm v})\,. \qquad (3.74)$$

Thus the gradients in the electron affinity, band gap and effective band densities of states provide an additional *effective* electric field to the electrostatic field F.

In compositionally invariant material, only the electrostatic field F is present and so

$$J_{\rm n}(\mathbf{r}) = qD_{\rm n}\nabla n + q\mu_{\rm n}Fn \qquad (3.75)$$

and

$$J_{\rm p}(\mathbf{r}) = -qD_{\rm n}\nabla p + q\mu_{\rm p}Fp\,. \qquad (3.76)$$

Above we have used the 'Einstein' relations which relate mobility to the diffusion constant for either carrier, and are valid under low field conditions

$$\mu_{\rm n} = \frac{qD_{\rm n}}{k_{\rm B}T}, \qquad \mu_{\rm p} = \frac{qD_{\rm p}}{k_{\rm B}T}\,. \qquad (3.77)$$

Now, $J_{\rm n}$ and $J_{\rm p}$ each resolve into two contributions: a *drift* current (sometimes called migration) where carriers are driven by an electric field, and a *diffusion* current where they are driven by a concentration gradient. In the first case the carriers move so as to minimise their electrostatic potential energy, and in the second to minimise their statistical potential energy.

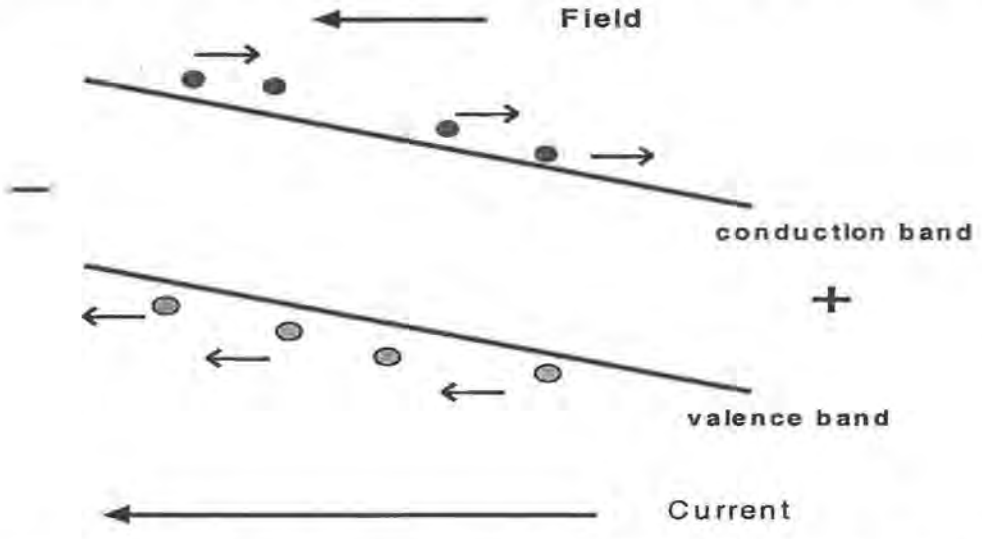


Fig. 3.14. Drift. Carriers flow under an *electric field* in order to reduce their electrical *potential* energy. Notice that electrons and holes drift in opposite directions, both producing currents in the direction of the field.

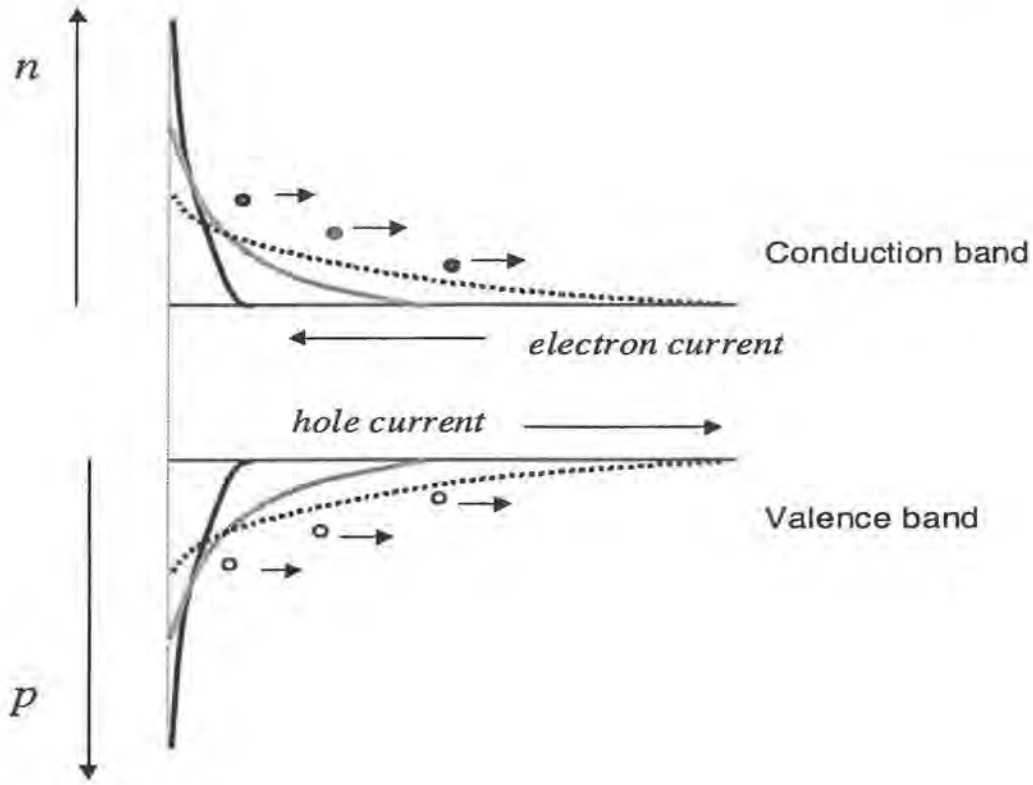


Fig. 3.15. In diffusion, the electron and hole populations spread out under a concentration gradient to reduce their statistical potential energy. The grey and dotted lines indicate how the electron and hole density may develop after generation close to the left hand surface. Note that the currents produced by electron and hole diffusion have opposite signs. Unless the electron and hole mobilities are different, the net current may be zero.

In the presence of a drift field only, the total current is

$$J = J_{\mathrm{n}} + J_{\mathrm{p}} = q(\mu_{\mathrm{n}} n + \mu_{\mathrm{p}} p)F \tag{3.78}$$

where the coefficient of F is the conductivity of the semiconductor, as given in Eq. 3.41. Note that the field drives electrons and holes in opposite directions.

In the presence of a concentration gradients only we have,

$$J_{\mathrm{n}} = q(D_{\mathrm{n}} \nabla n - D_{\mathrm{p}} \nabla p) \tag{3.79}$$

If the electron and hole gradients are similar, as may be the case under illumination, the electron and hole diffusion currents tend to cancel each other out. A *net* diffusion current usually arises only when the electron and hole carrier gradients are very different. This can be achieved in certain device configurations, such as a p–n junction. In that case, currents are

dominated by minority carrier diffusion. This will be discussed in detail in Chapter 6.

3.7.2. *Validity of the drift-diffusion equations*

The drift-diffusion current equations in the form 3.75 and 3.76 are very commonly used for the electron and hole currents in device physics. However, they depend upon several assumptions. It is worth reviewing these here, and considering some exceptions.

Assumptions used:

- Electron and hole populations each form a quasi thermal equilibrium with a characteristic Fermi level and temperature

This is one of the fundamental tenets of semiconductor physics. It would not be valid if for instance, the bands were too narrow, too close together, or if there was too high a density of intra band gap states. In those conditions we might not really have a semiconductor!

- Electron and hole temperatures are always the same as the lattice temperature

This fails in conditions where electrons or holes gain extra kinetic energy; for example, through acceleration through large electric fields. Then the hot electrons and holes need to be described by spatially varying temperatures T_n and T_p, and a term proportional to the temperature gradient appears in the electron and hole currents. Such high fields are not usually encountered in photovoltaic devices. However, it is theoretically possible that hot carrier effects may arise through quantum confinement in low dimensional structures. This will be raised in Chapter 10.

- Changes in state occupancy are much more likely to be due to scattering collisions within a band than to generation, recombination or trapping events which remove a carrier from that band. This is the relaxation time approximation.

This may fail in materials containing a high density of defects, where trapping, detrapping and recombination events are common. Then the relaxation time approximation is not valid. Another special case arises at an abrupt boundary between two different semiconductors (a heterojunction).

- Electron and hole states can be described by a quantum number **k**

This is used in the definition of the electron and hole currents above. It fails for non-crystalline materials, with the consequence that carrier mobilities cannot be rigorously defined from the expectation value of the scattering time. Nevertheless, analogous expressions for electron and hole currents can be used in amorphous and defective materials, but with empirical expressions for the mobilities.

- Boltzmann approximation ($E_c - E_{F_n} \gg k_B T$; $E_{F_p} - E_v \gg k_B T$)

This was made in obtaining the drift diffusion forms from the quasi Fermi level gradients. In degenerate semiconductors (where the Boltzmann approximation fails) the definitions in terms of Fermi level gradients are still valid and should be used.

- Compositional invariance

This was made in replacing the band edge gradient with F. For a compositionally varying material, contributions to the current will also arise from any variation in the electron affinity, band gap, and effective mass, and can be included by using Eqs. 3.73 and 3.74.

3.7.3. *Current equations for non-crystalline solids*

The drift diffusion picture is clearly not valid in the case of defective crystalline or amorphous materials, when the density of states in the band gap is high. Then the following effects may apply:

- the number density of electrons and holes may be differently sensitive to temperature, electric field and illumination;
- the mobility may be temperature, field and carrier density dependent;
- there may be a contribution to the current which comes from conduction between localised states in the band gap.

There are a number of different ways of describing the currents in defective materials. A thorough description is beyond the scope of this chapter. It is often assumed that the currents due to electrons and holes moving in delocalised states in the conduction and valence bands can be described by the Boltzmann transport theory, *i.e.*, drift diffusion equations apply. There may be an additional current due to electron (or hole) motion between

localised states in the band gap,

$$J = J_{\mathrm{n}} + J_{\mathrm{p}} + \sum_{\text{localised states } i} J_{\mathrm{i}} \,. \tag{3.80}$$

This localised state current can be described with the aid of an energy dependent mobility, $\mu(E)$, in analogy to the definition of current in a crystal above

$$\sum_i J_{\mathrm{i}} = \sum_i \mu(E_{\mathrm{i}}) f(E_{\mathrm{i}}) g(E_{\mathrm{i}}) \tag{3.81}$$

where $g(E_{\mathrm{i}})$ is the density of localised states and $f(E)$ is the (usually Fermi–Dirac) distribution function. The form of $\mu(E)$ depends on the mechanism of transfer between localised states. For example, for thermally assisted hopping it should be thermally activated.

Delocalised conduction or valence band states can be distinguished from localised, intra band gap states by a mobility edge. In practice, this is the energy at which the mobility becomes temperature activated. These will be discussed in more detail in Chapter 8.

3.8. Summary

Semiconductors are suitable materials for photovoltaics on account of the band gap between occupied and unoccupied bands which can be bridged by a visible photon, and the ease of charge transport through the crystal bands. In crystalline semiconductors, electrons promoted to the conduction band can be treated like nearly free particles with a well-defined wavevector and an effective mass which reflects the effect of the lattice atoms. Absent electrons in the valence band are known as holes and can be treated as positively charged particles with an effective mass determined by the valence band structure. Electrons and holes can be treated approximately as independent particles. The band structure is the relationship between the energy of the electron and its wavevector. The most important points are the valence band maximum and conduction band minimum. The form of the band structure is determined by the crystal structure but it is approximately parabolic near to these critical points.

In equilibrium, states are filled by electrons according to the Fermi Dirac distribution function. In a pure semiconductor the Fermi level lies within the band gap and the electron and hole densities depend exponentially upon E_{F}. Doping with impurities of different valence increases the density

of electrons or holes in the semiconductor at equilibrium and moves E_F towards the bands. Heavy doping fixes the density of the dominant carrier. At equilibrium, electron and holes currents are always zero.

The semiconductor can be disturbed from equilibrium by the action of light or an applied electric bias. Away from equilibrium the distribution function is changed but can be approximated by a Fermi Dirac function with separate, spatially varying Fermi levels, called quasi Fermi levels, for the electron and hole. In general, different electron and hole temperatures may apply but in practical conditions the ambient temperature is usually valid.

Application of bias produces electron and hole currents. In conditions where scattering events within bands being faster than transitions between bands (the relaxation time approximation), the currents are proportional to the gradients in the respective quasi Fermi levels and to the respective mobilities. The quasi Fermi level gradient may be considered as a driving force for conduction. By resolving the quasi Fermi level gradient into electrostatic potential and carrier density gradients, the current can be expressed as a sum of a field driven 'drift' current and a statistically driven 'diffusion' current. In non-crystalline materials the same concepts of quasi Fermi level and mobility can be applied, but the effective mobility may be energy dependent.

Chapter 4

Generation and Recombination

4.1. Introduction: Semiconductor Transport Equations

The essential function of a solar cell is the generation of photocurrent. The output is determined by a balance between light absorption, current generation and charge recombination. Currents due to electrons and holes in semiconductors were discussed in Chapter 3, and Chapter 5 will deal with the mechanisms of charge separation which drive the photocurrent. This chapter is concerned with the processes of charge carrier generation and recombination. We introduce the theoretical formalism for electronic transitions in semiconductors and use this to derive the rates of photogeneration and the principal recombination mechanisms. We will derive some of the commonly used formulae for recombination rates, and attempt to show how the rates are dependent upon the electronic structure, materials properties and operating conditions. The detailed derivations will not be relevant to all readers, but are included to suggest how the formalisms can be adapted to deal with other, novel photovoltaic materials and structures.

The basic equations of device physics, the semiconductor transport equations, are based on two simple principles: that the number of carriers of each type must be conserved; and that the electrostatic potential due to the carriers charges obeys Poisson's equation. For a semiconductor containing electrons and holes, conservation of electron number requires that

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla.\mathbf{J}_{\mathrm{n}} + G_{\mathrm{n}} - U_{\mathrm{n}} \tag{4.1}$$

for electrons and

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla.\mathbf{J}_{\mathrm{p}} + G_{\mathrm{p}} - U_{\mathrm{p}} \tag{4.2}$$

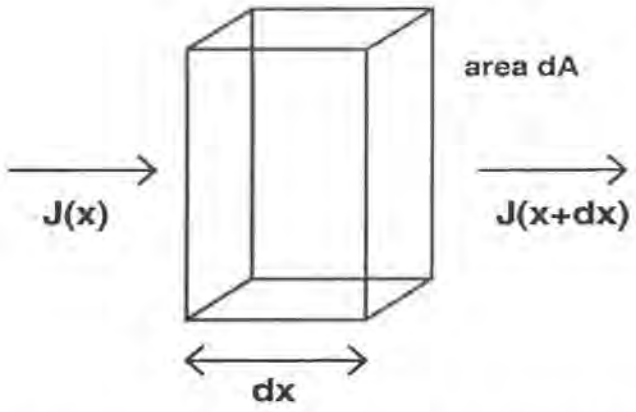


Fig. 4.1. Illustration of the continuity equations. On passing through a small volume, the charge changes by $(J(x+dx)-J(x)) \times dA$. This has to be supplied by the difference between the carrier generation and recombination rates, $(G-U) \times dx \times dA$.

for holes, where $G_{\mathrm{n/p}}$ is the volume rate of generation of electrons (holes) and $U_{\mathrm{n/p}}$ is the volume rate of recombination. Poisson's equation in the differential form is

$$\nabla^2\phi = \frac{q}{\varepsilon_{\mathrm{s}}}(-\rho_{\mathrm{fixed}} + n - p) \tag{4.3}$$

where ϕ is the electrostatic potential, ε_{s} is the dielectric permittivity of the semiconductor and ρ_{fixed} is the local density of fixed charge.

The continuity Eqs. 4.1 and 4.2 can be understood by considering a small volume $\delta\mathcal{V} = \delta A \times \delta x$ within the semiconductor (Fig. 4.1).

In unit time, $\delta\mathcal{V} \times G_{\mathrm{n}}$ electrons are generated within the volume and $\delta\mathcal{V} \times U_{\mathrm{n}}$ electrons are removed. Electrons may also be introduced or removed by currents flowing into and out of the volume. Consider only the x component of the current. In unit time $\frac{1}{q}J_{\mathrm{n}}(x+\delta x) \times \delta A$ electrons leave the volume through the right hand boundary while $\frac{1}{q}J_{\mathrm{n}}(x) \times \delta A$ enter from the left, causing a net change of $\frac{1}{q}\frac{\delta J_{\mathrm{n}}}{\delta x} \times \delta v$ in the number of electrons. Adding the three contributions we have for the rate of change of electron density per unit volume,

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_{\mathrm{n}}}{\partial x} + G_{\mathrm{n}} - U_{\mathrm{n}}\,.$$

In three dimensions this generalises quite readily to the form given in (4.1) above.

Equations (4.1) and (4.2) are completely general. Once we know how the J, G and U depend on n, p and ϕ, and other material or environment

parameters, we have a set a set of coupled differential equations which can be solved for the three unknowns. The particular form of J, G and U depends on the material and environment. In Chapter 3 we looked at how J and n are defined for a crystalline semiconductor. In this chapter we will look at the mechanisms which determine U and G.

4.2. Generation and Recombination

Generation is an electronic excitation event which increases the number of free carriers available to carry charge. Recombination is an electronic relaxation event which reduces the number of free carriers. Generation requires an input of energy while recombination releases energy. The energy input can be provided by the vibrational energy of the lattice (phonons), light (photons) or the kinetic energy of another carrier. The released energy is taken up by these same mechanisms. For every generation process there is an equivalent recombination process. This is due to microscopic reversibility, and is an important principle in understanding the function of photovoltaic devices.

Generation may be the promotion of an electron from valence to conduction band, which creates an electron-hole pair, or from valence band into a localised state in the band gap, which generates only a hole, or from a localised state into the conduction band, which generates only an electron. For the solar cell, the most important form of generation is optical, *i.e.*, by the absorption of a photon.

Recombination is the loss of an electron or hole through the decay of an electron to a lower energy state. Again this may be from band to band, destroying an electron-hole pair, or it may be from conduction band to trap state or from trap state to valence band, removing only an electron or a hole, respectively. The energy released can be given up as a photon (*radiative* recombination), as heat through phonon emission (*non-radiative* recombination) or as kinetic energy to another free carrier (*Auger* recombination).

Thermal generation and recombination

At absolute zero in the absence of any external bias, electrons occupy all of the energy levels available up to the Fermi level, and no recombination or generation processes occur. As the temperature is raised, the lattice gains vibrational kinetic energy and some of this may be given up to an electron

to promote it to a higher energy level. The promotion of an electron across the band gap is is called *thermal generation*. At the same time, electrons in excited states can relax down to vacant lower energy states and give up the energy difference as vibrational energy to the lattice. The loss of a mobile carrier in this way is *thermal recombination*. Like any other generation-recombination processes, these thermal processes can involve band to band transitions and localised state to band transitions. At finite temperatures these processes occur continually at a rate which increases with increasing temperature. In thermal equilibrium, the rate of every thermal generation process is matched exactly by the rate of the equivalent thermal recombination events. Thus the thermal generation $G^{\rm th}_{\rm n/p}$ is balanced by the equilibrium recombination rate $U^{\rm th}_{\rm n/p}$,

$$G^{\rm th}_{\rm n} = U^{\rm th}_{\rm n}$$

and

$$G^{\rm th}_{\rm p} = U^{\rm th}_{\rm p} .$$

Since we are interested in the *disturbance* of the populations from equilibrium, we need consider only the excess recombination and generation rates in Eq. (4.1), *i.e.*,

$$U_{\rm n} = U^{\rm total}_{\rm n} - U^{\rm th}_{\rm n} \tag{4.4}$$

and

$$G_{\rm n} = G^{\rm total}_{\rm n} - G^{\rm th}_{\rm n} .$$

Thus thermal generation is not included explicitly as a contribution to G.

For band-to-band generation and recombination processes, the rates are equal for electrons and holes and

$$U_{\rm n} = U_{\rm p} = U \tag{4.5}$$

and

$$G_{\rm n} = G_{\rm p} = G .$$

In this chapter we will first consider how electronic transition rates may be treated microscopically using Fermi's Golden Rule, in Sec. 4.3. In Sec. 4.4 we will address the process of photogeneration, and show how the absorption coefficient of a semiconductor is derived. In Sec. 4.5 we consider the different types of recombination: radiative, Auger, nonradiative via traps,

and surface recombination, and show how the recombination rates are related to the material properties. Section 4.6 addresses the semiconductor transport equations again and indicates some simplified cases.

4.3. Quantum Mechanical Description of Transition Rates

4.3.1. *Fermi's Golden Rule*

In many cases electronic transitions can be described by Fermi's Golden Rule. This is an approximation, based on first order perturbation theory, of the full quantum mechanical transition rate. According to Fermi's Golden Rule, the transition *probability* per unit time from an initial filled state $|i\rangle$ of energy E_i to a final empty state $|f\rangle$ of energy E_f which differs in energy by E under the action of some perturbing Hamiltonian $\mathbf{H}'$ is given by

$$\frac{2\pi}{\hbar}|\langle i|\mathbf{H}'|f\rangle|^2\delta(E_\mathrm{f}-E_\mathrm{i}\mp E)\,. \tag{4.6}$$

The term in brackets $|\langle\rangle|^2$ is a *matrix element* coupling initial and final states. It describes the size of the interaction between the two states under the given perturbation. $\mathbf{H}'$ may represent the effect of a light field, of phonon interactions, or carrier–carrier interactions. The delta function ensures conservation of energy. A negative sign applies for excitation events, such as absorption where $E_\mathrm{f} > E_\mathrm{i}$, and a quantum of energy E is removed from the perturbing field, and a positive sign for relaxation events, such as recombination, where $E_\mathrm{f} < E_\mathrm{i}$ and a quantum E is emitted.

Box 4.1. Derivation of Fermi's Golden Rule

To obtain the transition *rate* we need to multiply by the probabilities that the initial state is occupied, $f_\mathrm{i} = f(E_\mathrm{i}(\mathbf{k}_\mathrm{i}))$ and that the final state is available, $1 - f_f = (1 - f(E_f(\mathbf{k}_\mathrm{f})))$, where $f(E)$ is the electronic occupation function. Then the rate of transitions $i \rightarrow f$ is given by

$$r_{\mathrm{i}\rightarrow\mathrm{f}} = \frac{2\pi}{\hbar}|\mathbf{H}_\mathrm{if}|^2\delta(E_\mathrm{f}-E_\mathrm{i}-E)f_\mathrm{i}(1-f_\mathrm{f}) \tag{4.7}$$

where $\mathbf{H}_\mathrm{if}$ is shorthand for the matrix element. (In Eq. 4.7 an electron promotion event ($E_\mathrm{f} > E_\mathrm{i}$, is assumed. The analysis applies equally to the case $E_\mathrm{f} < E_\mathrm{i}$, in which case E takes the opposite sign.)

Provided that the transition $i \rightarrow f$ is energy conserving, then the same perturbing field can induce the *reverse* transition $f \rightarrow i$. (Stimulated

emission in a solid is an example of this.) As soon as state f is partially occupied and state i is partially empty, there is a finite probability that this reverse transition will occur. By the symmetry of quantum mechanics the transition $f \rightarrow i$ is governed by the *same* matrix element H_{if}, and it has rate

$$r_{f\rightarrow i} = \frac{2\pi}{\hbar}|\mathbf{H}_{if}|^2\delta(E_f - E_i - E)f_f(1 - f_i) \tag{4.8}$$

The *net* rate of transitions $i \rightarrow f$ is therefore given by the difference between the two rates

$$r_{if} = r_{i\rightarrow f} - r_{f\rightarrow i} = \frac{2\pi}{\hbar}|\mathbf{H}_{if}|^2\delta(E_f - E_i - E)(f_i - f_f)\,. \tag{4.9}$$

For all such transitions involving a quantum of energy E we should sum over all pairs of initial and final states which differ by E. For a crystal, where states are distinguished by wavevector, we can convert the sum over discrete states $r(E) = \sum_{i,f} r_{if}$, into an integral over all initial states $\mathbf{k}_i$ and final states $\mathbf{k}_f$ which differ in energy by E, by introducing the density of states functions $g_f(\mathbf{k}_f)$ and $g_i(\mathbf{k}_i)$ for the relevant crystal bands (Eq. 3.11). Then we have for the transition rate per unit crystal volume,

$$r(E) = \frac{2\pi}{\hbar}\iint |\mathbf{H}_{if}|^2\delta(E_f - E_i - E)(f(E_i(\mathbf{k}_i)) - f(E_f(\mathbf{k}_f))) \times g_i(\mathbf{k}_i)g_f(\mathbf{k}_f)d^3\mathbf{k}_i d^3\mathbf{k}_f \tag{4.10}$$

The form of $\mathbf{H}_{if}$ depends upon the nature of the interaction. For instance, for optical transitions the matrix element delivers the condition that $\mathbf{k} = \mathbf{k}'$, and the matrix element then depends only on E.

If the $\mathbf{k}$ dependence is not needed explicitly (for instance, if the band structure is isotropic) then the integral can be expressed in terms of energies,

$$r(E) = \frac{2\pi}{\hbar}\int |\mathbf{H}_{if}|^2(f(E_i) - f(E_i + E))g_i(E_i)g_f(E_i + E)dE_i \tag{4.11}$$

To find the *total* rate of band to band transitions, $r(E)$ should be summed over the transition energy E, weighted by the density of available photon states $g_{ph}(E)$ of the perturbing field

$$G = \int r(E)g_{ph}(E)dE \tag{4.12}$$

Optical transitions may involve additional photons, or phonons. In such cases the first order approximation given by Fermi's Golden Rule fails, and it is necessary to expand the quantum mechanical transition rate to higher order. An important example is the case of optical transitions in an indirect band gap semiconductor, which involve a photon and a phonon. The higher order terms result in a different energy dependence of the absorption coefficient for indirect and direct gap materials, which is discussed below. The higher order approximations are beyond the scope of this book.

4.3.2. *Optical processes in a two level system*

Here we consider the simple case of a steady state light field interacting with a two level system. The two level model allows us to derive the distribution function for photons, which will be used later. The photon energy is equal to the difference in energies E_v and E_c. Since the light biases the system the two levels will not be in thermal equilibrium, but we will assume that they are each in quasi thermal equilibrium with quasi Fermi levels of E_{F_p} and E_{F_n} respectively [de Vos, 1992].

The following events can occur:

- a photon can be absorbed to promote an electron from E_v to E_c;
- an electron can relax to E_v from E_c while emitting a photon (spontaneous emission);
- a photon can *stimulate* the relaxation of an electron from to E_v from E_c together with the emission of a second photon (stimulated emission) (Fig. 4.2).

The first and third of these processes are closely related. They can be distinguished *only* by the relative occupation probabilities of the two levels, and not by the strength of the photon field: any photon can cause either type of event. Therefore we will combine them into the *net* absorption with a rate given by

$$r_{abs} = \frac{2\pi}{\hbar}|\mathbf{H}_{cv}|^2 f \cdot (f_v - f_c) \tag{4.13}$$

where f_v is the (Fermi Dirac) probability that there is an electron is the state E_v,

$$f_v = \frac{1}{e^{(E_v - E_{F_p})/k_B T} + 1}, \tag{4.14}$$

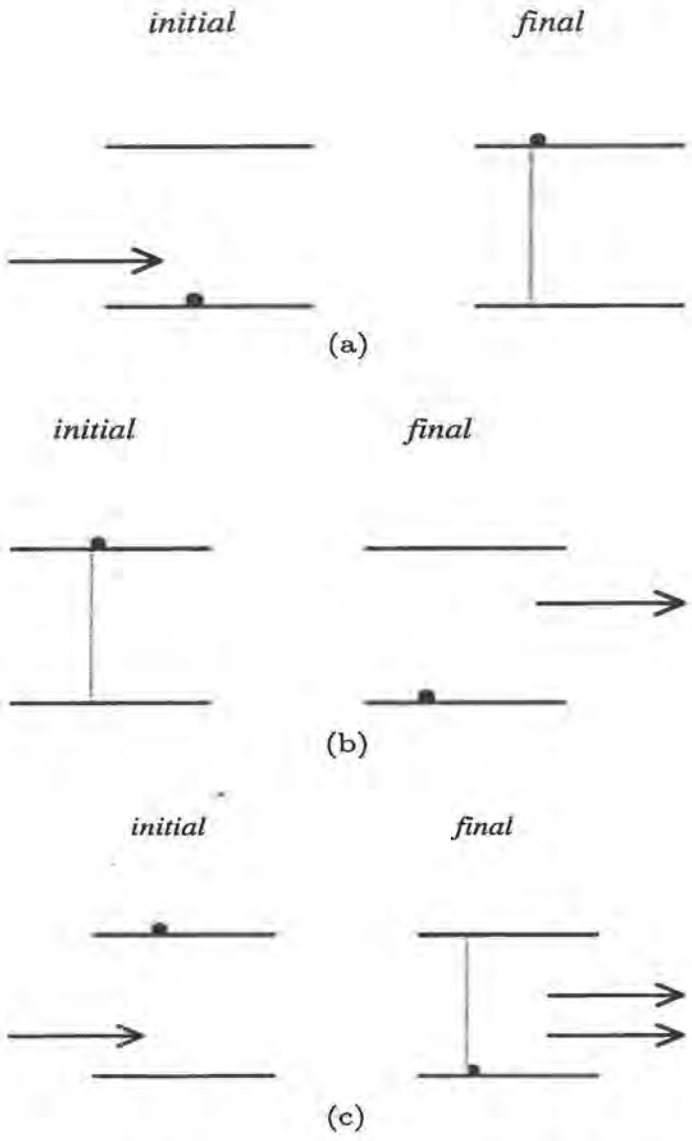


Fig. 4.2. (a) Absorption of a photon to promote an electron. (b) Relaxation of an electron accompanied by spontaneous photon emission. (c) Absorption of a photon to result in relaxation and stimulated photon emission.

f_c the probability that there is an electron in state E_c,

$$f_c = \frac{1}{e^{(E_c - E_{Fn})/k_B T} + 1}, \tag{4.15}$$

f the probability that there is a photon, and $\mathbf{H}_{cv}$ is the optical matrix element.

Spontaneous emission proceeds at a rate

$$r_{sp} = \frac{2\pi}{\hbar} |\mathbf{H}_{cv}|^2 f_c \cdot (1 - f_v). \tag{4.16}$$

The same matrix element controls absorption and spontaneous emission because of microscopic reversibility: although the initial and final states in Fermi's Golden Rule have been exchanged, the value of H_{cv}^2 remains the same.

In the steady state, the rates of all three processes must balance. Setting

$$r_{abs} - r_{sp} = 0\,, \tag{4.17}$$

we find that

$$f = \frac{1}{e^{(E-\Delta\mu)/k_B T} - 1} \tag{4.18}$$

where we have used the definition of $\Delta\mu$

$$\Delta\mu = E_{F_n} - E_{F_p}\,. \tag{4.19}$$

It is clear from Eq. 4.19 that a more intense light field will cause a greater separation in quasi Fermi levels. For this reason, $\Delta\mu$ is sometimes referred to as the 'chemical potential of radiation' [Wuerfel, 1982].

f is the quasi-equilibrium occupation function for photons. Thus, the number density of photons of energy E in quasi thermal equilibrium is given by

$$n_{ph}(E) = f(E)g_{ph}(E) \tag{4.20}$$

where $g_{ph}(E)$ is the density of photon states. Away from equilibrium $n_{ph}(E)$ will be determined by the strength of the electromagnetic field.

Equation 4.17 also gives us the useful result that:

$$f_c(1 - f_v) = f \cdot (f_v - f_c)\,. \tag{4.21}$$

4.4. Photogeneration

Photogeneration is by far the most important generation process in photovoltaic devices. By *photogeneration* we mean the generation of mobile electrons and holes through the absorption of light in the semiconductor. Other generation processes of relevance to photovoltaics are trap assisted and Auger generation, which will be mentioned in Chapters 8 and 10, respectively.

Though most relevant for photovoltaics, this is not the only optical process which occurs in semiconductors. Photons may alternatively be absorbed to increase the kinetic energy of mobile carriers (free carrier

absorption) or to generate phonons, or to promote electrons between localised states, or they may be scattered. The first two of these are usually important only at photon energies much smaller than the band gap (< 100 meV) and the first only at high carrier densities. Near to the band gap, band to band and localised state to band transitions are dominant.

Scattering, particularly by interfaces and by inhomogeneities in non uniform media, removes light from the incident beam without generating carriers, and so is undesirable for photovoltaics. However, it may be exploited in solar cell structures which are designed to trap the light and so amplify the photon field. This will be discussed in Chapter 9.

4.4.1. *Photogeneration rate*

The macroscopic absorption coefficient α describes how the light intensity is attenuated on passing through the material. $\alpha(E)$ may be considered as the sum of the absorption cross sections per unit volume of material for the various optical processes. Suppose a beam of photons of energy E and intensity I_0 is normally incident on a slab of absorbing material. A fraction $\alpha(E) \cdot dx$ of the photons of energy E entering a slab of thickness dx will be absorbed and the light intensity $I(x)$ will be attenuated by a factor $e^{-\alpha(E)dx}$. Hence

$$\frac{dI}{dx} = -\alpha I \,. \tag{4.22}$$

Integrating Eq. 4.22 for a material of non-uniform α the intensity at a depth x, $I(x)$, is given by

$$I(x) = I(0)e^{-\int_0^x \alpha(E,x')dx'} \tag{4.23}$$

where I(0) is the intensity just inside the surface (*i.e.*, after accounting for reflection). For uniform α, this reduces to the simple Beer–Lambert law, $I(x) = I(0)e^{-\alpha x}$ (Fig. 4.3). It can be shown that α is related to the imaginary part of the refractive index $\mathrm{Im}(n_s)$ through $\alpha = \frac{4\pi \mathrm{Im}(n_s)}{\lambda}$ where λ is the wavelength of light. α defined this way may also contain contributions from scattering; however, these are not considered in the microscopic derivation below.

If we can assume that all photons are absorbed to generate free carriers then the rate of carrier generation, per unit volume, at a depth x below the surface is given by

$$g(E,x) = b(E,x)\alpha(E,x) \tag{4.24}$$

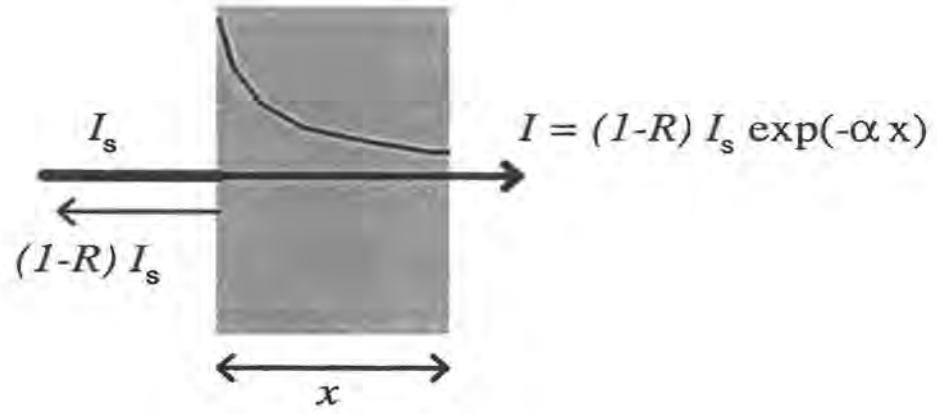


Fig. 4.3. Attenuation of light intensity in a slab of absorption α and thickness x. A fraction (1-R) of the light incident from the sun, I_s, is reflected at the front surface. The remaining intensity attenuates exponentially with distance in a uniform material, as shown by the thin black line. Photon flux density attenuates in the same way.

where b is the photon *flux* at x. Notice that it is the number and not the energy of the photons which determines the photogeneration rate. To relate g to the *incident* flux we need to allow for reflection of photons at the surface and attenuation within the material. Thus

$$g(E,x) = (1 - R(E))\alpha(E)b_s(E)e^{-\int_0^x \alpha(E,x')dx'} \tag{4.25}$$

where $b_s(E)$ is the incident flux and $R(E)$ is the reflectivity of the surface to normally incident light of energy E. This is the spectral photogeneration rate. To find the total generation rate at x we sum over photon energies

$$G(x) = \int g(E,x)dE\,. \tag{4.26}$$

The integral should be extended only over energies where photon absorption primarily results in free carrier generation.

4.4.2. *Thermalisation*

Photogeneration does not depend upon the energy of the absorbed photon, except in that the energy exceed the band gap. When higher energy photons are absorbed, they generate carriers with higher kinetic energy, but that energy is quickly lost and only E_g of potential energy remains to be collected, as shown in Fig. 4.4. The important quantity is the *number* of excitation events and not the amount of *energy* absorbed. This is one of the most important concepts in understanding photovoltaic devices.

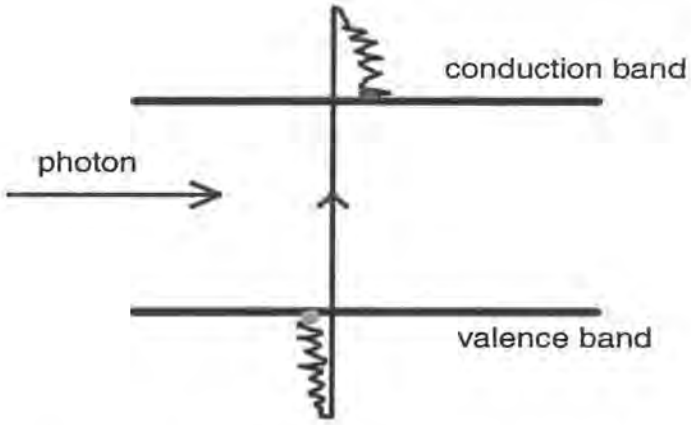


Fig. 4.4. Photogeneration of an electron-hole pair and loss of excess electron and hole kinetic energy by thermalisation to the respective band edges.

It is also the essential difference between photovoltaic and solar thermal action.

The photogenerated carriers lose any extra kinetic energy by *thermalisation*, or *cooling*. Microscopically, this means that they undergo repeated collisions with the lattice, giving up some of their kinetic energy to produce a phonon while they decay into a lower energy state, until they are in thermal equilibrium with the ambient. Phonons are the means by which energy is carried away to the outside world. This cooling happens very fast, on the order of picoseconds, partly because of the high density of final states which are available at lower kinetic energy. Cooling is faster than all but the most intense generation processes. In exceptional circumstances, for example for very high optical intensities or for electronic structures with reduced density of available states, then the carriers may not be able to lose energy fast enough and consequently are 'hot'. This will be discussed in detail in Chapter 10.

4.4.3. *Microscopic description of absorption*

Since photogeneration is the promotion of an electron from a valence to a conduction band energy level, the energy dependence of α must be strongly related to the density of valence and conduction band states and therefore to the band structure of the material. Here we will see how using Fermi's Golden Rule. Further details of the microscopic optical properties of semiconductors are given elsewhere [Bassani, 1975; Stern, 1963; Wuerfel, 1982; Bastard, 1986].

From Eq. 4.10, the net rate of transitions from an initial state in the valence band $|i\rangle = |v, \mathbf{k}_v\rangle$ to a final state in the conduction band $|f\rangle = |c, \mathbf{k}_c\rangle$ through an interval $E = E_c(\mathbf{k}_c) - E_v(\mathbf{k}_v)$ is given by

$$r(E) = \frac{2\pi}{\hbar} \iint |\mathbf{H}_{cv}|^2 \delta(E_c - E_v - E)(f_v(E_v(\mathbf{k}_v)) - f_c(E_c(\mathbf{k}_c))) \times g_v(\mathbf{k}_v) g_c(\mathbf{k}_c) d^3\mathbf{k}_v d^3\mathbf{k}_c \,. \tag{4.27}$$

Now, for an electromagnetic field of strength E_0, polarisation vector ε and angular frequency ω, where the wavelength of light is long compared to interatomic distances, the perturbation $\mathbf{H}'$ is given in the dipole approximation by

$$\mathbf{H}' = \frac{iqE_0}{2m_0\omega} \varepsilon \cdot \mathbf{p} \tag{4.28}$$

where $\mathbf{p}$ is the quantum mechanical momentum operator. Then

$$r(E) = \frac{2\pi}{\hbar} \frac{q^2 E_0^2 \hbar^2}{4m_0^2 E^2} \iint M_{cv}^2 \delta(E_c - E_v - E)(f_v(E_v(\mathbf{k}_v)) - f_c(E_c(\mathbf{k}_c))) \times g_c(\mathbf{k}') g_v(\mathbf{k}) d^3\mathbf{k} d^3\mathbf{k}' \tag{4.29}$$

where

$$M_{cv} = |\langle v, \mathbf{k}| \varepsilon \cdot \mathbf{p} | c, \mathbf{k}' \rangle| \tag{4.30}$$

is called the dipole matrix element.

Since $r(E)$ represents the net rate of photon absorption, the electromagnetic field is giving up *energy* to the semiconductor at a net rate

$$\frac{\partial U_E}{\partial t} = -E \times r(E) \tag{4.31}$$

per unit time, where U_E is the energy density of the radiation. Now, we use the fact that, for a plane wave, the cycle-averaged rate at which the beam loses energy with time is equal to the rate at which the beam loses intensity with distance,

$$\frac{\partial I}{\partial x} = \frac{\partial U_E}{\partial t} \tag{4.32}$$

if x is the direction of propagation. From Eq. 4.22 the absorption coefficient is given by

$$\alpha = -\frac{1}{I} \frac{dI}{dx} \,.$$

Now, since the intensity of radiation in a medium of refractive index n_s is related to U_E through

$$I = \frac{U_E c}{n_s} \tag{4.33}$$

and, from basic electromagnetic theory,

$$U_E = \frac{n_s^2 E_0^2}{8\pi} \tag{4.34}$$

we finally obtain for α:

$$\alpha = \frac{n_s}{c}\frac{E}{U_E} r_{abs}(E) . \tag{4.35}$$

Substituting for $r_{abs}(E)$, from Eq. 4.29, we obtain

$$\alpha(E) = \frac{A_\alpha}{m_0}\iint M_{cv}^2 \delta(E_c - E_v - E)(f_v(E_v(\mathbf{k}_v)) - f_c(E_c(\mathbf{k}_c))) \times g_c(\mathbf{k}')g_v(\mathbf{k})d^3\mathbf{k}d^3\mathbf{k}' \tag{4.36}$$

where

$$A_\alpha = \frac{4\pi^2 q^2 \hbar}{n_s c m_0 E} . \tag{4.37}$$

For low enough excitation levels, the valence band is effectively full ($f_v \approx 1$) and the conduction band empty ($f_c \approx 0$), so, to a good approximation

$$\alpha(E) = \frac{A_\alpha}{m_0}\iint M_{cv}^2 \delta(E_c - E_v - E) g_c(\mathbf{k}')g_v(\mathbf{k})d^3\mathbf{k}d^3\mathbf{k}' . \tag{4.38}$$

The macroscopic absorption coefficient may alternatively be considered as a sum of contributions from all valence (i) to conduction (f) band transitions which differ in energy by E

$$\alpha(E) = \sum_{i,f} \alpha_{if}(E) \tag{4.39}$$

where

$$\alpha_{if}(E) = \frac{A_\alpha}{m_0} M_{if}^2 \delta(E_f - E_i - E) \tag{4.40}$$

In this representation, the density of $\mathbf{k}$ states is included through the summation over different transitions.

Our expression for α will simplify when we evaluate the matrix element. At this stage it is worth distinguishing direct and indirect gap materials.

4.4.4. *Direct gap semiconductors*

In a crystal, electron and hole states can be written, using the Bloch theorem, as the product of a plane wave and a rapidly varying atomic part. Thus

$$|c, \mathbf{k}'\rangle = u_{\mathrm{c}}(\mathbf{r})e^{\mathrm{i}\mathbf{k}'\cdot\mathbf{r}} \tag{4.41}$$

and

$$|v, \mathbf{k}\rangle = u_{\mathrm{v}}(\mathbf{r})e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\,.$$

Now, acting on $|c, \mathbf{k}'\rangle$ with the differential operator $-i\hbar\varepsilon\cdot\nabla$, where the momentum operator has been written as $-i\hbar\nabla$, and writing the matrix element in integral notation we obtain for the matrix element

$$M_{\mathrm{cv}} = -i\hbar\varepsilon\cdot\int[e^{\mathrm{i}(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}u_{\mathrm{v}}^{*}(\mathbf{r})\nabla u_{\mathrm{c}}(\mathbf{r}) + i\mathbf{k}'e^{\mathrm{i}(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}u_{\mathrm{v}}^{*}(\mathbf{r})u_{\mathrm{c}}(\mathbf{r})]d^{3}\mathbf{r}\,. \tag{4.42}$$

First, the integral is non zero only when $\mathbf{k}' - \mathbf{k} = 0$. In other words, momentum must be conserved. This is the usual *selection rule* for optical transitions. It is the reason for the distinction between direct and indirect gap materials. Second, the contribution from the second term in the intergrand must vanish since the different bands c, and v, must be orthoghonal. This is a consequence of quantum mechanics. Then the dipole matrix element is simply given by $\langle c|\varepsilon\cdot\mathbf{p}|v\rangle$. M_{cv} is a property of the material and, in general, of the electric field direction. For isotropic materials $M_{\mathrm{cv}}^{2} = \frac{1}{3}|\langle c|\mathbf{p}|v\rangle|^{2}$.

Now α can be evaluated simply from the densities of states. When these depend only on energy we have

$$\alpha(E) = \frac{A_{\alpha}}{m_0}M_{\mathrm{cv}}^{2}\int g_{\mathrm{c}}(\mathbf{k}(E_{\mathrm{i}}+E))g_{\mathrm{v}}(\mathbf{k}(E_{\mathrm{i}}))dE_{\mathrm{i}}\,. \tag{4.43}$$

The quantity given by the integral is known as the joint density of states (JDOS). In the case of the parabolic band approximation (introduced in Chapter 3), where $g_{\mathrm{v}}(E)$ varies like $(E_{\mathrm{v0}} - E)^{1/2}$ and $g_{\mathrm{c}}(E)$ varies like $(E - E_{\mathrm{c0}})^{1/2}$, it is straightforward to show that

$$\alpha(E) = \alpha_0(E - E_{\mathrm{g}})^{1/2} \tag{4.44}$$

where α_0 is a material dependent constant, and $E_{\mathrm{g}} = E_{\mathrm{c0}} - E_{\mathrm{v0}}$. Thus, the absorption coefficient reflects the shape of the individual densities of states. In general the JDOS of a direct gap material will look like the product of the conduction band and valence band density of states (DOS) functions.

4.4.5. *Indirect gap semiconductors*

In indirect gap materials, this is not the case. Equation 4.42 showed that optical transitions cannot occur unless $\mathbf{k}' - \mathbf{k} = 0$. Therefore an electron cannot be excited from the valence band maximum to the conduction band minimum in an indirect gap material simply by the absorption of a photon. What can happen, however, is that the electron can be excited simultaneously with the absorption or emission of a *phonon*. In the first case the ground state is a composite state of electron in the valence band plus a phonon, and the final state is a composite of an electron in the conduction band and no phonon:

$$|i\rangle = |v, \mathbf{k}; \omega_\mathrm{p}, \mathbf{k}_\mathrm{p}\rangle$$

$$|f\rangle = |c, \mathbf{k}'; 0\rangle$$

where the phonon has energy $E_\mathrm{p} = \hbar\omega_\mathrm{p}$ and momentum $\hbar\mathbf{k}_\mathrm{p}$. Conservation of energy requires that

$$E_\mathrm{p} = E_\mathrm{c}(\mathbf{k}') - E_\mathrm{v}(\mathbf{k}) - \hbar\omega\,. \tag{4.45}$$

Evaluation of the matrix element delivers the condition due to momentum conservation that $\mathbf{k}' - \mathbf{k}_\mathrm{p} - \mathbf{k} = 0$. Transitions from the valence band maximum to the conduction band minimum are allowed if a phonon of suitable $\mathbf{k}$ is available (Fig. 4.5). The sum over all initial and final states must now involve the probability density of phonons of energy E_p and wavevector $\mathbf{k}_\mathrm{p}$, $n_\mathrm{p}(E_\mathrm{p})$, as well as the JDOS of conduction and valence bands.

We will estimate the absorption coefficient for an indirect transition with phonon absorption. See also ([Pankove, 1971]). The transition rate

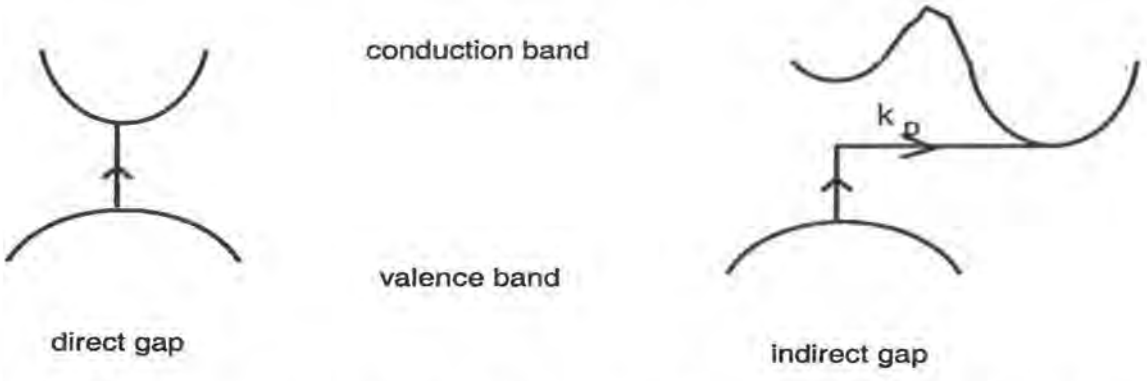


Fig. 4.5. Absorption in a direct and indirect gap band structure. In the indirect gap material, a phonon of momentum $\mathbf{k}_\mathrm{p}$ is needed to conserve momentum.

depends upon the availability of a phonon of suitable energy, the density of occupied valence band states and of unoccupied conduction band states. According to Bose–Einstein statistics the probability of finding a phonon of energy E_p is given by $\frac{1}{e^{E_\mathrm{p}/k_\mathrm{B}T}-1}$. If the various matrix elements are energy independent, the conduction band is initially empty and the valence band full, then the transition rate varies like

$$r \propto \frac{1}{e^{E_\mathrm{p}/k_\mathrm{B}T}-1} \int g_\mathrm{c}(E_\mathrm{i}+E+E_\mathrm{p}) g_\mathrm{v}(E_\mathrm{i}) dE_\mathrm{i} \tag{4.46}$$

where the integral can be taken over energies, rather than $\mathbf{k}$, since the phonon distribution is isotropic in $\mathbf{k}$. Using the densities of states for parabolic bands such that $g_\mathrm{v}(E) \propto (E_\mathrm{v0}-E)^{1/2}$ and $g_\mathrm{c}(E) \propto (E-E_\mathrm{c0})^{1/2}$, the integral can be evaluated to find

$$\alpha_{\mathrm{p}-} \propto \frac{(E_\mathrm{g}-E-E_\mathrm{p})^2}{e^{E_\mathrm{p}/k_\mathrm{B}T}-1} .$$

A second type of indirect transition, where the photon absorption causes the *generation* of a phonon, should be included. In this case the transition rate is proportional to the probability of phonon emission, $\frac{1}{1-e^{-E_\mathrm{p}/k_\mathrm{B}T}}$ and the contribution to α varies like

$$\alpha_{\mathrm{p}+} \propto \frac{(E_\mathrm{g}-E-E_\mathrm{p})^2}{1-e^{-E_\mathrm{p}/k_\mathrm{B}T}} .$$

The net absorption coefficient is the sum of contributions from each type of event. At photon energies which are such that $E-E_g \gg k_\mathrm{B}T$, E_p can be neglected in the numerator and the absorption coefficient has the quadratic form

$$\alpha \propto (E-E_\mathrm{g})^2 . \tag{4.47}$$

The absorption coefficient of an indirect material has very different behaviour to a direct gap material and does not reflect the JDOS at photon energies close to the band edge. $\alpha(E)$ is generally smaller and rises more smoothly from the band edge than in direct gap materials. Similar behaviour is observed in highly doped semiconductors where electron scattering, rather than phonon emission and absorption, assists the optical transitions.

At higher photon energies, direct optical transitions are allowed in the indirect gap material. The contribution of these direct transitions should reflect the JDOS without any phonon contributions and so rise more rapidly

with energy than the indirect contributions. The direct contributions usually appear as a change in curvature in $\alpha(E)$.

4.4.6. *Other types of behaviour*

In practice, the $(E-E_g)^{1/2}$ behaviour predicted by Eq. 4.44 is seldom seen. At energies above E_g the DOS functions diverge from the parabolic approximation, while at energies just below E_g *excitons* influence the absorption.

The parabolic band approximation is only good close to the band extrema, typically within 100 meV. At values of $\mathbf{k}$ far from the band extrema the $E(\mathbf{k})$ curve begins to flatten out, and becomes stationary at other high symmetry values of $\mathbf{k}$. This leads to local maxima and other features in in the DOS, which can often be identified from peaks in the absorption spectrum.

As a better approximation to $\alpha(E)$ than $(E-E_g)^{1/2}$, parametric forms may be derived for groups of semiconductors sharing a similar crystal structure. These will depend typically upon the energies of the important symmetry points, band gaps and dipole matrix elements. Various models have been developed for the III–V and II–VI families of semiconductors [Adachi, 1992].

Multiple step photogeneration

As well as direct band to band photogeneration, free charge carriers may be generated indirectly, where the photon is absorbed to create an excited state which subsequently dissociates to release one or more free carriers. Examples are absorption by excitons and by sensitisers. Excitons are Coulombic bound states of electron-hole pairs, and were introduced in Chapter 3. An exciton can be created by a photon of energy smaller than E_g. The difference between the exciton energy and the band gap, called the exciton binding energy, is usually less than k_BT and so most excitons dissociate at room temperature. However, the binding energy will be strong in cases where the photogenerated electron and hole wavefunctions are localised, as in the case of low dimensional semiconductor structures and molecular semiconductors. In those cases the excitons are important features of the room temperature absorption. The magnitude of the excitonic absorption, its oscillator strength, is generally larger for more strongly bound excitons. Note that only stationary ($\mathbf{k} = 0$) excitons can be generated optically.

The important point for application to photovoltaics is that excitonic absorption does not automatically generate mobile carriers. The exciton needs to dissociate first, and it may recombine before that happens. Therefore in modelling device behaviour we should use for G the *net* generation rate of free carriers from exciton generation and dissociation, and not simply the optical excitation rate.

Sensitisers are analogous to excitons and are usually deliberately introduced as optical absorbers. The sensitiser may be a molecular species such as a dye molecule, or a small solid state particle in contact with a semiconductor surface. Such systems are widely used in photography. The photon creates an excited state, which may then dissociate into a charged pair, following the sequence

$$S + h\nu \rightarrow S^* \Leftrightarrow S^+ + e^-$$

where S, S^* and S^+ represent the ground state, excited state, and ionised state of the sensitiser. The final stage involves injection of the free carrier into the semiconductor. Sensitisers at heterojunctions may inject charges of opposite sign into the two different media forming the junction [Hagfeldt, 2000].

In the case of absorption by excitons or sensitisers the optical generation rate should be replaced by

$$g(E,x) = (1 - R(E))\eta_{\mathrm{diss}}(E)\alpha(E)b_s(E)e^{-\int_0^x \alpha(E,x')dx'}$$

where $\eta_{\mathrm{diss}}(E)$ is the quantum efficiency for dissociation, *i.e.*, the probability that one absorbed photon of energy E will generate a free charge.

Amorphous materials

In amorphous (non-crystalline) material, the lack of long-range order means that crystal momentum need not be conserved in an optical transition. The band gap is always 'direct'. Absorption events are therefore more likely than in the equivalent crystalline material; for example, the absorption coefficient is larger in amorphous than in crystalline silicon. The form of $\alpha(E)$ is dominated by the JDOS without $\mathbf{k}$ conservation restrictions. This will be discussed for the case of amorphous silicon in Chapter 8.

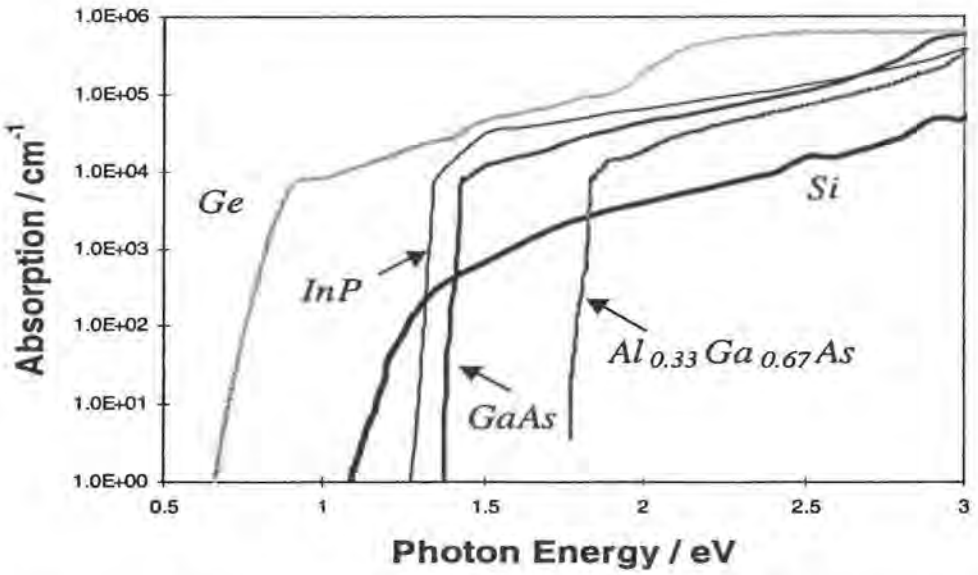


Fig. 4.6. Absorption spectra of some common photovoltaic semiconductors. Notice how the band edge of the direct semiconductors, GaAs, InP and $Al_xGa_{1-x}As$ is sharper than for the indirect semiconductors silicon and germanium. The sharp edges of the GaAs and InP absorption are influenced by excitonic effects.

4.4.7. *Examples and data*

Figure 4.6 shows the absorption spectra of a number of semiconductors which are important for photovoltaics. Notice how the absorption edge for the direct band gap semiconductors GaAs and InP, is sharper than for the indirect band gap materials, silicon and germanium. Notice how the shape of the curves for GaAs, InP and $Al_xGa_{1-x}As$ are similar; this is due to their similar crystal structure.

The absorption length of a photovoltaic material is a useful quantity. This is defined as the distance light of a particular wavelength must travel before the intensity is attenuated by a factor e, and is given by $\frac{1}{\alpha}$. At visible wavelengths the direct band gap materials GaAs and InP have absorption lengths of less than one micron. This means that only a few microns of material are needed to absorb virtually all of the light. In contrast, the indirect gap material Si has an absorption length of tens of microns, so that wafers tens or hundreds of μm thick are needed for good absorption.

Reflectivity has not been treated explicitly in this section. For typical semiconductors R(E) is $\approx$ 30–40% at visible wavelengths, and so it is an important factor. The net reflectivity of a semiconductor surface can be reduced using anti-reflection coatings or surface texturing; these are treated in Chapter 9.

4.5. Recombination

4.5.1. *Types of recombination*

By *recombination* we refer to the loss of mobile electrons or holes by any of a number of removal mechanisms. Unlike generation, where one mechanism is dominant, several different recombination mechanisms are important for the photovoltaic device.

We should distinguish two categories: *unavoidable* recombination processes which are due to the essential physical processes in the intrinsic material, and *avoidable* processes which are largely due to imperfect material.

Amongst the unavoidable recombination processes are the processes which result from optical generation, spontaneous and stimulated emission (see Sec. 4.3). For photovoltaics, the most important of these is spontaneous emission, which is also known as radiative recombination.

The other important unavoidable process is the interaction of an electron or hole with a second similar carrier, resulting in the decay of one carrier across the band gap and the increase in the kinetic energy of the other carrier by an amount equal to the band gap. This is called *Auger* recombination. It is the reverse of a rare generation process where a carrier with kinetic energy greater than the band gap is able to give up some of its kinetic energy to excite an electron across the gap (discussed in Chapter 10). Auger recombination is important in low band gap materials with high carrier densities, where carrier–carrier interactions are stronger.

Avoidable recombination processes usually involve relaxation by way of a localised trap state. These trap states are due to impurities in the crystal or defects in the crystal structure. These are often known as non-radiative recombination processes (although Auger is also non-radiative) and are usually the dominant mechanisms. Recombination in semiconductors is well covered by Landsberg [Landsberg, 1990] and many semiconductor textbooks such as Shur [Shur, 1990] or Tyagi [Tyagi, 1991].

4.5.2. *Radiative recombination*

Box 4.2. Derivation of the radiative recombination rate

Here we are going to derive an expression for the rate of radiative recombination using our results for absorption coefficient.

From Sec. 4.3 above we know that the rate of spontaneous relaxation events from an initial state $|f\rangle = |c, \mathbf{k}_c\rangle$ of energy E_c to a final state $|i\rangle =$

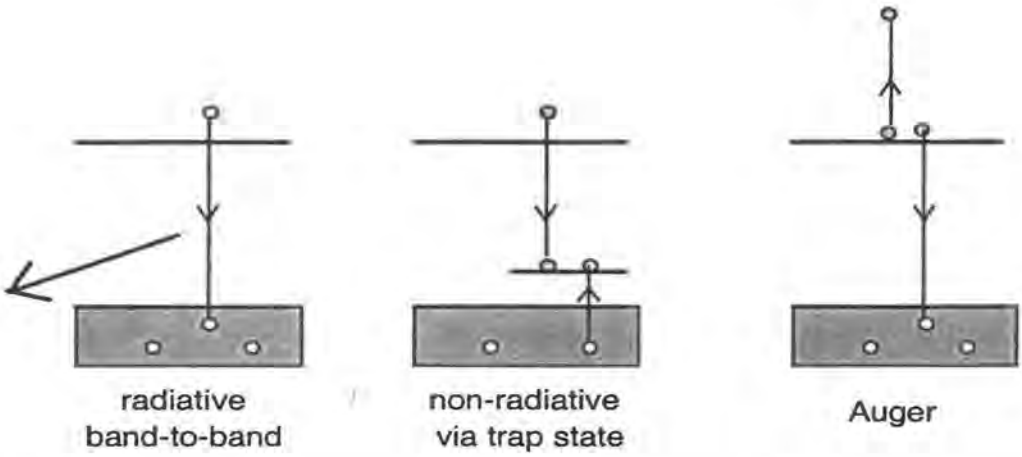


Fig. 4.7. Radiative band-to-band, non-radiative and Auger recombination.

$|v, \mathbf{k}_v\rangle$ of energy E_v is given by

$$r_{sp} = \frac{2\pi}{\hbar}\mathbf{H}_{cv}^2 f_c \cdot (1 - f_v)$$

(Eq. 4.13) while the net rate of absorption is

$$r_{abs} = \frac{2\pi}{\hbar}\mathbf{H}_{cv}^2 f \cdot (f_v - f_c)$$

(Eq. 4.16) where f is the probability that there is a photon of energy $E = (E_c - E_v)$. At quasi thermal equilibrium the two rates match and we obtain Eq. 4.18 for f. We will call this $f_{eq}(E)$. Away from equilibrium

$$r_{sp} = r_{abs} \times \frac{f_{eq}}{f}. \tag{4.48}$$

Now, under an incident field the probability that there is a photon of energy E is related to the (non-equilibrium) number density of photons n_{ph} by

$$f = \frac{n_{ph}}{g_{ph}} \tag{4.49}$$

where g_{ph} is the density of photon states in the energy range E to $E + dE$. In an optically isotropic medium of refractive index n_s

$$g_{ph}(E) = \frac{8\pi n_s^3 E^2}{h^3 c^3} \tag{4.50}$$

n_{ph} is related to the field energy density through

$$n_{ph} = \frac{U_E}{E} \tag{4.51}$$

Now we can relate the spontaneous emission rate to the absorption coefficient, using Eq. 4.35 to relate $r_{\rm abs}$ to α,

$$r_{\rm sp} = \frac{r_{\rm abs}E}{U_{\rm E}} \times g_{\rm ph} f_{\rm eq} = \frac{c}{n_{\rm s}} \alpha(E) g_{\rm ph}(E) f_{\rm eq}(E) \,. \tag{4.52}$$

Substituting for $g_{\rm ph}$ and $f_{\rm eq}$ we have

$$r_{\rm sp} = \frac{8\pi n_{\rm s}^2}{h^3 c^2} \frac{\alpha(E)E^2}{e^{(\rm E-\Delta\mu)/k_B T} - 1} \,. \tag{4.53}$$

This is the volume rate of radiative recombination, taken over all angles of emission. For applications in devices we are most often interested in the recombination resolved along a particular direction. In that case, the geometric factor is given by π rather than 4π and the recombination rate is reduced by a factor 1/4:

$$r_{\rm sp} = \frac{2\pi n_{\rm s}^2}{h^3 c^2} \frac{\alpha(E)E^2}{e^{(\rm E-\Delta\mu)/k_B T} - 1} \,. \tag{4.54}$$

This rate can be expressed in terms of the photon flux from a (biased) black body, $b(E, \Delta\mu)$

$$r_{\rm sp} = b_{\rm e}(E, \Delta\mu)\alpha(E) \tag{4.55}$$

where

$$b_{\rm e}(E, \Delta\mu) = F_{\rm a} \frac{2}{h^3 c^2} \frac{E^2}{e^{(\rm E-\Delta\mu)/k_B T} - 1}$$

as shown in Eq. 2.14, where $F_{\rm a} = \pi$ for emission normal to the cell surface and $n_{\rm s} = 1$.

We obtain the total radiative recombination rate by integrating 4.55 over photon energies:

$$U_{\rm rad}^{\rm total} = \int_0^\infty \alpha(E) b_{\rm e}(E, \Delta\mu) dE \,.$$

As explained in Sec. 4.1, to obtain the *net* recombination rate we must subtract the rate at thermal equilibrium, which is the rate when $\Delta\mu = 0$,

$$U_{\rm rad} = \int_0^\infty \alpha(E) b_{\rm e}(E, \Delta\mu) dE - \int_0^\infty \alpha(E) b_{\rm e}(E, 0) dE \,. \tag{4.56}$$

Spatial variations in the recombination rate enter through spatial variations in $\Delta\mu(x)$ (see also Wuerfel [Wuerfel, 1982] and Stern [Stern, 1963]).

4.5.3. *Simplified expressions for radiative recombination*

In order to use the recombination rate into the transport equations, it would be useful to be able to express $U_{\rm rad}$ in terms of the carrier densities. In a non-degenerate semiconductor, $E - \Delta\mu \gg k_{\rm B}T$ for all energies E at which the rate, Eq. 4.55, is non-negligible. Then, to a good approximation,

$$\frac{E^2}{e^{(\rm E-\Delta\mu)/k_B T}-1} \approx e^{-(\rm E-\Delta\mu)/k_B T} E^2 .$$

Using the relation for a semiconductor under bias, Eq. 3.53,

$$np = e^{\Delta\mu/k_B T} n_{\rm i}^2 ,$$

we can write Eq. 4.56 as

$$U_{\rm rad} = B_{\rm rad}(np - n_{\rm i}^2) \tag{4.57}$$

where

$$B_{\rm rad} = \frac{1}{n_{\rm i}^2}\frac{2\pi}{h^3c^2}\int_0^\infty n_{\rm s}^2 \alpha(E) e^{-\rm E/k_B T} E^2 dE . \tag{4.58}$$

The radiative recombination coefficient $B_{\rm rad}$ is carrier density independent and is a property of the material.

The expression simplifies further for doped material. If the density of photogenerated electrons and holes is each equal to Δn, then in p type material with a doping density $N_{\rm a}$,

$$(np - n_{\rm i}^2) = (n_0 + \Delta n)(p_0 + \Delta n) - n_0 p_0 \approx \Delta n N_{\rm a}$$

where n_0 and p_0 are the electron and hole densities in equilibrium, given by $n_{\rm i}^2/N_{\rm a}$ and $N_{\rm a}$, respectively. This means that $U_{\rm rad}$ is proportional to the excess minority carrier density,

$$U_{\rm rad} = \frac{n - n_0}{\tau_{\rm n,rad}} \tag{4.59}$$

where

$$\tau_{\rm n,rad} = \frac{1}{B_{\rm rad} N_{\rm a}} \tag{4.60}$$

is the *minority carrier radiative lifetime*. A similar analysis shows that in n type material of doping density $N_{\rm d}$,

$$U_{\rm rad} = \frac{p - p_0}{\tau_{\rm p,rad}} \tag{4.61}$$

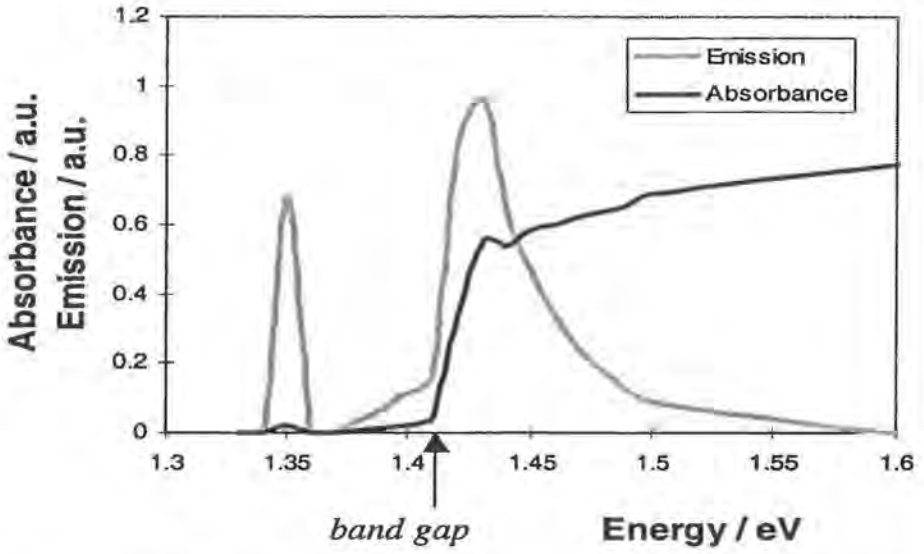


Fig. 4.8. Absorption and calculated emission spectrum for a slab of GaAs with a small density of defect levels below the band edge. The emission spectrum is strongly influenced by the shape of the absorption near the band edge. Notice how a small density of impurity states below the band edge, which is barely visible in the absorption spectrum, is strongly visible in the emission spectrum.

with

$$\tau_{\mathrm{p,rad}} = \frac{1}{B_{\mathrm{rad}} N_{\mathrm{d}}} \quad (4.62)$$

The radiative lifetime can be measured from the time resolved spontaneous emission following instantaneous optical excitation of the semiconductor (called photoluminescence or fluorescence). B_{rad} can be determined experimentally from the variation of τ_{rad} with doping.

Equation 4.58 shows that B_{rad} is larger for materials with a high absorption coefficient, and therefore radiative recombination is more important in direct band gap materials. The exponential term means that, relative to absorption, contributions from energy levels closer to the band edges are much more important. It also means that radiative recombination from either band to impurity states inside the band gap can be very important, and can dominate over band-to-band events.

In cases where the chemical potential is uniform, the radiative recombination rate can be directly related to bias. If $\Delta\mu(x) \approx qV$ throughout the material then, using Eq. 4.57,

$$U_{\mathrm{rad}} = U_{\mathrm{rad,0}}(e^{\mathrm{qV/k_B T}} - 1) \quad (4.63)$$

where $U_{\mathrm{rad,0}}$ is a material dependent constant.

Radiative recombination is unimportant for practical cells at operating point but in the limit of perfect material it is the mechanism which limits efficiency.

Box 4.3. Bimolecular recombination

The detailed treatment of radiative recombination above shows how the rate of recombination is proportional to the densities of both types of carrier. In fact, it is usual for band-to-band recombination processes which do not depend on the presence of a third quantity to vary like

$$U = Bnp$$

where B is a material and process dependent constant. We show here that this is consistent with Fermi's Golden Rule.

For transitions from valence to conduction band, we may replace $(f_\mathrm{v} - f_\mathrm{c})$ with $(1 - f_\mathrm{v}).f_c$ in Eq. 4.27. Since under normal conditions the conduction band is mainly empty while the valence band is mainly full, the Fermi Dirac functions vary exponentially with energy over the relevant ranges of E_v and E_c. The factors f_c and f_v are therefore likely to be the most rapidly varying factors. If the matrix element is only weakly dependent on initial and final state energies, it can be taken outside the integral in the expression for the transition rate in and the integrations over the conduction and valence bands separate. In the case where the band structure is isotropic,

$$U = \frac{2\pi}{\hbar}|\mathbf{H}'_{\mathrm{cv}}|^2 \int_{-\infty}^{\mathrm{E_{v,0}}} (1 - f_\mathrm{v}(E_\mathrm{v}))g_\mathrm{v}(E_\mathrm{v})dE_\mathrm{v} \int_{E_{\mathrm{c0}}}^{\infty} f_\mathrm{c}(E_\mathrm{c})g_\mathrm{c}(E_\mathrm{c})dE_\mathrm{c}\,. \quad (4.64)$$

The two integrals in Eq. 4.64 are in fact the density of holes in the valence band and electrons in the conduction band. Thus, the transition rate is proportional to the product of the free carrier densities, $U \propto np$. (Remember that the thermal recombination rate must be subtracted to obtain the net recombination rate!)

For transitions from conduction band to a localised site, or from localised site to valence band, we replace the conduction (or valence) band DOS function with the DOS for the localised state, and n (or p) with the density of electrons (or holes) in the localised state. For conduction band to localised state transitions we have

$$U = BnN_\mathrm{t}(1 - f_\mathrm{t}) \quad (4.65)$$

where N_t is the density of localised states and f_t the probability that the localised state is occupied.

Relationships between rates and carrier densities are more complicated when the initial to final state transition is a multi step process and when the transition depends on the availability of a third entity. Shockley Read Hall recombination via a trap state in the band gap is an example of a multi-step process and Auger recombination, considered next, is a three carrier process.

4.5.4. *Auger recombination*

In *Auger* recombination, a collision between two similar carriers results in the excitation of one carrier to a higher kinetic energy, and the recombination of the other across the band gap with a carrier of opposite polarity. The energy which is released through recombination is given up as kinetic energy to the other carrier. Ultimately that extra energy will be lost as heat as the excited carrier relaxes to the band edge.

For band-to-band Auger recombination, an electron and two holes or a hole and two electrons are involved. By a similar argument to that used above, the rate is proportional to the densities of all three carriers, so that the net rate varies like

$$U_{\mathrm{Aug}} = A_{\mathrm{p}}(n^2 p - n_0^2 p_0)$$

for two-electron collisions and

$$U_{\mathrm{Aug}} = A_{\mathrm{n}}(np^2 - n_0 p_0^2) \tag{4.66}$$

for two-hole collisions. Auger processes are most important where carrier densities are high, for instance in low band gap and doped materials, or at high temperature. The dependence on doping density is strong. In p-type doped material the electron lifetime for band-to-band Auger recombination is given by

$$\tau_{\mathrm{n,Aug}} = \frac{1}{A_{\mathrm{n}} N_{\mathrm{a}}^2} . \tag{4.67}$$

Similarly, the hole lifetime in n-type material is

$$\tau_{\mathrm{p,Aug}} = \frac{1}{A N_{\mathrm{d}}^2} .$$



Fig. 4.9. Energy-momentum representation of Auger recombination in direct (left) and indirect (right) band gap materials. A collision between two electrons near to the minimum of the conduction band results in the promotion of one electron to a higher energy state and recombination of the other with a hole. Both energy and momentum must be conserved. The right-hand figure shows that Auger events may occur between conduction band minimum and valence band maximum in indirect band gap materials, since the difference in momentum can be taken up by the promoted electron.

Auger recombination can also occur via a trap state. An electron colliding with an occupied trap state close to the conduction band can stimulate the recombination of the electron in the trap state with a valence band hole, whilst gaining kinetic energy. Similarly a hole colliding with an empty trap close to the valence band can stimulate Auger recombination. In these cases the minority carrier lifetimes vary like $\frac{1}{N_a N_t}$ and $\frac{1}{N_d N_t}$.

Auger events conserve momentum as well as energy. An electron with energy E and momentum $\mathbf{k}$ can recombine with a hole of energy $E - E'$ and momentum $\mathbf{k}'$, provided that there is an electron state available at $(E + E')$, $\mathbf{k} + \mathbf{k}'$ (see Fig. 4.9). This means that Auger recombination *can* occur in indirect band gap materials, unlike radiative recombination which is suppressed. Auger recombination is therefore much more important in indirect than direct band gap materials, and is the dominant loss mechanism in very pure silicon and germanium.

4.5.5. *Shockley Read Hall recombination*

By far the most important recombination processes in real semiconductors are those which involve defect or *trap* states in the band gap. Since a trap state is spatially localised whilst the free electron or hole is delocalised, we can think of the free carrier as being captured by the trap. The carrier can subsequently be released by thermal activation. Alternatively, if the trap captures a carrier of the opposite polarity before the first carrier is released,

then the trap has been emptied again and the two carriers have in effect recombined. This is illustrated in Fig. 4.11. Localised states which serve mainly to capture and release only one type of carrier are usually called traps. Those which capture both types of carrier are called recombination centres. Usually recombination centres lie deeper into the band gap than traps.

Box 4.4. Derivation of the Shockley Read Hall (SRH) recombination rate

Consider a semiconductor containing a density N_t trap states at an energy E_t in the band gap. Empty traps can capture electrons from the conduction band, and filled traps can capture holes from the valence band. The rate at which electrons are captured in this bimolecular process is given by Eq. 4.65

$$U_{nc} = B_n n N_t (1 - f_t)$$

where f_t is the probability that the trap is occupied. The coefficient can be expressed as

$$B_n = v_n \sigma_n \tag{4.68}$$

where v_n is the mean thermal velocity of the electron, and σ_n the capture cross section of the trap for electrons. It will also be useful to define a lifetime for electron capture by the trap

$$\tau_{n,SRH} = \frac{1}{B_n N_t} . \tag{4.69}$$

The rate at which electrons are released from the trap depends on the occupation of the traps and can be written

$$G_{nc} = \frac{N_t f_t}{\tau_{esc}} \tag{4.70}$$

where the release time τ_{esc} is determined by the condition that $U_{nc} = G_{nc}$ in equilibrium. Then it follows that

$$G_{nc} = B_n n_t N_t f_t \tag{4.71}$$

where n_t is the value of the electron density when the electron Fermi level is equal to the trap level,

$$n_t = n_i e^{(E_t - E_i)/k_B T} . \tag{4.72}$$

In a similar way, holes are captured at a rate

$$U_{\mathrm{pc}} = B_{\mathrm{p}} p N_{\mathrm{t}} f_{\mathrm{t}} \tag{4.73}$$

where

$$B_{\mathrm{p}} = \nu_{\mathrm{p}} \sigma_{\mathrm{p}} \tag{4.74}$$

and

$$\tau_{\mathrm{p,SRH}} = \frac{1}{B_{\mathrm{p}} N_{\mathrm{t}}}\,, \tag{4.75}$$

and released at a rate

$$G_{\mathrm{pc}} = B_{\mathrm{p}} p_{\mathrm{t}} N_{\mathrm{t}} (1 - f_{\mathrm{t}}) \tag{4.76}$$

where p_{t} is the value of the hole density when $E_{\mathrm{F_p}} = E_{\mathrm{t}}$,

$$p_{\mathrm{t}} = n_{\mathrm{i}} e^{(\mathrm{E_i} - \mathrm{E_t})/\mathrm{k_B T}}\,. \tag{4.77}$$

In general, the rates of capture and release for electrons and holes will be different, depending on the affinity of the trap for electrons or holes and its position in the band gap. However, in the steady state, the net rate of electron capture by the traps $U_{\mathrm{nc}} - G_{\mathrm{nc}}$ must be equal to the net rate of hole capture $U_{\mathrm{pc}} - G_{\mathrm{pc}}$, since charges cannot be allowed to build up on the traps. This condition fixes the value of f_{t},

$$f_{\mathrm{f}} = \frac{B_{\mathrm{n}} n - B_{\mathrm{p}} p_{\mathrm{t}}}{B_{\mathrm{n}}(n + n_{\mathrm{t}}) + B_{\mathrm{p}}(p + p_{\mathrm{t}})}\,. \tag{4.78}$$

f_{t} can then be eliminated from the expressions for $U - G$. Finally we find for the net recombination rate

$$U_{\mathrm{SRH}} = \frac{np - n_{\mathrm{i}}^2}{\tau_{\mathrm{n,SRH}}(p + p_{\mathrm{t}}) + \tau_{\mathrm{p,SRH}}(n + n_{\mathrm{t}})} \tag{4.79}$$

where we have used the definitions of $\tau_{\mathrm{n,SRH}}$ and $\tau_{\mathrm{p,SRH}}$. This is the Shockley Read Hall expression for recombination through a single trap state.

For doped semiconductors, U_{SRH} simplifies. In p type material, provided that $\tau_{\mathrm{n}} N_{\mathrm{a}} \gg \tau_{\mathrm{p}} n_{\mathrm{t}}$ and $N_{\mathrm{a}} \gg p_{\mathrm{t}}$, U_{SRH} becomes proportional to the excess carrier density,

$$U_{\mathrm{SRH}} \approx \frac{(n - n_0)}{\tau_{\mathrm{n,SRH}}}\,. \tag{4.80}$$

In n type material

$$U_{\mathrm{SRH}} \approx \frac{(p - p_0)}{\tau_{\mathrm{p,SRH}}} . \tag{4.81}$$

However, when n_t and p_t, or when $\tau_{\mathrm{n,SRH}}$ and $\tau_{\mathrm{p,SRH}}$ differ by orders of magnitude, these limiting forms may not apply. Then the full expression must be used.

SRH recombination is strongest when n and p are of similar magnitude. By examining Eq. 4.79 can see that for a mid gap trap with equal capture times, U has its maximum when $n = p$. (See Fig. 4.10.) This means that in undoped regions, where n and p may be similar, SRH recombination is more important relative to radiative recombination. (Equation 4.57 shows that radiative recombination depends only on the np product, which is constant for uniform $\Delta\mu$.) This dependence on the n/p ratio influences the bias dependence of the SRH recombination rate. We will see in Chapter 6 that SRH recombination varies like e^{qV/k_BT} in doped material but like $e^{qV/2k_BT}$ in the 'depleted' layer between two differently doped materials.

In real semiconductors, there may be several trap levels in the band gap and carriers may recombine by more than one step. However, the largest contribution is from traps which are located close to the centre of the band gap, for $n \approx p$. So, for bulk material with a uniform density of traps, the approximation is usually made that recombination through one particular trap level is dominant. Multi-level recombination is discussed by Landsberg [Landsberg, 1990].

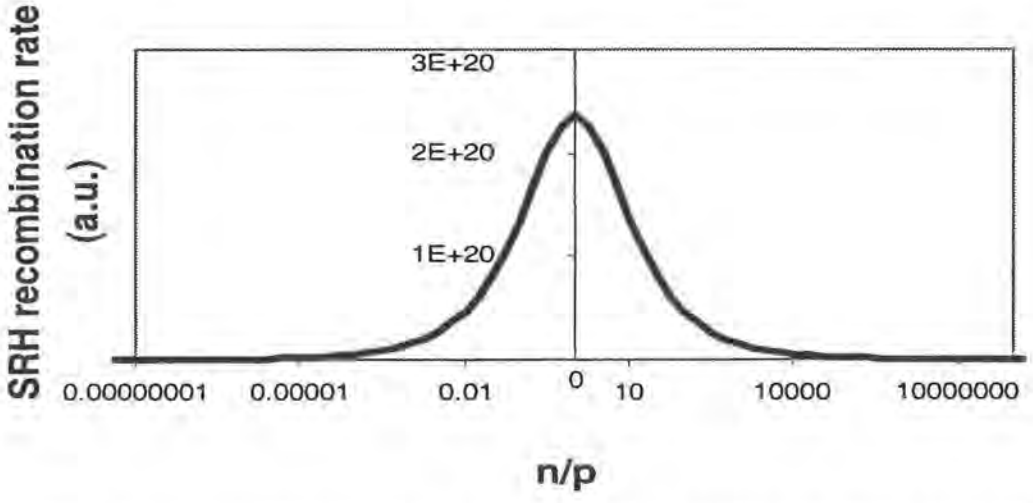


Fig. 4.10. SRH recombination rate as a function of the ratio n/p for a trap at mid-gap.

4.5.6. *Surface and grain boundary recombination*

According to Eq. 4.79, spatial variations in U_{SRH} can be caused by variations in n and p, or by spatial variations in the nature or number density of the dominant trap. A higher density of trap states shortens the electron and hole capture times. In real materials, defects are much more likely to occur at surfaces and at the interfaces between different crystal regions in a multicrystalline or heterostructured material. Localised states at surfaces and interfaces include both crystal defects due to broken bonds, and extrinsic impurities which are deposited from the external environment, or which are concentrated at interfaces during growth. In such cases the trap states responsible for recombination are concentrated in a two-dimensional rather than three-dimensional space and it is much more meaningful to express the recombination in terms of the trap density per unit area of the surface or interface, than per unit volume. The relevant quantity will be a recombination flux — the number of carriers recombining at the interface per unit area per unit time, rather than a volume recombination rate.

If a surface contains a density N_{s} traps per unit area, then within a very thin layer δ_{x} around the surface, the surface recombination flux will be

$$U_{\mathrm{s}}\delta x = \frac{n_{\mathrm{s}}p_{\mathrm{s}} - n_{\mathrm{i}}^2}{\frac{1}{S_{\mathrm{n}}}(p_{\mathrm{s}} + p_{\mathrm{t}}) + \frac{1}{S_{\mathrm{p}}}(n_{\mathrm{s}} + n_{\mathrm{t}})} \tag{4.82}$$

per unit area, where n_{s}, p_{s} are the electron and hole densities at the surface. S_{n} is the *surface recombination velocity* for electrons, defined by,

$$S_{\mathrm{n}} = B_{\mathrm{n}}N_{\mathrm{s}} \tag{4.83}$$

and S_{p} the surface recombination velocity for holes,

$$S_{\mathrm{p}} = B_{\mathrm{p}}N_{\mathrm{s}}\,. \tag{4.84}$$

In this definition, both S_{n} and S_{p} are directed towards the surface from the bulk.

In p type material, Eq. 4.82 reduces to

$$U_{\mathrm{s}}\delta x \approx S_{\mathrm{n}}(n_{\mathrm{s}} - n_0)\,.$$

This leakage of minority carriers to the surface results in a *surface recombination current.* The magnitude of the current can be obtained from the electron continuity equation. In the dark, at steady state, Eq. 4.1 requires that $\nabla \cdot J_n = qU_{\mathrm{n}}$. Integrating this across the interface layer, we find that

the electron current density has changed by

$$\Delta J = J\left(x_s + \frac{1}{2}\delta x\right) - J\left(x_s - \frac{1}{2}\delta x\right) = q\int_{x_s-\delta x}^{x_s+\delta x} U\,dx = qS_n(n_s - n_0)\,.$$

If the interface is a surface, then this condition determines the current density *at* the surface

$$J_n(x_s) = -qS_n(n_s - n_0)\,. \tag{4.85}$$

Similarly, the change in hole current density at an interface in n type material is given by

$$\Delta J_p = -q\int_{x_s-\delta x}^{x_s+\delta x} U\,dx = -qS_p(p_s - p_0)\,,$$

and the current density at the surface by

$$J_p(x_s) = qS_p(p_s - p_0)\,. \tag{4.86}$$

4.5.7. *Traps versus recombination centres*

When a carrier is captured by a trap, it may then be released or it may be annihilated by the capture of the opposite type of carrier. When the time for electron release by thermal activation

$$\frac{1}{\tau_{esc}} = B_n n_i e^{(E_t - E_i)/k_B T}$$

is much shorter than the time for capture of a hole

$$\frac{1}{\tau_{cp}} = B_p p\,,$$

the state can be considered an *electron trap* rather than a recombination centre. This may happen if the state is close in energy to the conduction band edge, or if the cross section for electron capture is much larger than for hole capture, as, for instance, for a positively polarised defect. Similarly, localised states which are close to the valence band or which have a higher cross section for hole capture act as *hole traps*. Traps serve to slow down the transport of carriers but they do not remove them.

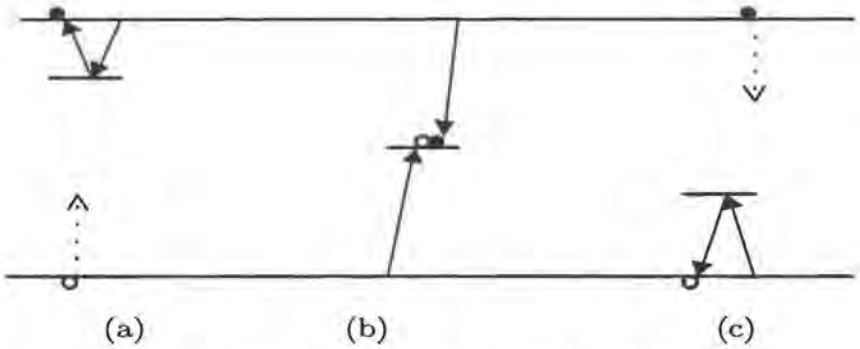


Fig. 4.11. (a) Electron trapping and detrapping; (b) electron-hole recombination; (c) hole trapping and detrapping.

4.6. Formulation of the Transport Problem

Finally we can proceed to solve the set of equations for n, p and ϕ set out in the introduction to this chapter. In one dimension,

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_{\mathrm{n}}}{\partial x} + G_{\mathrm{n}} - U_{\mathrm{n}} \tag{4.87}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial J_{\mathrm{p}}}{\partial x} + G_{\mathrm{p}} - U_{\mathrm{p}} \tag{4.88}$$

and

$$\frac{d^2\phi}{dx^2} = \frac{q}{\varepsilon_{\mathrm{s}}}(-\rho_{\mathrm{fixed}}(x) + n - p)\,. \tag{4.89}$$

For G_{n} and G_{p} we provide the photogeneration rate at that point. For band-to-band generation, $G = G_{\mathrm{n}} = G_{\mathrm{p}}$ is given by Eqs. 4.25 and 4.26. Photogeneration is the only generation process normally considered explicitly for solar cells; *thermal* generation of carriers is taken into account in each of the expressions for recombination. To evaluate $G(x)$ we need to know the absorption coefficient $\alpha(E, x)$, the reflectivity $R(E)$, and the incident photon flux density $b_{\mathrm{s}}(E)$.

For U_{n} or U_{p} we should, in general, provide the sum of all the recombination processes — radiative, Auger, and trap assisted — as a function of n, p, ϕ and x. For radiative recombination we also need to know the coefficient B_{rad} (Eqs. 4.57 and 4.58). For Auger recombination we need the Auger coefficients (Eq. 4.66). For SRH recombination in the bulk we need the density and position of the dominant trap state and the lifetimes for electron and hole capture (Eq. 4.79); at the surface we need the recombination velocities (Eqs. 4.83 and 4.84).

For the fixed charge density in Poisson's equation we need the doping profile

$$\rho_{\text{fixed}}(x) = (-N_{\text{a}}(x) + N_{\text{d}}(x)) \tag{4.90}$$

where N_{a} and N_{d} represent the densities of *ionised* acceptors and donors.

J_{n} and J_{p} are derived for a crystalline material in Chapter 3. According to Eqs. 3.65 and 3.66, the currents can be expressed in terms of the electron and hole quasi Fermi levels, in one dimension,

$$J_{\text{n}} = \mu_{\text{n}} n \frac{dE_{\text{F}_{\text{n}}}}{dx}$$

$$J_{\text{p}} = \mu_{\text{p}} p \frac{dE_{\text{F}_{\text{p}}}}{dx} .$$

For a crystalline material, these can be expressed in terms of n and p using the definitions of $E_{\text{F}_{\text{n}}}$, $E_{\text{F}_{\text{p}}}$ (Eqs. 3.31 and 3.34). For non-crystalline materials we must substitute alternative appropriate expressions for J_{n} and J_{p} such as Eq. 3.81.

When J_{n} and J_{p} are substituted into Eqs. 4.87 and 4.88, we have a set of three equations which can be solved to deliver n, p and ϕ as a function of x and t assuming that all material parameters are known. Boundary conditions are provided by the external electrical conditions and by the surface recombination conditions, and by the time dependence of the electrical and optical conditions for the transient problem.

4.6.1. *Comments on the transport problem*

Steady state solutions

Since solar cells operate in the steady state we are usually interested in the case where

$$\frac{\partial n}{\partial t} = \frac{\partial p}{\partial t} = 0$$

and solve the transport equations for a steady state illumination and electrical conditions. In the steady state, the electron and hole densities in each band and in localised states must be constant. This means that the generation and capture terms for the exchange of carriers between band and *trap* states (as opposed to recombination centres) must cancel out, and so trapping can be left out of the continuity equations. The consequence is

that the generation and recombination processes are all effectively band-to-band: $G_n = G_p$ and $U_n = U_p$. The trapped charge density will influence the solution through the fixed charge term in Poisson's equation.

In the transient case where traps are being filled or emptied, then the terms for capture and release from traps must be included and net generation rate for holes and electrons are nonzero.

Photon continuity

Until now, we have assumed that the density of photons at a point is determined solely by the absorption within the material. That is, we have assumed that the density of photons resulting from radiative recombination is negligible compared to the incident flux density. In materials with a high radiative efficiency, or under high illumination conditions, these 'recycled' photons may be significant. Then a further continuity equation is required, for photons, and a further unknown needs to be found, the photon flux density $b(x)$. The case of photon recycling is treated in Chapter 9.

4.6.2. *Transport equations in a crystal*

The most relevant context for conventional photovoltaics is a one-dimensionally varying crystalline material in the steady state. The quantity of interest is the net current at either of the terminals, $J_n + J_p$, and its dependence on applied bias and illumination. To find this we first need to set up the transport equations and solve for n, p and ϕ, as described above.

We proceed by combining the current equations for a compositionally invariant crystal (Eqs. 3.75 and 3.76) with the steady state continuity equations to obtain a pair of second order differential equations governing n and p. In one dimension

$$D_n \frac{d^2 n}{dx^2} + \mu_n F \frac{dn}{dx} + \mu_n n \frac{dF}{dx} - U + G = 0 \tag{4.91}$$

and

$$D_p \frac{d^2 p}{dx^2} + \mu_p F \frac{dp}{dx} + \mu_p p \frac{dF}{dx} - U + G = 0\,. \tag{4.92}$$

While G is usually a function of position only, U generally depends upon both carrier densities. This couples the transport equations for electrons and holes. In certain situations, however, U depends on n or p only and the equations can be solved independently. In particular, when

(i) one carrier type greatly exceeds the other and the recombination rate simplifies to the monomolecular form $U \approx (n - n_0)/\tau_n$ (Eq. 4.80), and
(ii) the electric field F is zero or constant, the transport equations for minority carriers simplify to the analytically soluble form:

$$\frac{d^2n}{dx^2} + \frac{qF}{k_B T}\frac{dn}{dx} - \frac{(n - n_0)}{L_n^2} + \frac{G(x)}{D_n} = 0 \tag{4.93}$$

for electrons in the p region, and

$$\frac{d^2p}{dx^2} + \frac{qF}{k_B T}\frac{dp}{dx} - \frac{(p - p_0)}{L_p^2} + \frac{G(x)}{D_p} = 0 \tag{4.94}$$

and for holes in the n region.

Here we have used the Einstein relation $\mu = \frac{qD}{k_B T}$ (Eq. 3.77) for μ and have introduced the *diffusion length* for electrons

$$L_n = \sqrt{\tau_n D_n} \tag{4.95}$$

and holes

$$L_p = \sqrt{\tau_p D_p}\,. \tag{4.96}$$

The diffusion length is a measure of the average distance a minority carrier will diffuse before recombining. In steady state, the one-dimensional diffusion equation becomes

$$\frac{(n - n_0)}{\tau_n} = D_n \frac{d^2n}{dx^2} \tag{4.97}$$

which has solutions for the *excess* minority carrier concentrations $(n - n_0)$ of the form $e^{\pm x/\sqrt{D\tau}}$. This makes $L = \sqrt{D\tau}$ a natural unit of length to characterise diffusion. Minority carrier concentration and current profiles have different behaviour depending on whether L is large or small compared to the layer width, and to the absorption length.

Finally, we solve the transport equations for $(n - n_0)$, $(p - p_0)$ subject to appropriate boundary conditions. In the next chapters, we will meet several examples where the simplified approach of Eqs. 4.93 and 4.94 can be used.

4.7. Summary

Photocurrent generation by a solar cell is linked to charge carrier generation and recombination by conservation of the numbers of electrons and

holes. Photogeneration is the primary carrier generation process in photovoltaic cells. For direct gap semiconductors, the absorption coefficient can be described mathematically by Fermi's Golden Rule, and near to the band edge it depends on photon energy approximately as $(E - E_g)^{1/2}$. For indirect gap materials, light absorption requires phonon absorption or emission, the absorption coefficient is generally smaller and rises more gradually, like $(E - E_g)^2$, near the band edge. Photogeneration formally requires the dissociation of an excited state and is not identical to light absorption in some compound or organic materials where this dissociation is incomplete.

The main recombination mechanisms are radiative, Auger, and trap assisted recombination in the bulk or at the surface. Radiative recombination is the relaxation of an electron across the band gap together with the emission of a photon. It is unavoidable in a light absorbing material and is most important when absorption is strong, in direct gap semiconductors. The radiative recombination rate is described by a generalised Planck formula and varies approximately as np. Auger recombination is the relaxation of a charge carrier to excite a second carrier to a higher energy state within the band. The rate depends on carrier densities to third order and is strongest when the charge carrier densities are high. Trap assisted recombination is a multiple step relaxation process, usually involving intermediate states in the band gap. The energy lost by relaxation is given up as heat. The rate depends upon the density and position of these intermediate states and on the relative densities of electrons and holes. It is strongest for deep traps when n and p are similar. In doped material it becomes linear and can be characterised by a single recombination time. Trap assisted recombination is particularly important at a surface on account of surface states. Surface recombination becomes linear in doped material and can be characterised by a surface recombination velocity which depends on the density of surface defects. In real materials trap assisted recombination is dominant. In the limit of perfect material, radiative recombination is dominant and Auger becomes important for low band gap and indirect gap materials.

Formulae for the generation and recombination rates can be used to set up transport equations for each of the charge carriers. Together with Poisson's equation, these form a set of coupled differential equations which can be solved for the charge carrier densities, currents and the electrostatic potential. In general, the problem is complex but can be greatly simplified through assuming a linear recombination rate and neglecting electric field, conditions which are valid in many device structures.

References

S. Adachi, *Physical Properties of III-V Semiconductor Compounds* (Chichester: Wiley, 1992).

F. Bassani and G. Pastori Parravicini, *Electronic States and Optical Transitions in Solids* (Oxford: Pergamon, 1975).

G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Editions de Physique, 1986)

A. de Vos, *Endoreversible Thermodynamics of Solar Energy Conversion* (Oxford University Press, 1992).

S.J. Fonash, *Solar Cell Device Physics* (New York, London: Academic, 1980).

M.A. Green, *Silicon Solar Cells: Advanced Principles and Practice* (Sydney: Centre for Photovoltaic Engineering, 1995).

A. Hagfeldt and M. Grätzel, "Molecular photovoltaics", *Acc. Chem. Res.* **33**, 269–277 (2000).

P.T. Landsberg, *Recombination in Semiconductors* (New York: Cambridge University Press, 1991).

J.I. Pankove, *Optical Processes in Semiconductors* (Englewood Cliffs: Prentice Hall, 1971).

M. Shur, *Physics of Semiconductor Devices* (Englewood Cliffs: Prentice Hall, 1990).

F. Stern, "Elementary theory of the optical properties of solids", *Solid State Physics* **15**, 300 (1963)

M.S. Tyagi, *Introduction to Semiconductor Materials and Devices* (Chichester: Wiley, 1991).

P. Wuerfel, "The chemical potential of radiation", *J. Phys. C — Solid State Physics* **15**, 3967–3985 (1982).

Chapter 7

Monocrystalline Solar Cells

7.1. Introduction: Principles of Cell Design

The most common solar cell design is a monocrystalline p–n junction. p–n homojunctions have the significant advantage over heterojunction designs that there is no material interface at the junction, and so losses due to interface states can be avoided. In this chapter we will discuss the design of p–n junction solar cells in the two best performing photovoltaic materials, monocrystalline silicon and gallium arsenide, illustrating the issues relevant to PV in weakly absorbing and strongly absorbing materials. We will look at the techniques which are used to improve performance in practice.

Efficient photovoltaic energy conversion requires efficient light absorption, efficient charge separation and efficient charge transport. From Chapter 6 we can translate these requirements into demands on the material and design parameters of a p–n junction solar cell.

- For good optical absorption, the optical depth of the device, $\alpha(E)(x_\mathrm{p} + x_\mathrm{n})$ should be high for energies E above the band gap, and the reflectivity of the surface, $R(E)$, should be small.
- For good charge separation, the built in bias V_bi should be large, demanding a high doping gradient across the junction (a large $N_\mathrm{a} \times N_\mathrm{d}$ product). Charge recombination in the junction region should be as slow as possible. And, as we have seen in Chapter 6, the junction should be located close to the surface for effective charge separation over a range of wavelengths.
- For efficient minority carrier transport the minority carrier lifetimes $(\tau_\mathrm{n}, \tau_\mathrm{p})$ and diffusion lengths $(L_\mathrm{n}, L_\mathrm{p})$ should be long and surface recombination velocities $(S_\mathrm{n}, S_\mathrm{p})$ should be small. Good majority carrier transport requires small series resistance R_s, and the shunt resistance R_sh

should be as high as possible to avoid leakage of carriers back across the junction.
- The band gap should be close to the optimum for the intended solar spectrum, $b_s(E)$.

Some of these demands are contradictory and a compromise must be found. For instance, high doping levels needed for a large built in bias are likely to reduce minority carrier lifetimes, and strong optical absorption will generally mean strong recombination.

7.2. Material and Design Issues

7.2.1. *Material dependent factors*

There are certain basic materials requirements. For high theoretical conversion efficiencies in AM1.5 ($\geq 30\%$) the band gap should lie in the range 1–1.6 eV; or in the range 0.7 to 2 eV for efficiencies over 20% (see Fig. 7.1). p and n type varieties of the semiconductor material must be available, and we need to be able to produce a junction of good quality. For acceptable photocurrent levels the quantum efficiency should be high over a broad range of wavelengths. That means that the diffusion lengths must be long compared to the absorption depth, and this places high demands on the crystal quality.

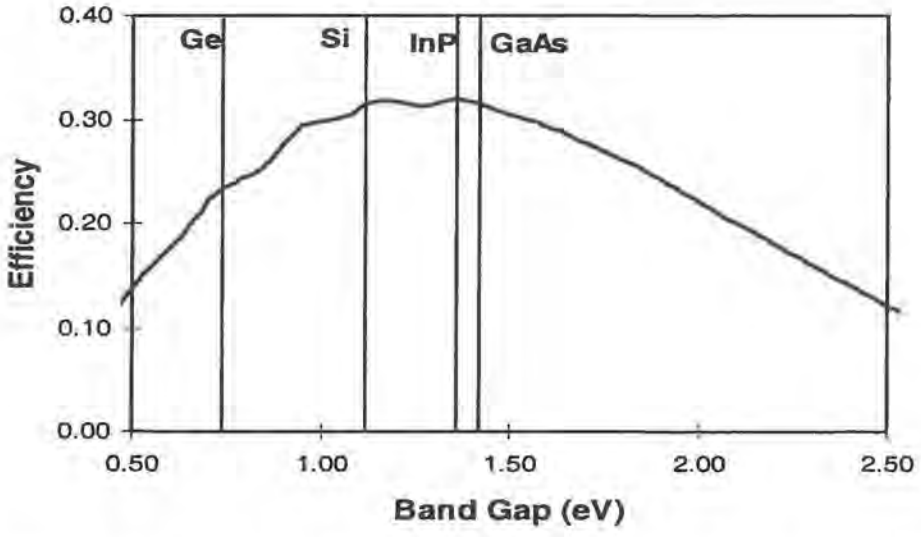


Fig. 7.1. Limiting efficiency as a function of band gap for the air mass 1.5 solar spectrum, and the band gaps of some common photovoltaic materials.

These requirements limit the range of materials which can be used. Figure 7.1 shows the band gaps of some common semiconductor materials compared to the limiting efficiency in a standard solar spectrum. Of the elemental semiconductors, only silicon has a suitable band gap. Germanium and selenium were used in early photovoltaic cells but the band gaps are too small. Of the compound semiconductors, III–V crystalline materials have been best developed. Gallium arsenide and indium phosphide have suitable band gaps. Several II–VI and I–III–V semiconductors are strong optical absorbers with suitable band gaps and some, notably CdTe and $CuInSe_2$, have been developed for photovoltaics. These materials cannot always be doped both n and p type, so that p–n homojunctions cannot be made. Amorphous silicon and related alloys have suitable band gaps, but diffusion lengths are too short compared to the absorption depth, and so again they cannot be used for p–n homojunction cells.

Materials can be prepared as single crystals (monocrystalline) or as polycrystalline or multicrystalline wafers. If grain sizes are smaller than device thickness, grain boundary effects are likely to dominate transport and diffusion lengths are likely to be short. In such cases, much of the analysis of Chapter 6 is not valid. We will distinguish 'multicrystalline' as material where the crystal grain size is comparable with or larger than the device thickness, and 'polycrystalline' or 'microcrystalline' as material where the grain size is much smaller. Devices made from multicrystalline material can be considered in same way as single crystal devices, treated below. Junctions in polycrystalline materials will be considered in the Chapter 8.

Intrinsic materials variables are the optical properties, the available dopants, the carrier mobilities, and the dominant recombination processes, all of which are related to the crystal structure. Process variables are the means of wafer and junction fabrication.

7.2.2. *Design factors*

The requirement of a good quality junction generally favours the p–n homojunction for the design. Once the material is chosen, a limited number of design factors remain which can be controlled. These include junction polarity; junction depth; doping levels; doping gradients; cell thickness; surface treatments; and contact design. The simple p–n junction can also be modified in a number of ways to improve performance. Strategies include compositional variations such as windows and graded layers, and light trapping structures.

The main focus of cell design depends upon the optical properties of the material. In a weakly absorbing material like silicon, strategies to maximise the optical absorption are important, and recombination at the rear surface needs to be avoided. In a highly absorbing material like gallium arsenide, good optical absorption is achieved easily but more attention must be paid to minimising recombination near the front surface and around the junction. Strategies for these two groups of materials are discussed below.

7.2.3. *General design features of p–n junction cells*

A number of general design features apply to p–n homojunction cells independent of the material. These are that:

(i) the thickness should exceed the absorption length, for efficient light absorption;
(ii) the junction should be shallow compared to both the diffusion length in the emitter and the absorption length, to avoid having a 'dead layer' at the front of the cell where light is absorbed unproductively, as noted in Chapter 6;
(iii) the emitter should be doped heavily, to improve conductivity to the fine metallic contacts on the front of the cell. Heavy emitter doping also allows the base to be doped lightly, which improves collection in the neutral base region without limiting the open-circuit voltage.
(iv) Reflection of light should be minimised. All crystalline semiconductor cells are treated with an anti-reflection (AR) coat. This is a thin layer of material of refractive index between that of the semiconductor and the air which couples light of preferred wavelengths into the semiconductor by matching the optical impedance. The AR coat material and thickness are chosen to maximise photocurrent generation for the relevant incident spectrum. (This is discussed further in Chapter 9.)

7.3. Silicon Material Properties

7.3.1. *Band structure and optical absorption*

Silicon is a group IV element which adopts the tetrahedral crystal structure at room temperature and pressure. In this arrangement every valence electron is involved in bonding and the material is a semiconductor. As discussed in Chapter 3, the band gap is indirect. That is, although the lowest point in the conduction band and the highest point in the valence band

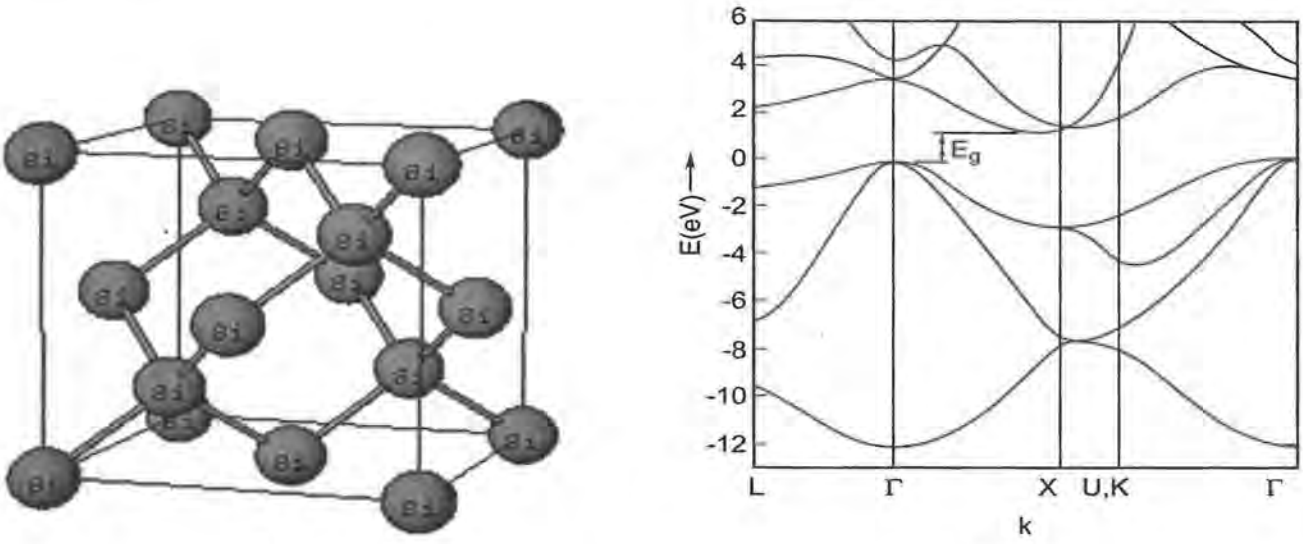


Fig. 7.2. (a) Tetrahedral crystal structure of crystalline silicon; (b) Band structure of silicon, showing energy levels as function of wavevector, $\mathbf{k}$. Levels are normally filled up to the energy $E = 0$, which marks the top of the valence band. Notice that the valence band maximum is at the Γ point, where $\mathbf{k} = 0$, while the conduction band minimum is near the X point, where $\mathbf{k}$ is directed along (100). The fundamental band gap lies between these extrema and is marked as E_g.

— the fundamental band gap — are separated by only 1.1 eV at room temperature, these points occur at different values of the crystal momentum and a phonon is needed as well as a photon in order for an electron to be promoted. This reduces the optical absorption compared to direct gap semiconductors. *Direct* optical transitions occur in silicon at photon energies above 3 eV though these are not useful for photovoltaics.

The band gap is fairly close to the optimum for solar energy conversion. At 1.1 eV, the theoretical limiting efficiency in AM1.5 is a few percent smaller than the maximum.

Silicon has a refractive index of around 3.4 and a natural reflectivity of about 40% over visible wavelengths. Introducing a single or multilayer anti-reflection coat reduces this to less than 5%.

7.3.2. *Doping*

Another criterion for electronic materials is whether they can be easily doped. Silicon can be readily doped n type, usually by the addition of pentavalent phosphorus impurity atoms, which introduce an extra loosely bound electron when substituted for a tetravalent silicon atom. p type

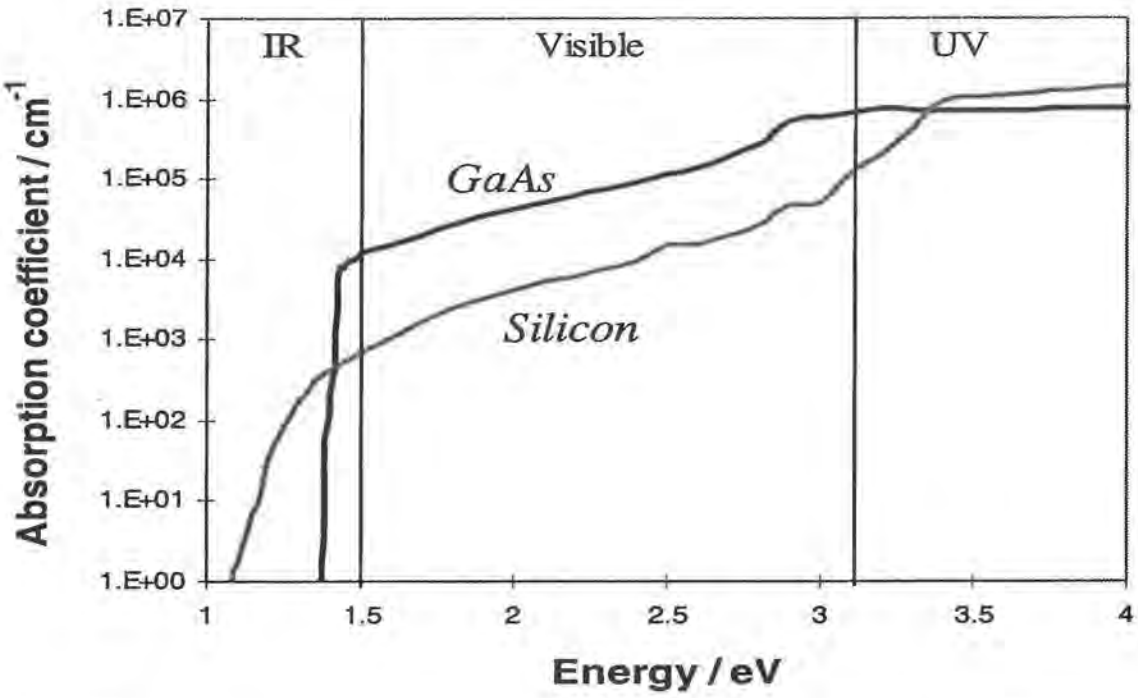


Fig. 7.3. Silicon absorption spectrum. The absorption spectrum of gallium arsenide is shown for comparison, to emphasise the low absorption in silicon due to its indirect band gap. The vertical lines denote the wavelengths 800 nm and 300 nm, the approximate limits of the visible part of the spectrum.

silicon may be produced by doping with trivalent boron, which introduced an electron vacancy, or hole, into the lattice when substituted for silicon. Addition of these impurity atoms generally degrades the material quality. However relatively high doping levels are needed in the emitter to reduce series resistance and increase V_{bi}, which in turn increases V_{oc}. In silicon the influence of doping on V_{oc} is limited by the shrinkage of the band gap, due to the introduction of tail states, at high doping levels. This means that V_{oc} cannot in fact exceed 81% of E_g at room temperature.

7.3.3. *Recombination*

We will see below that nearly all of the volume of a typical crystalline silicon cell is provided by p type silicon. The effect of recombination in the depletion region and emitter layer is rather low because of the low levels of photogeneration in those thin layers, and can usually be neglected. This leaves electron recombination in the p region as the dominant volume recombination process. This includes radiative, Auger and trap-assisted mechanisms.

The net recombination rate, U, will be the sum of all processes,

$$U = U_{\text{rad}} + U_{\text{Aug}} + U_{\text{SRH}} . \tag{7.1}$$

From Eq. 4.59 we expect the radiative recombination rate in p type material to follow

$$U_{\text{rad}} = \frac{n - n_0}{\tau_{\text{n,rad}}} \tag{7.2}$$

where the radiative lifetime $\tau_{\text{n,rad}}$ varies like $1/N_{\text{a}}$. Since silicon is an indirect band gap semiconductor, radiative recombination is slow and $\tau_{\text{n,rad}}$ is typically of the order of milliseconds, so this contribution is negligible.

Auger recombination can be important in silicon, if the carrier density is high. For electrons in p type material, from Eq. 4.67

$$U_{\text{Auger}} = A_{\text{n}} N_{\text{a}}^2 (n - n_0) \tag{7.3}$$

so that the lifetime varies like

$$\tau_{\text{n,Aug}} \propto N_{\text{a}}^{-2} . \tag{7.4}$$

Auger processes are most important in silicon at high doping densities (over 10^{17} cm^{-3}) and can be the dominant recombination pathway mechanism.

Trap assisted nonradiative processes dominate in all but the purest silicon. For Shockley Read Hall recombination through a single trap state we expect, from Eq. 4.80,

$$U_{\text{SRH}} = \frac{n - n_0}{\tau_{\text{n,SRH}}} \tag{7.5}$$

where the minority carrier lifetime $\tau_{\text{n,SRH}}$ depends upon the density of trap states, as well as their position within the band gap. From Eq. 4.69 we know that

$$\tau_{\text{n,SRH}} = \frac{1}{B_{\text{n}} N_{\text{t}}} \tag{7.6}$$

and there is evidence that $\tau_{\text{n,SRH}}$ decreases as doping increases, suggesting that the number of active traps increases with doping. Lifetimes vary from ms for very pure material to a fraction of a μs for high defect densities.

Combining the bulk recombination processes in Eq. 7.1 and noting that in the neutral p region all recombination mechanisms are linear in the excess electron density, $(n - n_0)$, we obtain the net electron lifetime τ_{n} such that

$$U = \frac{n - n_0}{\tau_{\text{n}}} \tag{7.7}$$

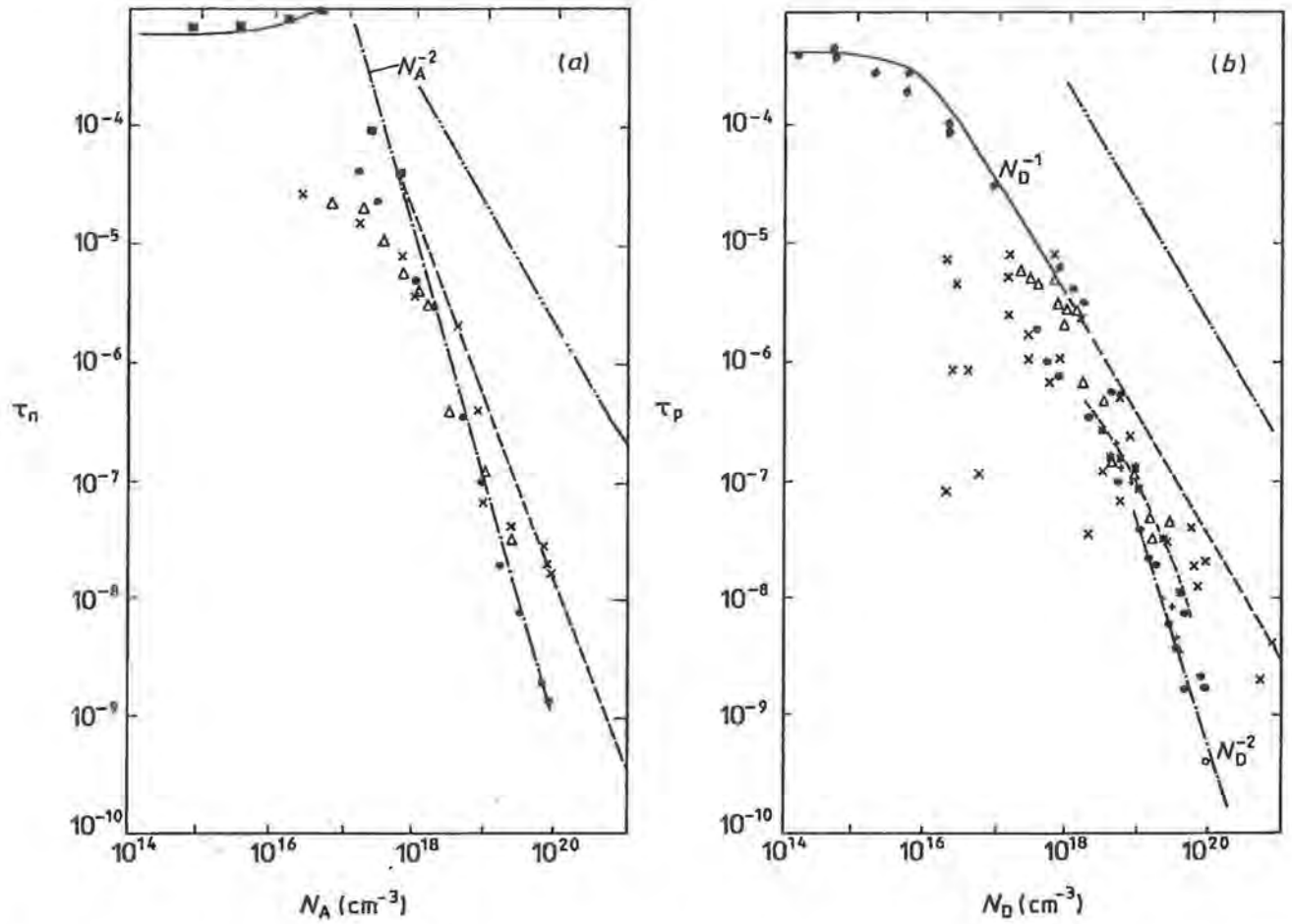


Fig. 7.4. Lifetime as a function of doping density in p type (left panel) and n type (right panel) silicon. At high doping densities the lifetime varies like the inverse square of the doping density, indicating that Auger processes are dominant. Lifetimes for other recombination processes are expected to vary like the simple inverse of the doping. From van Overstraeten and Mertens [van Overstraeten, 1986].

and

$$\frac{1}{\tau_n} = \frac{1}{\tau_{SRH}} + \frac{1}{\tau_{rad}} + \frac{1}{\tau_{Aug}}. \tag{7.8}$$

Different processes will have different temperature and doping dependence. In lightly doped p-type silicon at room temperature, SRH processes dominate, giving rise to a τ value of about 10 μs. Lifetime in lightly doped n type material is rather shorter, around 1 μs. In more heavily doped silicon or at higher temperatures Auger dominates. This is evident from the dependence of τ on N_a shown in Fig. 7.4. Radiative lifetimes in silicon are extremely long and never dominate the recombination in a practical solar cell.

As we saw in Chapter 6, surface recombination is important when bulk recombination is low, *i.e.*, when minority carrier diffusion lengths are long. For electrons in p type silicon, the surface recombination velocity at untreated surfaces, and at interfaces with metallic contacts, is in the range of 10^3–10^5 cm s^{-1} [Hovel, 1975]. When the surface is passivated with a layer of silicon dioxide, the oxide shields minority carriers from defects at the surface and reduces S_n to less than 100 cm s^{-1} [Green, 1995].

7.3.4. *Carrier transport*

In practice, the electron mobility in p type silicon tends to be higher than the hole mobility in n type silicon doped to the same level, and the minority electron diffusion length is longer. This is because the fundamental impurities in silicon are of acceptor type and so are more important in n type than p type material. It means that carrier collection is more efficient in p type material than in an n type layer of the same doping density and thickness, and so cells are generally designed as n–p cells with a thin n type emitter on top of a thick p type base.

Electron and hole mobilities are determined by the frequency of scattering events within the conduction or valence band. At low doping levels scattering is dominated by the silicon lattice. At high doping mobility decreases because the impurity atoms generate scattering centres. To a first approximation the mobility does not depend on whether doping is n type or p type. In silicon, electrons have a mobility of around 1500 cm^2 V^{-1} s^{-1} at low doping, falling to around 70 cm^2 V^{-1} s^{-1} at impurity concentrations of over 10^{19} cm^{-3}, while holes have a low-doping mobility of 500 cm^2 V^{-1} s^{-1}, falling to around 50 cm^2 V^{-1} s^{-1}. The corresponding diffusion constants in lightly doped material are 30–40 cm^2 s^{-1} for electrons and around 10 cm^2 s^{-1} for holes [Shur, 1990; Green, 1995].

Minority carrier diffusion lengths are defined from $L = \sqrt{\tau D}$ (Eq. 4.95) where τ refers to the *net* recombination and do not distinguish between processes. For a commercial silicon cell, τ_n is several μs and L_n is around 100 μm.

The resistivity of any semiconductor material, ρ, depends upon the level of doping. From the definition of conductivity (Eq. 3.41),

$$\rho = \frac{1}{\sigma} = \frac{1}{q(\mu_n n + \mu_p p)} . \tag{7.9}$$

In doped material, resistivity will therefore be dominated by the majority carrier density and mobility. ρ varies roughly like $(10^{16}/N_a)$ Ohm-cm in p type silicon, and is a factor of three smaller in n type. With an acceptor concentration of 10^{16} cm^{-3}, a p region several hundred μm thick has a sheet resistance of around 10^{-1} Ohm-cm^2 which is unimportant for the current densities under standard solar illumination. However, series resistance effects can be more serious in the emitter, where the funnelling of current from a large semiconductor area into a very small contact area creates high current densities. This is one reason why the emitter is highly doped relative to the base. When the contact resistance is included the equivalent series sheet resistance for a commercial silicon cell is typically 20–50 Ohm-cm^2.

Silicon materials properties are discussed in detail by [Green, 1995; Bube 1998; van Overstraeten, 1986]; and many semiconductor texbooks.

7.4. Silicon Solar Cell Design

7.4.1. *Basic silicon solar cell*

A typical silicon solar cells is an n–p junction made in a wafer of p type silicon a few hundred microns thick and around 100 cm^2 in area. The p type wafer forms the *base* of the cell and is thick (300–500 μm) in order to absorb as much light as possible, and lightly doped ($\sim 10^{16}$ cm^{-3}) to improve diffusion lengths. The n type emitter is created by dopant diffusion and is heavily doped ($\sim 10^{19}$ cm^{-3}) to reduce sheet series resistance. This layer should be thin to allow as much light as possible to pass through to the base, but thick enough to keep series resistance reasonably low. Carrier collection from the emitter is negligible because of high recombination in this heavily doped layer. The front surface is anti-reflection coated and both front and back surfaces are contacted before encapsulation in a glass covering.

7.4.2. *Cell fabrication*

Single crystal silicon may be grown by a number of methods. In the common *Czochralski* process a single crystal is drawn slowly out of a melt. In the *float zone* process a single crystal is gradually formed from a polycrystalline rod by passing a molten zone through it. This is more costly but produces higher purity material. In either case the dopant (usually boron) is introduced during growth to produce a p type crystal. The solid crystal is sliced into wafers and etched to smooth the rough surfaces.

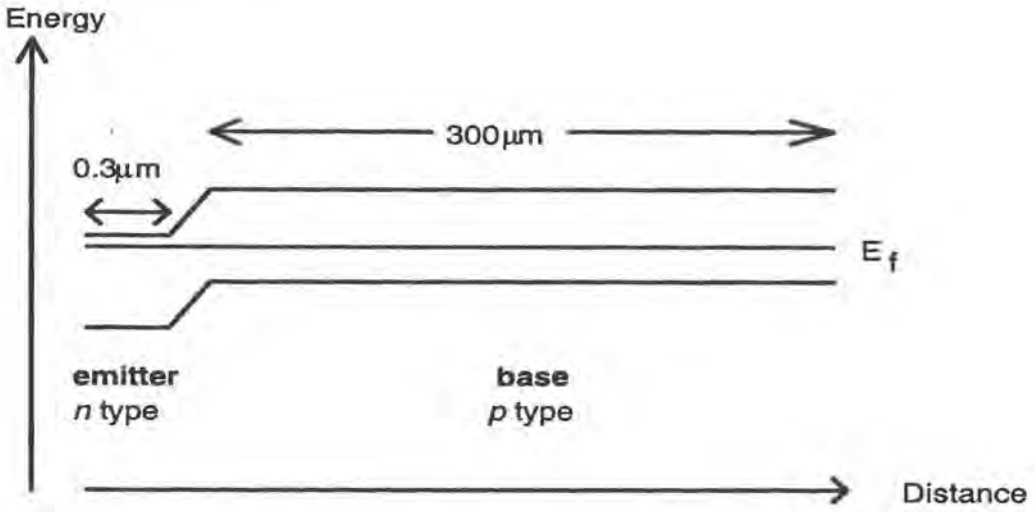


Fig. 7.5. Energy band diagram of a n–p junction in silicon.

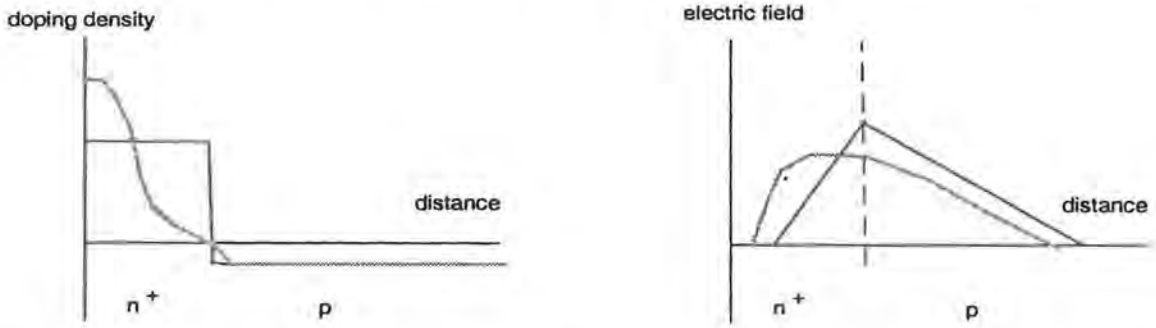


Fig. 7.6. Doping profiles for an abrupt (black line) and diffused (grey line) junction.

The junction is prepared by diffusing the n type dopant — usually phosphorus — on to the p type wafer. Phosphorus may be deposited either from the vapour phase by exposure to nitrogen gas bearing $POCl_3$; from solid phase, for example by chemical vapour deposition of phosphorus oxide; or directly by ion implantation. The latter method allows greater control of the doping profile but is more costly. Note that the dopant profile in the diffused n layer will not be uniform and the junction not abrupt, unlike the ideal p–n junction in Chapter 6. This has the consequence that the electric fields extend further from the junction (and the depletion approximation is less accurate). The diffused doping profile may be modelled with an error function.

Multicrystalline silicon, which is used in most commercial silicon cells, is made by a variety of methods such as casting and ribbon growth. (For

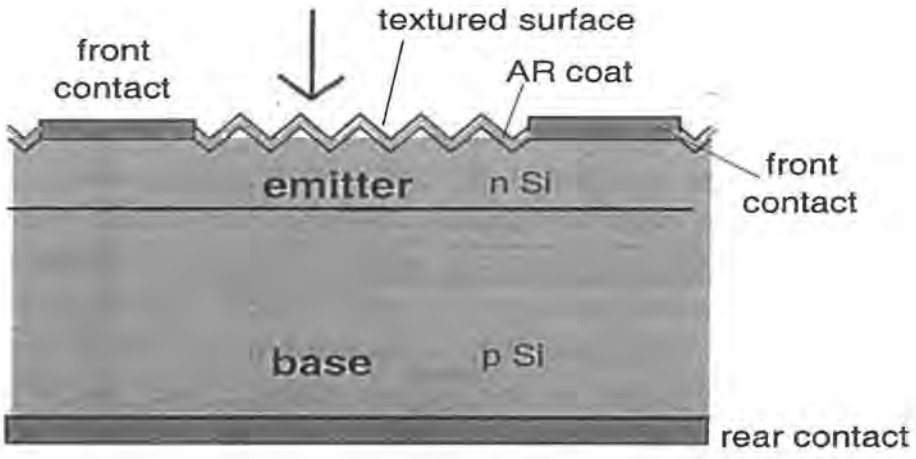


Fig. 7.7. Layer structure of basic silicon cell.

details see Green [Green, 1995].) The relatively large sizes of the grains (0.1–10 cm) mean that moderately efficient devices can be prepared from multicrystalline material using techniques similar to those used for monocrystalline silicon.

The front surface is usually textured to reduce reflectivity and an anti-reflection coating is deposited from liquid or vapour phase added. For silicon the AR coating should have a refractive index of around 2 and thickness of 80–100 nm. Suitable materials for silicon are tantalum oxide (Ta_2O_5), titania (TiO_2) and silicon nitride (Si_3Ni_4).

The rear surface is doped more heavily to create a back surface field, which helps to reduce the loss of carriers through surface recombination. This is discussed below.

Finally the front and back contacts are added. In the early silicon cells, aluminium was used as the rear contact. In large scale production, AR coat, front and back contacts are usually deposited by screen printing and then fired. Screen printing of contacts is cheap but obscures a relatively large area of the cell and degrades conductivity.

7.4.3. *Optimisation of silicon solar cell design*

Figure 7.8 illustrates the absorption and recombination profiles in a typical silicon cell, calculated using the p–n junction theory of Chapter 6 and the materials parameters in Table 7.1. From the graphs we can make the following observations:

- absorption of light close to the band gap (near infrared) is poor

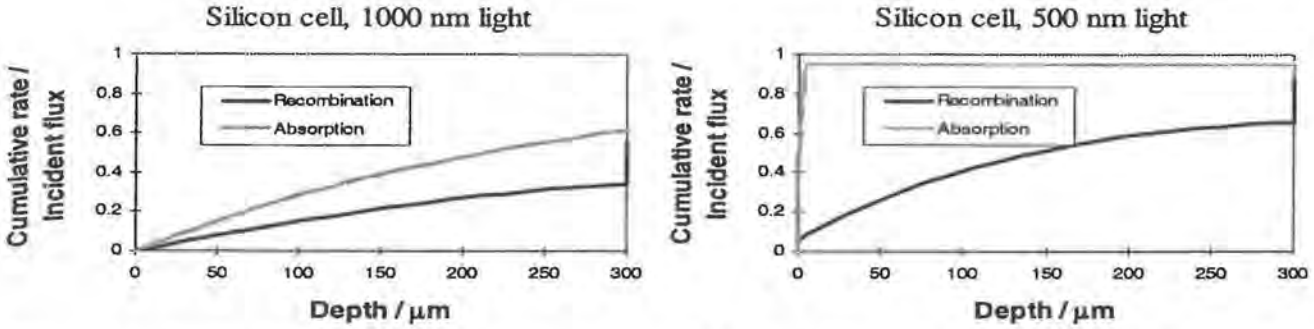
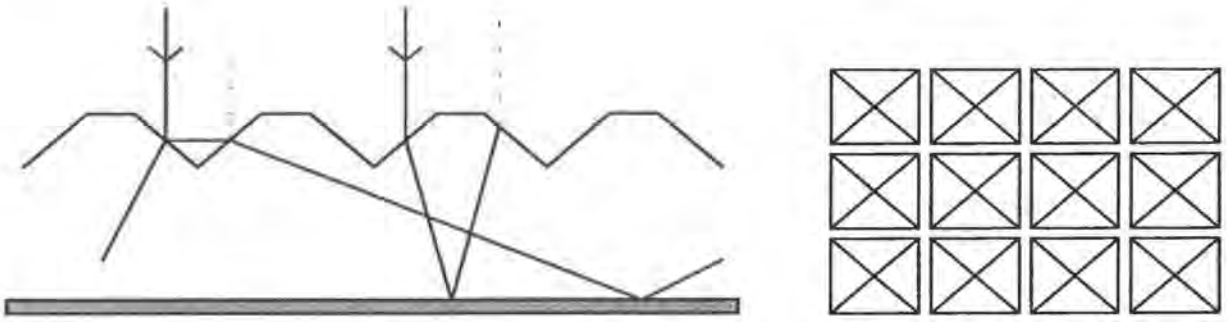


Fig. 7.8. Simulated cumulative absorption and recombination rates for a silicon solar cell with the parameters in Table 7.1 under monochromatic light at two different wavelengths at a bias close to open circuit. In each graph the cumulative absorption and recombination, normalised to the total incident flux, are plotted against depth through the cell. For a good cell at short circuit, recombination should be zero and cumulative absorption should be close to one. At open circuit (illustrated) the final values of recombination and absorption are equal.

Table 7.1. Typical silicon material and cell parameters, used for the calculations in Fig. 7.8.

	EMITTER (N TYPE)	BASE (P TYPE)
thickness/μm	0.5	300
doping/cm^{-3}	1×10^{19}	1×10^{16}
minority carrier diffusion constant/cm^2 s^{-1}	2	40
minority carrier lifetime/s	1×10^{-6}	5×10^{-6}
minority carrier diffusion length/μm	14	140
surface recombination velocity/cm s^{-1}	10000	10000
reflectivity	0.05	
absorption: at 500 nm/cm^{-1}	15000	
at 1 μm/cm^{-1}	35	

- bulk recombination in the p region is the most important recombination process
- rear surface recombination is important, particularly for photogeneration by red and infrared light
- front surface recombination and recombination in the junction region are relatively unimportant for photogeneration by long wavelengths



(a) (b)

Fig. 7.9. Surface texturing: (a) shows how the refraction of incident light rays at the textured surface increases the optical path length relative to light normally incident on a plane surface; (b) shows a plan view of regular pyramids.

To improve the performance of the cell it is necessary to maximise the absorption of red light, minimise recombination at the rear surface, and minimise series resistance. Bulk recombination is determined mainly by the method of wafer growth, and for good quality silicon it is already as low as can be expected.

Thus the main challenges in crystalline silicon cell design are to:

- Maximise absorption
- Minimise rear surface recombination
- Minimise series resistance

These are discussed below.

7.4.4. *Strategies to enhance absorption*

- Texturing of front surface. This reduces the net reflection of light and increases the optical depth of the cell. Texturing can be achieved by treating with an anisotropic chemical etc. which acts preferentially along the (111) crystal planes and leaves a pattern of pyramids on the surface. Regular pyramids can be produced on a monocrystalline surface by photolithographic definition. Light trapping is improved by using *inverted* pyramids, which improve the total internal reflection of light reflected from the back surface, by asymmetric pyramids, or by texturing the rear surface. Light trapping strategies are discussed further in Chapter 9.
- Optimisation of contacts. Shading of the front surface by metal contacts reduces the surface area available to the incident light by as much as 10%. Reduced contact area increases the available surface area but increases

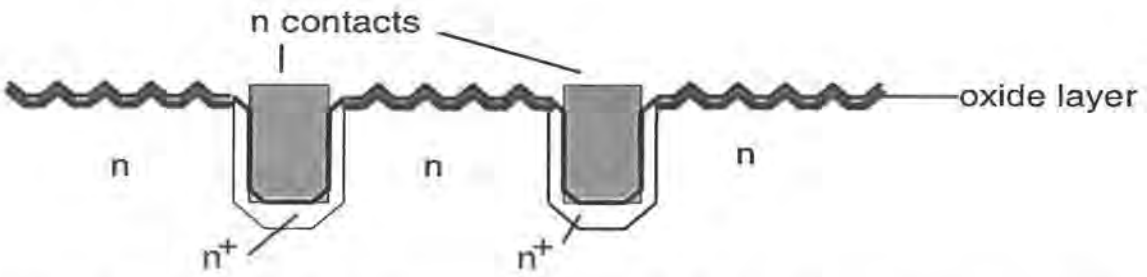


Fig. 7.10. Buried contacts. Differential n^+ doping in the contact grooves reduces series resistance.

the resistance either in the emitter, if the contacts are too sparse, or in the metal, if the fingers are too narrow. The optimum arrangement is a grid of narrow, dense, highly conducting fingers. One solution is to use narrow, deep contacts partly buried in the surface of the cell. This may reduce shading to a fraction of a percent of the surface. By embedding the contacts in the semiconductor, a larger contact area can be achieved without increasing the surface shading. The grooves are created by laser or mechanical etching, and are doped more heavily than the main emitter to improve conductivity. However, the large scale preparation of such contacts is more costly than screen printing.

7.4.5. *Strategies to reduce surface recombination*

- Back surface field. A more heavily doped layer is formed at the back surface of the p type base by alloying with aluminium or by diffusion. This introduces a p^+–p junction and presents a potential barrier to the minority electrons. This *back surface field* reflects electrons and reduces the effective rear surface recombination velocity, to less than 100 cm s^{-1}. (See Box 7.1). The extra p^+–p junction also adds to the built in bias of the cell, and may enhance V_{oc} [Hovel, 1975]. Front surface fields have also been used in some cell designs, but are less effective since the ratio of doping levels will be smaller.

Box 7.1. Calculation of effective surface recombination velocity with a back surface field

When a back surface field layer is present in p type material, the change in doping introduces a step in the conduction band edge of height $\Delta E_c =$

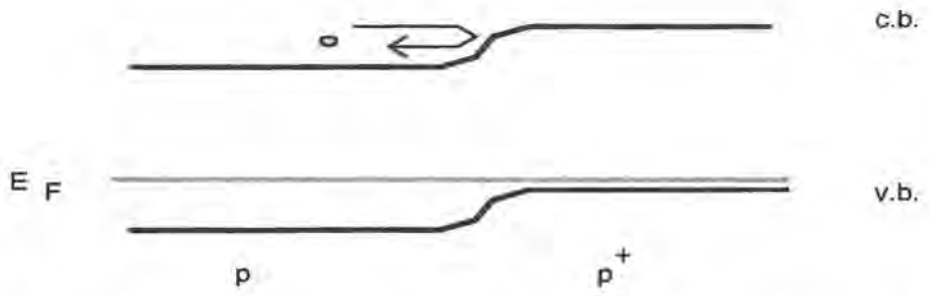


Fig. 7.11. Back surface field in p type material. Electrons are reflected from the doping barrier, effectively reducing the surface recombination velocity.

$k_{\mathrm{B}}T\ln(\frac{N_{\mathrm{a}+}}{N_{\mathrm{a}}})$. Electron concentration is reduced on the p^+ side of the low-high junction ($x_{\mathrm{j}+}$) relative to the main p side ($x_{\mathrm{j}-}$) by the ratio of the doping levels, $N_{\mathrm{a}}/N_{\mathrm{a}+}$. This follows from continuity of the Fermi level across the junction. If $x = x_{\mathrm{j}}$ at the p–p^+ junction we have

$$\Delta n(x_{\mathrm{j}-}) = \Delta n(x_{\mathrm{j}+}) \times e^{\Delta \mathrm{E_c}/\mathrm{k_B T}} = \Delta n(x_{\mathrm{j}+}) \frac{N_{\mathrm{a}+}}{N_{\mathrm{a}}}$$

for the excess electron concentration Δn. Current continuity at the junction provides a second boundary condition,

$$D_{\mathrm{n}-} \frac{dn(x_{\mathrm{j}-})}{dx} = D_{\mathrm{n}+} \frac{dn(x_{\mathrm{j}+})}{dx}$$

where $D_{\mathrm{n}-}$ and $D_{\mathrm{n}+}$ are the diffusion constants in the p and p^+ layers. Now, if we define an effective surface recombination velocity at the junction, S_{eff}, in analogy with Eq. 4.85,

$$S_{\mathrm{eff}} = -\frac{D_{\mathrm{n}-}}{\Delta n(x_{\mathrm{j}-})} \frac{dn(x_{\mathrm{j}-})}{dx} ,$$

it follows that

$$S_{\mathrm{eff}} = -\frac{D_{\mathrm{n}+}}{\Delta n(x_{\mathrm{j}+})} \frac{N_{\mathrm{a}}}{N_{\mathrm{a}+}} \frac{dn(x_{\mathrm{j}+})}{dx} .$$

The electron density in the highly doped layer is fixed by the boundary condition at the rear surface

$$-D_{\mathrm{n}+} \frac{dn(x_{\mathrm{j}} + x_{\mathrm{b}})}{dx} = S_{\mathrm{n}+}(\Delta n(x_{\mathrm{j}} + x_{\mathrm{b}})) ,$$

where x_{b} is the thickness of the highly doped layer and $S_{\mathrm{n}+}$ is the rear surface recombination velocity within the highly doped layer. If recombination

is linear in that layer and characterised by a diffusion length L_{n+} we can solve for the ratio $\frac{\frac{dn}{dx}(x_{j+})}{\Delta n(x_{j+})}$ and find, in the dark, that

$$S_{\text{eff}} = \frac{D_{n+}}{L_{n+}} \frac{N_a}{N_{a+}} \left\{ \frac{\frac{S_{n+}L_{n+}}{D_{n+}} \cosh \frac{x_b}{L_{n+}} + \sinh \frac{x_b}{L_{n+}}}{\frac{S_{n+}L_{n+}}{D_{n+}} \sinh \frac{x_b}{L_{n+}} + \cosh \frac{x_b}{L_{n+}}} \right\}. \tag{7.10}$$

If the highly doped layer is thin enough, Eq. 7.10 shows that S_{eff} will be reduced by the ratio of doping levels.

- Passivation of front surface with thin oxide coating. The high surface recombination velocity at a free silicon surface tends to create a dead layer, where photogenerated carriers are not collected, at the surface of an unpassivated cell. Oxidising the surface creates a thin layer of the wide band gap insulator, silicon dioxide, which prevents carriers from reaching the surface and hence reduces the effective surface recombination velocity. The interface between silicon and silicon dioxide is much less defective than a free silicon surface. This reduces the loss of carriers in the emitter through surface recombination, and improves the response to blue light.
- Use of point contacts at rear. Since the silicon–metal interface is more defective than the silicon–silicon dioxide interface, rear surface recombination can be reduced by contacting only part of the rear p layer with metal, using 'point' contacts. The rest of the surface can then be passivated with oxide, and the overall surface recombination losses greatly reduced. In order to avoid problems with series resistance, the region of semiconductor close to the point contacts is differentially doped p^+. This innovation followed from the rear point contact solar cell, described below.

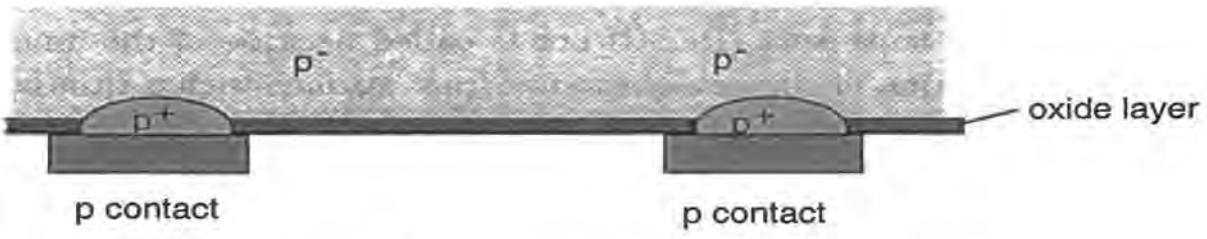


Fig. 7.12. Rear point contacts in p type material.

7.4.6. *Strategies to reduce series resistance*

- Optimisation of the n region doping. Reduced doping improves collection from the n region, giving a better response to blue light. Increased n doping increases V_{bi} and reduces series resistance, although very high n doping is unhelpful for increasing V_{oc} because of Auger recombination and band gap narrowing.
- Differential doping of the area around the contacts. This is achieved by exposing the areas to be contacted to dopant rich gases before deposition of the contacts. For point and grid contacts, the current density through the material close to the contacted area will be high. Doping this volume heavily reduces the losses to series resistance.
- Narrow but deep fingers in front contact, as above. The high aspect ratio reduces surface area blocked by contacts without reducing finger cross sectional area, and the relatively high contact area between fingers and semiconductor reduces the current density at the contact.

7.4.7. *Evolution of silicon solar cell design*

Let's look at how these strategies have been incorporated into the design of silicon solar cells. A summary of crystalline silicon cell performance data is given in Table 7.2.

Black cells

Typical 'black cell' designs (so called because of their almost zero reflectivity) were developed in the early 1980s and exhibited efficiencies of up to 17%. Black cells incorporated the innovation of the surface texturing as well as the features of the basic cell described above.

Passivated emitter cells

Passivated emitter solar cells (PESC) are so called because of the innovation of the passivation of the non-contacted front surface with a thin layer of silicon dioxide. Improvements such as these make it worthwhile using more expensive float zone produced silicon, which is better quality than Czochralski and has a longer diffusion length. The PESC cell was designed at the University of New South Wales and achieved an efficiency of 20% in 1985.

Table 7.2. Record efficiencies for various silicon cells and the best GaAs cell under standard test conditions (Air Mass 1.5 direct, 25°, 1000 W m^{-2}). [Green *et al.*, 2001; Green, 1995.]

DESCRIPTION	WHERE DEVELOPED	AREA (CM^2)	V_{OC} (V)	J_{SC} (MA/CM^2)	FF	EFFICIENCY (%)
monocrystalline Si cells:						
PESC	UNSW					20.8
Buried contact	UNSW	12				21.3
Back point contact	Stanford	10				22.3
PERL	UNSW	4.0	0.696	42.0	83.6	24.4
polycrystalline Si	UNSW/Eurosolare	1.0	0.628	36.2	78.5	19.8
commercial Si		100				13
monocrystalline GaAs	Kopin	3.91	1.022	28.2	87.1	25.1

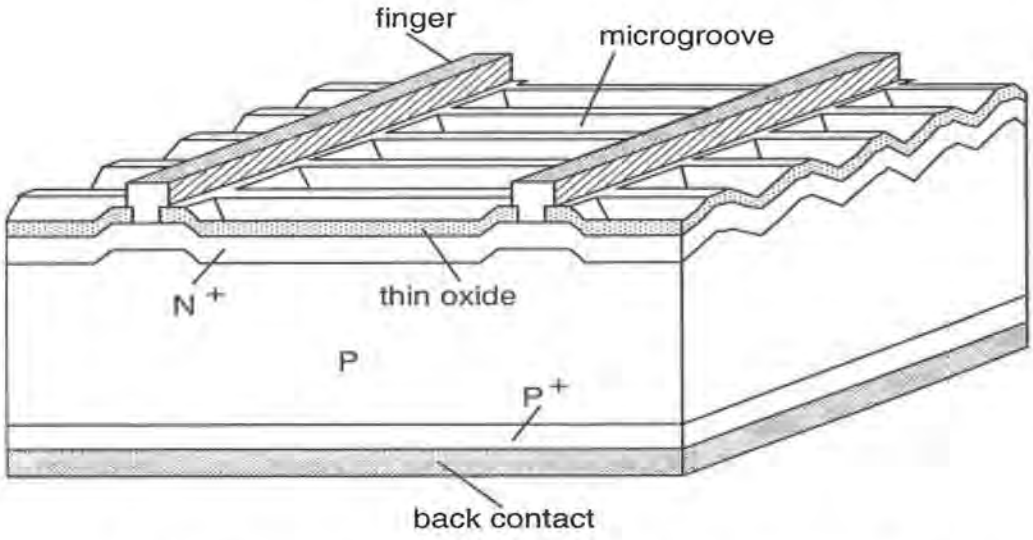


Fig. 7.13. Passivated emitter solar cell (PESC) [Green, 1995].

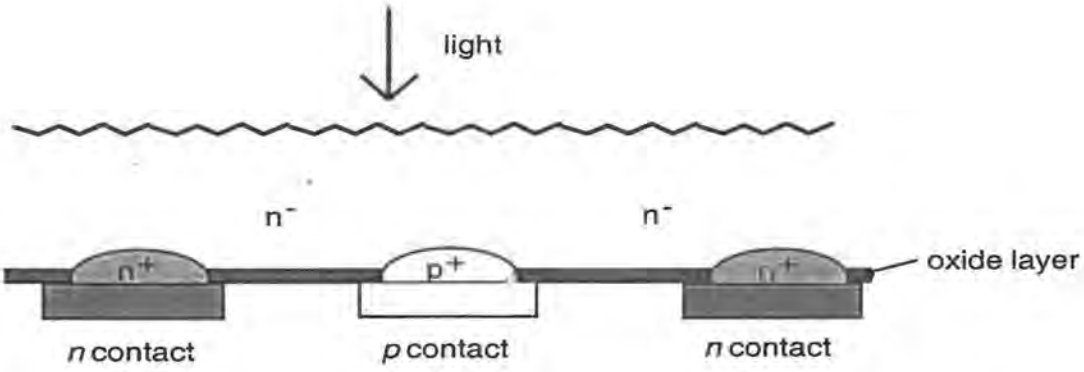


Fig. 7.14. Schematic cross section of rear point contact cell.

Rear point contact cell

By placing both the n and p contacts on the rear of the cell, this design eliminates shading losses entirely. This cell was introduced at Stanford in 1992, with an efficiency of 22%. The original design was intended for use in concentrators [Sinton, 1986]. The cell is made from lightly doped n type silicon with heavily doped n and p type regions close to point contacts on the rear surface. The front surface is passivated and textured as usual. The cell is thin (100 μm) and is intended to operate at high injection levels, so light trapping is important.

Extremely high purity material is needed, because photogenerated carriers have to *diffuse* to the rear of the cell. Small space charge regions will

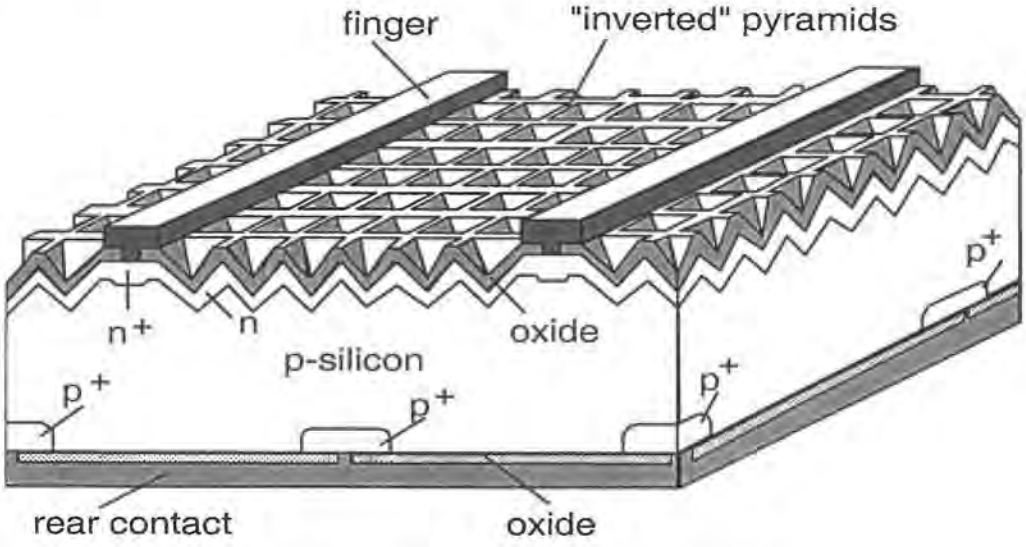


Fig. 7.15. PERL cell [Green, 1995].

develop at the rear of the cell between contacts of opposite polarity rather than at the front. Another difficulty is the risk of shorting out between contacts of opposite polarity on a single surface.

PERL cell

The passivated emitter, rear locally diffused (PERL) solar cell was developed at UNSW, with an efficiency of 24% in 1994 [Zhao, 1994]. This design exploits the advantage of point contacts in reducing recombination at the rear surface. It has the following features:

- Rear point contacts reduce the area of the semiconductor-metal interface, where recombination is high, so that most of the rear surface may be contacted with oxide.
- Grooved front contacts as with the passivated emitter solar cell.
- Differential heavy doping of n layer near contacts
- Surface texturing using inverted pyramids

7.4.8. *Future directions in silicon cell design*

The performance of silicon solar cells is now fairly close to the theoretical maximum of 29%. Continuing refinements to the design, mainly aimed at reducing shading and series resistance losses, may increase efficiencies of lab cells to 26% or 27% in AM1.5. The main challenges are now in improving

cell production techniques in order to mass-produce efficient cells more cheaply. For example, with the buried contact cell, efforts have focused on producing grooves more cheaply, for example by mechanical etching.

A quite different direction is the thin film microcrystalline silicon cell. Here the objective is to reduce bulk recombination losses without losing absorption and effective light trapping is required. This design works in the 'high injection' limit where different physics applies. It is considered briefly in Chapter 8.

7.4.9. *Alternatives to silicon*

Silicon is not an ideal solar cell material, for two main reasons. One is that its band gap (1.1 eV) is smaller than the optimum (1.4 eV) for terrestrial solar energy conversion. The other is that since its absorption coefficient is low, a relatively thick layer of silicon is needed (in conventional silicon designs) to absorb sunlight effectively. The significant requirement for high purity silicon increases the cost, as well as the weight, of the cell. Another consideration is the temperature dependence of efficiency which makes silicon less suitable for applications under concentrated light and in space. Therefore a number of alternative semiconductor materials have been developed for photovoltaic applications. Cell designs aimed at reducing costs by using less or less pure semiconductor material are considered in Chapter 8. In the next section we will briefly consider materials, from the III–V group of compound semiconductors, with more favourable materials properties for high efficiency single crystal cells.

7.5. III–V Semiconductor Material Properties

7.5.1. *III–V semiconductor band structure and optical absorption*

A III–V semiconductor is an alloy containing equal numbers of atoms from groups III and V in the periodic table. The group III atom contributes three valence electrons to bonding and the group V element contributes five. For many compounds the atoms arrange themselves into the *zincblende* crystal structure — two interlocking face centred cubic lattices — where each atom forms four bonds to neighbouring atoms of the other type. All of the valence electrons are used up in bonding, and, as in the case of silicon, the lowest energy configuration for the crystal is a semiconductor where

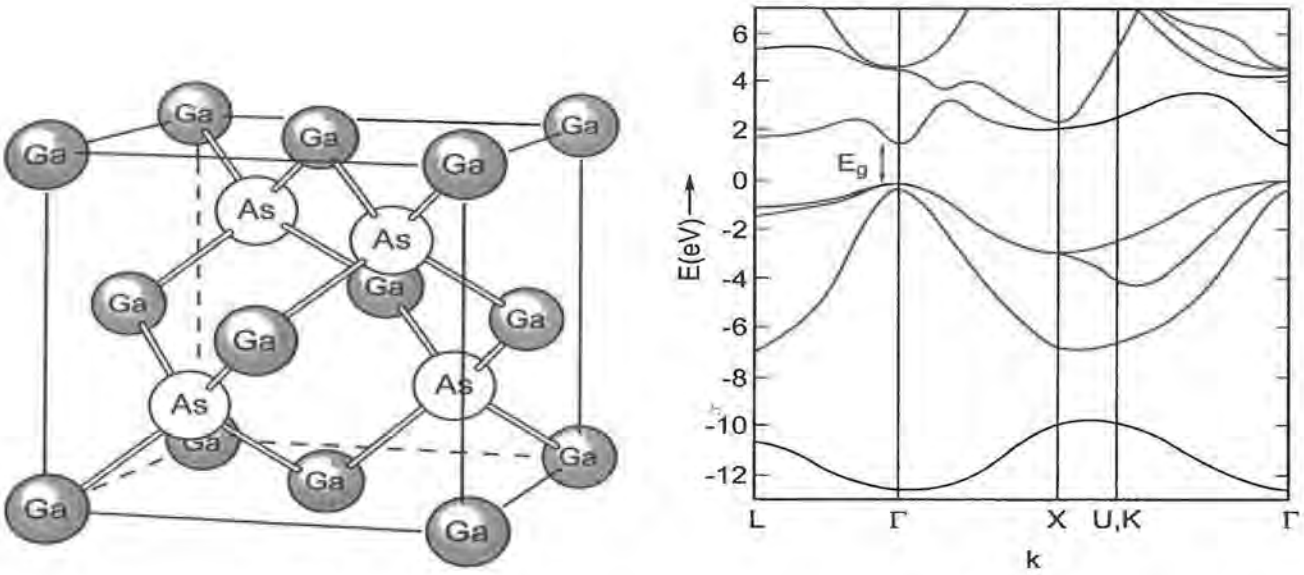


Fig. 7.16. GaAs crystal structure and band structure. Notice that the conduction band maximum occurs at the same point as the valence band minimum (the Γ point, where $k = 0$, giving a direct band gap).

there is a band gap between the (normally empty) conduction band and the (normally filled) valence band.

Relative to silicon, III–V's have several advantages as electronic materials. One is the possibility of varying the crystal composition by replacing some of the group III atoms with another group III element in order to vary the band gap in a controlled way. Another is that for many compositions these materials are *direct* gap semiconductors, and so are much more effective optical absorbers (see Fig. 7.3).

III–Vs have been widely developed for applications in optoelectronics. They are grown by a number of epitaxial techniques such as liquid phase epitaxy (LPE), molecular beam epitaxy (MBE), metal-organic chemical vapour deposition (MOCVD) and metal-organic vapour phase epitaxy (MOVPE) which allow minute control of the composition and layer thickness. Epitaxial growth has been developed in particular so that heterostructures — layered structures of materials of different band gap which enable spatial confinement of carriers for applications such as lasers — may be fabricated. The best understood and most widely used of these III–V semiconductors is *gallium arsenide* (GaAs). It is also the most suitable for solar energy conversion. Other relevant materials are the binary alloys indium phosphide (InP), gallium antimonide (GaSb) and ternary alloys such

as aluminium gallium arsenide ($Al_xGa_{1-x}As$), where a fraction x of the gallium atoms in GaAs have been replaced by aluminium atoms, indium gallium phosphide ($In_xGa_{1-x}P$) and indium gallium arsenide ($In_xGa_{1-x}As$). Indium phosphide has a suitable band gap for photovoltaic conversion and is particularly attractive for space applications on account of its resistance to degradation under radiation, or 'radiation hardness'. Properties of GaAs are covered elsewhere [Lush, 1986; Adachi, 1992; Bube, 1998].

7.5.2. *Gallium arsenide*

GaAs has a band gap of 1.42 eV at room temperature. This is close to the optimum for the standard solar spectrum and means a conversion efficiency of 31% is theoretically possible. It is a direct gap material and absorbs strongly above its band gap. Over visible wavelengths the absorption coefficient of GaAs is about ten times that of silicon (see Fig. 7.3), and only a few μm rather than hundreds of μm are needed for the active region of the solar cell. This most important for space applications where the priority is to reduce the cell weight.

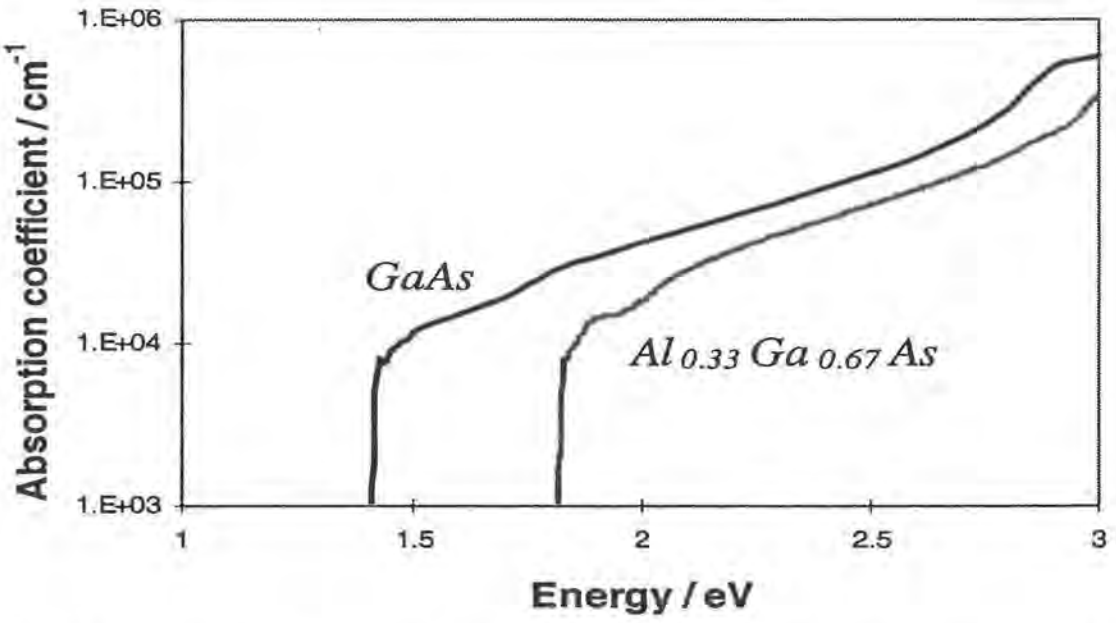


Fig. 7.17. Absorption coefficient of GaAs and $Al_{0.33}Ga_{0.67}As$. Note the strong absorption at energies just above the band gap, and the similar shape of the spectra. As the aluminium fraction x is increased, the $Al_xGa_{1-x}As$ band edge moves to higher energies. At compositions above $x = 0.4$ the band gap becomes indirect and $Al_xGa_{1-x}As$ is too poor an absorber to be useful for photogeneration. However, high aluminium content material is used as a 'window', much as oxide is used in silicon cells.

GaAs enjoys other advantages for solar cell applications. One is a better temperature coefficient than silicon. Solar cell efficiency tends to decrease as temperature increases, because of increasing carrier recombination and a decreasing band gap. The first effect is more pronounced in silicon where recombination depends upon the availability of phonons, which increases with temperature. This means that GaAs performs better in situations where the cell operates at high temperature, under concentration and in space.

Another factor relevant to space is the radiation resistance of the material. Exposure to extraterrestrial radiation degrades the efficiency of a solar cell over its lifetime. GaAs has a better radiation hardness than silicon, although indium phosphide is better still.

High purity GaAs is much more costly than pure silicon, and GaAs cells are some 5 to 10 times as expensive in spite of reduced cell thickness. However, GaAs production technology is still maturing, and production costs are decreasing as the scale of manufacture of GaAs based devices for applications in optoelectronics expands.

These factors mean that GaAs cells have been developed primarily for use in space. The most likely terrestrial application is for power generation under concentrated light. (Concentration is discussed in Chapter 9.) Although most of the design features mentioned below are general, some were developed with concentration in mind.

7.5.3. *Doping*

III–V semiconductors can be doped by replacing one of the elements with one of different valence. In GaAs, n type doping may be achieved by introducing controlled amounts of silicon during growth. The tetravalent silicon atoms normally replace some of the trivalent gallium atoms in the lattice, introducing a donor state associated with the extra valence electron. Tin, which is also tetravalent, has also been used as a donor impurity. For p type doping, carbon is the most commonly used impurity. Carbon, like silicon, is tetravalent but under certain growth conditions it prefers to substitute for *arsenic* atoms in the lattice, and so introduces a deficiency of valence electrons and hence acceptor states. Alternatively the group II element beryllium can be used, which introduces an acceptor state by substituting for gallium atoms. Dopants may be introduced by diffusion or directly during growth by epitaxial techniques.

7.5.4. *Recombination*

Radiative recombination is faster in GaAs than in silicon, and may dominate in very pure material. The direct band gap means that direct transitions from conduction to valence band are more likely and the radiative lifetime is a factor of 10^4 shorter. For example, at acceptor doping levels of 10^{16} cm^{-3}, τ_{rad} is around 1 μs in pure p type GaAs, compared to around 10 ms in silicon [Lush, 1986; Green, 1995]. The large difference is due to the difference in the absorption coefficients at energies just above the respective absorption edges. We know from Sec. 4.5.1 that the recombination rate is proportional to the value of α within k_BT of the absorption edge.

Auger recombination, on the other hand, is much slower than in silicon due to the much smaller intrinsic carrier concentration ($n_i = 2\times10^6$ cm^{-3} in GaAs, compared to 1×10^{10} cm^{-3} in silicon) and is negligible in photovoltaic devices.

In practical materials, nonradiative Shockley Read Hall recombination through defect states dominates. SRH lifetimes depend on the nature and concentration of impurities and the growth conditions. Electron lifetime in good quality p type GaAs varies from around 1 μs at low doping levels to approximately $(10^{19}/N_a)$ ns at higher doping. For holes in n type GaAs τ_p varies from 10–100 ns at low doping to a few ns at $N_d > 10^{18}$ cm^{-3}. τ_p is up to one order of magnitude shorter than τ_n in p type GaAs doped to the same level.

In p–n devices, SRH recombination in the space charge region dominates performance. This is because the SRH recombination rate is greatest in this region where n and p are similar, and because the carrier densities in the junction region of a GaAs device are high because of the high absorption. The recombination rate therefore involves both carrier types, and is not simply proportional to the minority carrier density, as in silicon, nor is it possible to resolve the lifetime into factors due to different processes, as in Eq. 7.8.

The effect of space charge region recombination is evident in the dark current characteristics of GaAs solar cells. The dark current tends to vary with the ideality factor of 2 characteristic of SRH recombination at the junction, $J_{dark} \sim e^{qV/2k_BT}$, at low positive biases. In silicon cells, for comparison, where the dark current is dominated by recombination in the neutral regions, the ideality factor is close to 1. At higher forward biases in good quality GaAs cells J_{dark} begins to vary like e^{qV/k_BT}, as radiative recombination begins to dominate.

Table 7.3. Typical GaAs material and cell parameters, used in calculations in Fig. 7.19.

	EMITTER (P TYPE)	BASE (N TYPE)
Thickness/μm	0.5	4
Doping/cm^{-3}	1×10^{18}	1×10^{17}
Minority carrier diffusion constant/cm^2 s^{-1}	25	10
Minority carrier lifetime/s	1×10^{-9}	1^{-8}
Minority carrier diffusion length/μm	1.5	3
Surface recombination velocity/cm s^{-1}	1000000	1000000
Reflectivity	0.05	
Absorption: at 500 nm/cm^{-1}	110000	
at 800 nm/cm^{-1}	13000	

Like bulk recombination, surface recombination is higher in GaAs than in silicon. At untreated surfaces the minority carrier surface recombination is of the order of 10^6 cm s^{-1} [Hovel, 1975] but this can be reduced to under 10^4 cm s^{-1} using window layers, discussed below [Bube, 1998].

7.5.5. *Carrier transport*

As in silicon, electrons have a higher mobility than holes in GaAs. Minority electrons have a mobility of around 5000 cm^2 V^{-1} s^{-1} at low doping which falls to around 1000 cm^2 V^{-1} s^{-1} at doping levels of 10^{18} cm^{-3}. The minority hole mobility varies from 300–400 cm^2 V^{-1} s^{-1} at low doping levels to less than 100 cm^2 V^{-1} s^{-1} at doping levels of more than 10^{18} cm^{-3}. Together with minority carrier lifetimes, these lead to diffusion lengths of a few μm in moderately doped GaAs, longer for p than n type. In ternary alloys, the higher defect density leads to much faster recombination. For example, in n doped $Al_xGa_{1-x}As$, L_p may be less than 0.1 μm and lifetimes may be a few tens of ps.

The doping dependent resistivity of p type GaAs is around $(10^{15}/N_a)$ Ohm cm. It is about an order of magnitude greater for n type material doped to the same level.

7.5.6. *Reflectivity*

GaAs has a refractive index of 3.3 and a natural reflectivity of 30–40% over visible wavelengths. Suitable anti-reflection coatings can be made from silicon nitride.

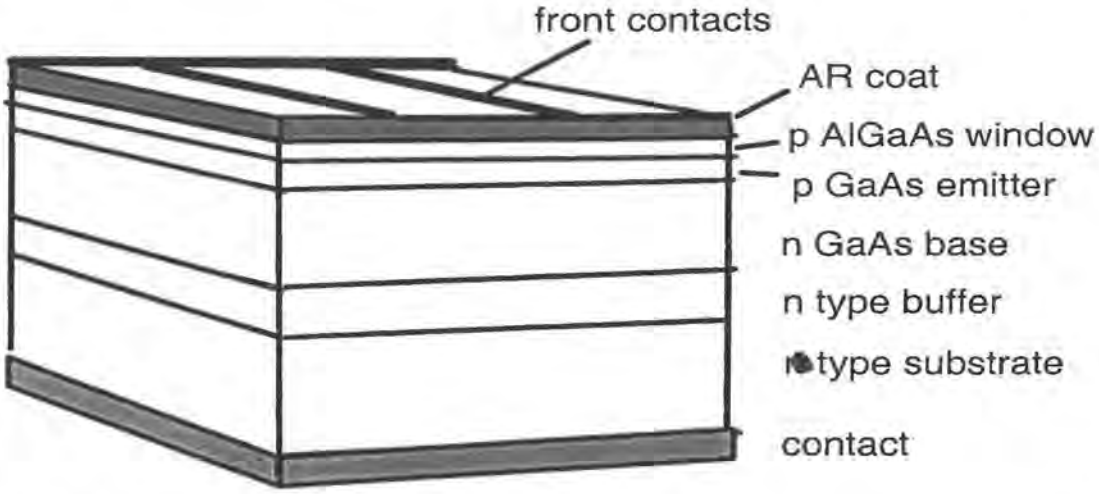


Fig. 7.18. Layer structure of a windowed monocrystalline GaAs solar cell.

7.6. GaAs Solar Cell Design

7.6.1. *Basic GaAs solar cell*

In GaAs, because diffusion lengths greater than the absorption depth can be achieved for either doping type, cells can be prepared as either p–n or n–p designs. In either case the emitter should be as thin as possible without increasing series resistance too much. For the p–n design, a 0.5 μm emitter doped to 10^{18} cm^{-3} is typical; for the n–p design, the emitter can be as thin as 0.2 μm because of the higher n type conductivity. In practice p^+–n designs seem to perform better than n^+–p designs. The base is much shorter than in silicon cells, typically 2–4 μm and comparable with the diffusion length.

7.6.2. *Optimisation of GaAs solar cell design*

Figure 7.19 illustrates absorption and recombination profile in a typical GaAs p–n junction cell, calculated as Fig. 7.8 and using the materials parameters in Table 7.3. The graphs show that for GaAs the influences on recombination are rather different. We can make the following observations:

- absorption of light is good at all wavelengths
- front surface recombination is important for long wavelengths
- recombination in the junction region is dominant

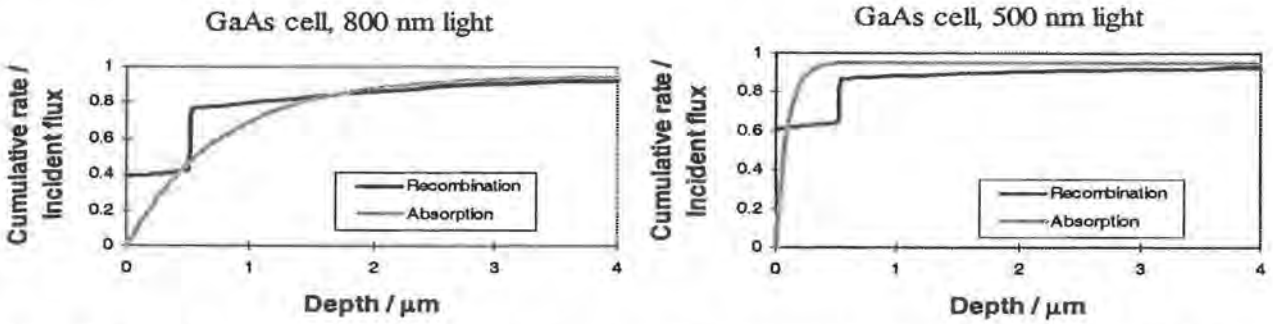


Fig. 7.19. Simulated cumulative absorption and recombination rates for a GaAs solar cell with the parameters in Table 7.3 under monochromatic light at two different wavelengths and at a bias close to open circuit.

- bulk recombination is unimportant relative to junction and surface recombination
- rear surface recombination is negligible, because of the high absorption

Therefore, the objectives in optimising GaAs cell design should be to

- Minimise front surface recombination
- Minimise junction recombination
- Minimise series resistance

We will add one more very practical objective:

- Minimise substrate cost.

This arises because the GaAs layers are extremely thin, and must be grown on a substrate for mechanical stability, yet depositing GaAs cells on GaAs substrates is prohibitively expensive.

Strategies to address these issues are discussed below.

7.6.3. *Strategies to reduce front surface recombination*

Because of the high absorption coefficient, more carriers are generated in the emitter in a GaAs than a silicon cell. This means the contribution of the emitter to the photocurrent is not negligible and also that surface and bulk recombination in this region are important.

- Front surface fields. A heavily doped layer is introduced by diffusion. The principle is the same as for the back surface field discussed for silicon.
- Window layers. Front surface recombination can be reduced by introducing a front surface *window* of a higher band gap material to reflect minority carriers away from the surface. The principle is similar to the back surface field. The higher band gap of the window layer presents a potential barrier to electrons generated in the p region (Fig. 7.20). The window layer is transparent to most visible light, but the interface between the window and the bulk GaAs is much less defective than untreated GaAs. The consequence is that the effective front surface velocity is greatly reduced. An analysis similar to that given above for back surface field shows that, if interface recombination is negligible, the effective surface velocity at a window layer is given by

$$S_{\text{eff}} = \frac{D_{\text{n+}}}{L_{\text{n+}}} e^{\Delta E_g/kT} \left\{ \frac{\frac{S_{\text{n+}} L_{\text{n+}}}{D_{\text{n+}}} \cosh \frac{x_b}{L_{\text{n+}}} + \sinh \frac{x_b}{L_{\text{n+}}}}{\frac{S_{\text{n+}} L_{\text{n+}}}{D_{\text{n+}}} \sinh \frac{x_b}{L_{\text{n+}}} + \cosh \frac{x_b}{L_{\text{n+}}}} \right\} \qquad (7.11)$$

  where ΔE_g is the difference in band gaps between GaAs and the window and $n+$ denotes the window layer. Typical window materials are $Al_xGa_{1-x}As$ and $In_xGa_{1-x}P$, both of which are lattice matched to GaAs and so result in good quality interfaces with a low density of interface states [Hovel, 1975].
- Heterojunctions and graded emitters. Other options are to fabricate the whole of the emitter from a wider gap material than the base, producing a *p–n heterojunction* cell, where the p–n junction actually occurs at the junction between two materials of different composition. The advantage is that the emitter still absorbs short wavelengths but recombination is suppressed, so improving response to blue light. The wider gap emitter also offers the possibility of increases in V_{oc} (though in practice these are negligible) and flexibility to maintain high doping in emitter without the recombination losses which would result in a homojunction cell. However, the interface is likely to introduce defects, and since recombination is already highest in the depleted region where n and p are similar, there is a danger that defects at the heterojunction will assist recombination. A compromise is to locate the high band gap — low band gap interface away from the depleted region in the neutral p layer.

Another idea is to grade the composition of the p layer, from GaAs near to the junction, to a high band gap alloy at the front surface. The compositional grade introduces an electric field which assists electron migration to

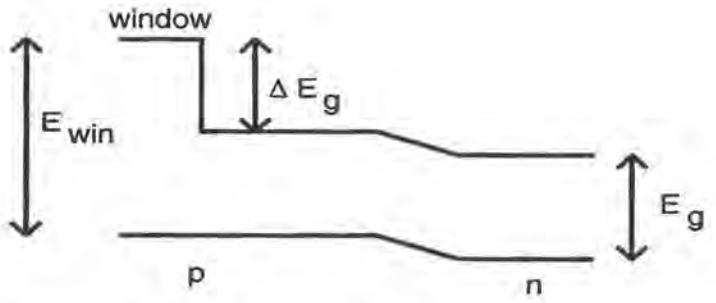


Fig. 7.20. Band diagram of a p–n GaAs cell with a p type AlGaAs window of band gap E_{win}.

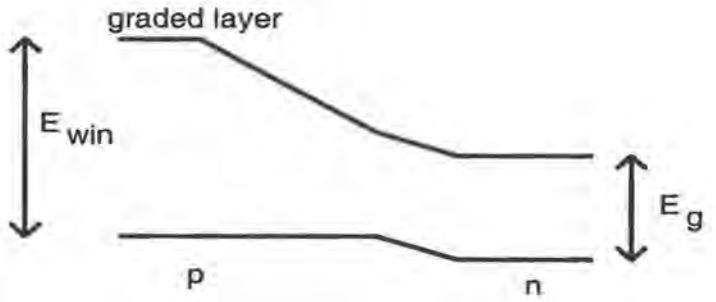


Fig. 7.21. Band diagram of a GaAs cell with an $Al_xGa_{1-x}As$ emitter of graded composition. The compositional grade provides an effective electric field which accelerates electrons towards the junction region and assists charge separation.

the junction, as shown in Fig. 7.21. This can be done in GaAs cells using $Al_xGa_{1-x}As$ of varying aluminium fraction and has resulted in enhanced photocurrent through improved minority carrier collection. The wider gap acts like a window layer, shifting the absorption of longer wavelength photons towards the junction where they are more likely to be utilised. These strategies would certainly be useful for silicon cells, but with silicon there is no convenient wider band gap semiconductor with a similar lattice constant which could be used.

7.6.4. *Strategies to reduce series resistance*

Series resistance is an issue with all solar cell designs because of the need to expose as much of the front surface as possible to the light, yet make electrical contact to it to collect the current. In GaAs concentrator cells it is a particularly important issue. For a cell to operate at optimal efficiency under 100 suns, the sheet resistance of the cell should be less that 10^{-3} Ohm-cm^2. GaAs cells for use in concentration are therefore designed

to have minimal resistance at high injection levels, and are not necessarily the optimum design for one sun. This means special grid patterns with high metallisation, and very high emitter doping, which can be achieved with the help of a window layer and a front surface field.

7.6.5. *Strategies to reduce substrate cost*

Because the GaAs cell is so thin, it must be grown on a substrate, and it is better to choose a substrate of the same lattice constant as the cell material in order not to introduce crystal defects at the rear interface. GaAs substrates are ideal, but prohibitively expensive, and for large scale production germanium, which has a similar lattice constant, is used. Ge has the same lattice constant as GaAs, but is rather rare and in the end the cost of GaAs on Ge solar cells is likely to be limited by the cost of Ge. Growth of GaAs on Si substrates suffers from the difference in lattice constants, and poor material quality inevitably results.

One solution is to re-use the substrate. Techniques have been developed in which epitaxially grown GaAs wafers are removed from a GaAs substrate by etching or cleaving along an intermediate plane. Cells produced this way are brittle and still require mechanical support. Another approach to cost reductions is to use *polycrystalline* GaAs, in which case lattice matching with a substrate is less important. 20% efficient poly-GaAs cells have been grown on Germanium substrates. A long term objective would be to deposit poly-GaAs cells on silicon, metal or glass.

7.7. Summary

The most efficient single junction photovoltaic cells are based on p–n junctions in monocrystalline semiconductors. Crystalline silicon is the most important photovoltaic material and is widely used for terrestrial applications. Crystalline GaAs is more suitable for applications in space and under concentration. Both materials have suitable band gaps leading to a high theoretical efficiency.

The objectives in solar cell design are to maximise light absorption, charge separation and charge transport, and to maximise the photovoltage. The design factors which may be varied in a p–n junction solar cell are the layer thicknesses, the doping levels, the surface treatments and the front surface contact pattern. High light absorption is achieved by using a high optical depth, minimising shading and reflection, and by using light

trapping structures. Charge separation is maximised by using a good quality junction with a high built-in bias. Efficient charge transport requires good quality crystal with low bulk and surface recombination and good majority carrier conductivity. For any material the ideal p–n structure has a thin heavily doped emitter on top of a thicker, heavily doped base. The remaining priorities depend upon the material used.

Silicon has a low absorption coefficient and most photogeneration takes place in the base. p type silicon is used as the base because the minority carrier transport characteristics are better than in n type material. The base should be several hundred μm thick for good light absorption, while the n type emitter is typically less than 0.5 micron thick. The base is lightly doped so that electron diffusion lengths are long enough (several hundred μm) for good charge collection. Because of the low absorption, the priorities in silicon solar cell design are to minimise light reflection and shading, and to reduce recombination at the rear surface. Reflection is reduced by texturing the front surface with a pattern of micron sized pyramids and by adding an anti-reflection coat. Shading is reduced by using thin contact fingers, and by burying the contacts into the emitter, while the series resistance which results from thin contacts can be reduced by differential doping of the semiconductor near to the contacts. Surface recombination is reduced by passivating the surface with a layer of silicon dioxide, and by the use of point contacts at the rear. The best monocrystalline silicon cell is 24% efficient in AM1.5.

GaAs is a strong absorber and devices only a few microns thick absorb most of the light. Both n–p and p–n structures may be used, with typical emitter thicknesses of less than 0.5 μm and base thicknesses of 2–4 μm. Significant photogeneration takes place in the emitter and junction as well as the base. This means that recombination at the front surface and in the junction is important. Front surface recombination can be reduced by building in doped window layers from a wider band gap semiconductor. The window separates the active region from the metallic contact, and helps to drive minority carriers towards the junction. Junction recombination is minimised mainly by control of the crystal quality and doping levels. Reflection is minimised using anti-reflection coats; texturing is not suitable because of the larger size of light trapping structures compared to the device thickness, and the sensitivity of the front surface quality. As in silicon, the front surface contact pattern should be designed to minimise shading without enhancing series resistance. The best single cell efficiency, of 25.1%, has been achieved in GaAs.

References

S. Adachi, *Physical Properties of III-V Semiconductor Compounds* (Chichester: Wiley, 1992).

R. Bube, *Photovoltaic Materials* (Imperial College Press, 1998).

M.A. Green, *Silicon Solar Cells: Advanced Principles and Practice* (Sydney: Centre for Photovoltaic Engineering, 1995).

M.A. Green, K. Emery, D.L. King, S. Igari and W. Warta, "Solar cell efficiency tables", *Progress in Photovoltaics* **9**, 287–293 (2001).

H.J. Hovel, "Solar cells", in *Semiconductors and Semimetals* **11**, eds. R.K. Willardson and A.C. Beer (1975).

G.B. Lush, in *Properties of Gallium Arsenide* (INSPEC, 1986).

M. Shur, *Physics of Semiconductor Devices* (Prentice Hall International, 1990).

R. Sinton *et al.*, "27.5% silicon concentrator solar cells", *Electron Device Letts.* **EDL-7**, 567 (1986).

R.J. van Overstraeten and R. P. Mertens, *Physics Technology and Use of Photovoltaics* (Bristol: Adam Hilger, 1986).

J. Zhao and M.A. Green, "23.5% efficient silicon solar cell", *Prog. Photovoltaics* **2**, 227–230 (1994).

Chapter 9

Managing Light

9.1. Introduction

This chapter is about the effect of increasing photon flux density within the solar cell. The main reason for doing this is that by increasing the photon flux per unit volume, we decrease the volume of cell material which is needed to harness a certain amount of light. This decreases the *cost* per unit of energy converted, since the volume of cell material is the main determinant of cost. Another reason is that cells may perform more efficiently under increased photon flux levels.

We start by reviewing how light is utilised in a photovoltaic cell. Consider a semiconductor layer of active width w, supported by a substrate of a similar semiconductor, as illustrated in Fig. 9.1. Light normally incident on the cell passes through encapsulating materials before meeting the cell surface, where it may strike semiconductor or cell contacts. Some fraction, R, of the light will be reflected from the semiconductor surface, contacts and outer layers. Of the remainder, some will be absorbed in the active

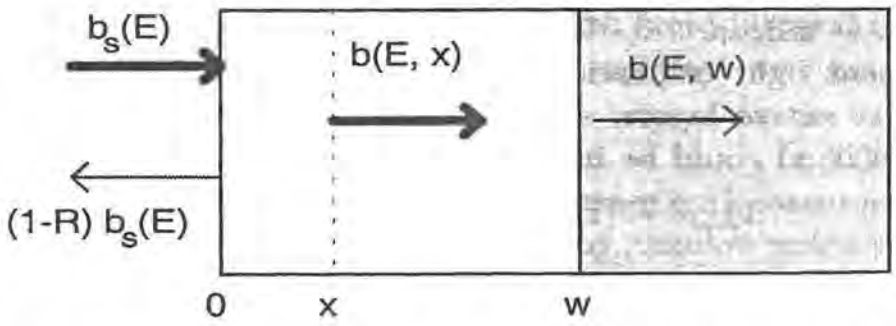


Fig. 9.1. Attenuation, transmission and reflection of incident radiation in a semiconductor layer of active width w. $b_s(E)$ is the incident photon flux density normal to the surface, and $b(E, x)$ is the transmitted photon flux density at depth x below the surface.

layers and some will pass straight through into the substrate. Within the cell, the flux attenuates so that at a depth x the flux density of photons of energy E is given by

$$b(E,x) = (1 - R(E))b_s(E)e^{-\int_0^x \alpha(E,x')dx'} \tag{9.1}$$

where $b_s(E)$ is the incident photon flux density normal to the surface, $R(E)$ is the reflectivity of the surface, and $\alpha(E,x)$ is the absorption coefficient of the semiconductor at x. Assuming that all absorbed photons result in band gap excitation, electron-hole pairs are generated at a rate

$$g(E,x) = \alpha(E,x)b(E,x) = (1 - R(E))\alpha(E,x)b_s(E)e^{-\int_0^x \alpha(E,x')dx'} \tag{9.2}$$

per unit volume, in accordance with Eq. 4.25. For a cell of thickness w, a fraction

$$f_{abs} = 1 - R(E) - \frac{b(E,w)}{b_s(E)}$$

of incident photons are absorbed while a fraction

$$f_{trans} = \frac{b(E,w)}{b_s(E)}$$

pass through. For a uniform absorber the absorbed and unused fractions are

$$f_{abs} = (1 - R(E))(1 - e^{-\alpha(E)w}) \tag{9.3}$$

and

$$f_{trans} = (1 - R(E))e^{-\alpha(E)w} \tag{9.4}$$

respectively. The unabsorbed fraction is high whenever αw is small, for instance, for wavelengths close to the band gap and for indirect semiconductors such as silicon.

For efficient light utilisation, as much light as possible should be absorbed in the active layers, so $R(E)$ and $b(E,w)$ should be as small as possible. If $b(E,x)$ could be increased within the cell, the photogeneration rate would increase. This means we could achieve the same photogeneration rate with a smaller volume, reducing materials requirements.

The effect of increased $b(E,x)$ on cell efficiency is discussed below in Sec. 9.4. We will show that if the flux density is increased by a factor X, the cell current should increase, to a first approximation, in proportion to X while the cell voltage increases as $\log(X)$. This means that the power

delivered by the cell, $J(V) \times V$, increases *faster* than X, so the *efficiency* increases under increased photon flux levels. Note that this is independent of the result that for some defective semiconductors, current collection can be improved at higher light intensities.

This chapter is organised as follows: first we review the role of photon flux in a solar cell and discuss the ways in which light management can improve the performance of a solar cell. Then we treat the following strategies: anti-reflection coatings, concentrator systems, light trapping structures and photon recycling systems. In each case the effect on device physics is briefly discussed.

9.2. Photon Flux: A Review and Overview of Light Management

Box 9.1. Understanding photon flux: some definitions

The spectral photon flux density $\beta(E, \mathbf{r}, \theta, \phi)$ is the number of photons of energy E which pass through unit area per unit solid angle per unit time. In general, β is defined by a direction (θ, ϕ) and is position as well as energy dependent. Conservation of photons requires that, for some volume $\mathcal{V}$ of material with surface S,

$$\int_{\mathrm{S}} \beta(E, \mathbf{s}, \theta, \phi) d\Omega \cdot d\mathbf{S} = \int_{\nu} (g_{\mathrm{ph}}(E, \mathbf{r}) - u_{\mathrm{ph}}(E, \mathbf{r})) dV \tag{9.5}$$

where $\mathbf{s}$ is a point on the surface, $g_{\mathrm{ph}}(E, \mathbf{r})$ is the photon emission rate and $u_{\mathrm{ph}}(E, \mathbf{r})$ the photon absorption rate per unit volume. We define $u_{\mathrm{ph}}(E, \mathbf{r})$ as the *net* absorption, that is, the difference between absorption and stimulated emission. The left hand integral in Eq. 9.5 is taken over solid angle Ω and points $\mathbf{s}$ on the surface S, while the right hand integral is taken over all points $\mathbf{r}$ within the volume $\mathcal{V}$. Using Gauss's law we find that at any point within $\mathcal{V}$

$$\int_{\Omega} \nabla \cdot \beta(E, \mathbf{r}, \theta, \phi) d\Omega = g_{\mathrm{ph}}(E, \mathbf{r}) - u_{\mathrm{ph}}(E, \mathbf{r}) \,. \tag{9.6}$$

The rate of absorption for photons for a given flux direction is proportional to the magnitude of the photon flux, β, in that direction, and the emission rate proportional to the magnitude of emitted flux. Summing up

contributions over solid angle, we find

$$u_{\mathrm{ph}}(E,\mathbf{r}) = \int_{\Omega} \alpha(E,\mathbf{r})\beta(E,\mathbf{r},\theta,\phi)d\Omega \tag{9.7}$$

where α is the absorption coefficient at $\mathbf{r}$, and

$$g_{\mathrm{ph}}(E,\mathbf{r}) = \int_{\Omega} \varepsilon(E,\mathbf{r})\beta_{\mathrm{e}}(E,\mathbf{r},\theta,\phi)d\Omega \tag{9.8}$$

where ε is the emission coefficient and and β_{e} is the emitted flux density. Note that, in general, ε and α may be direction dependent. If they are direction independent, then u_{ph} and g_{ph} are proportional to the sum of magnitudes of the fluxes from different directions. Combining Eqs. 9.6 to 9.8 and resolving along the direction (θ,ϕ), we have

$$\nabla \cdot \boldsymbol{\beta} = \varepsilon\beta_{\mathrm{e}} - \alpha\beta$$

which can also be written as

$$\frac{d\beta}{dl} = \varepsilon\beta_{\mathrm{e}} - \alpha\beta \tag{9.9}$$

where l is the lenth coordinate along (θ,ϕ) [Chandrasekhar, 1950].

In real materials the emitted flux, which arises from the radiative recombination of carriers, is usually negligible. That leaves us with the soluble linear equation

$$\frac{d\beta}{dl} + \alpha\beta = 0 \tag{9.10}$$

In the case where the incident flux is directed normal to the cell surface ($\theta = 0$, $\phi = 0$), we set $\beta = b(E,x)$ and have the one dimensional equation

$$\frac{db}{dx} + \alpha b = 0 \tag{9.11}$$

which has the solution

$$b(E,x) = b(E,0)e^{-\int_0^x \alpha(\mathrm{E},\mathrm{x}')\mathrm{dx}'} \tag{9.12}$$

usually known as the Beer–Lambert law. Henceforth, $b(E,x)$ is used to represent the flux travelling at depth x normal to the surface of a planar structure.

9.2.1. *Routes to higher photon flux*

An increased photon flux can be achieved in several ways, by *concentration* of the light, *trapping* of the light, or by exploitation of *photon recycling*. These are illustrated schematically in Fig. 9.2.

All of these processes aim to improve the efficiency of utilisation of photons in a given amount of photovoltaic material. They are useful in different conditions, for instance light trapping strategies are particularly important for weak absorbers and thin cells, while concentration is important in very efficient, crystalline cells.

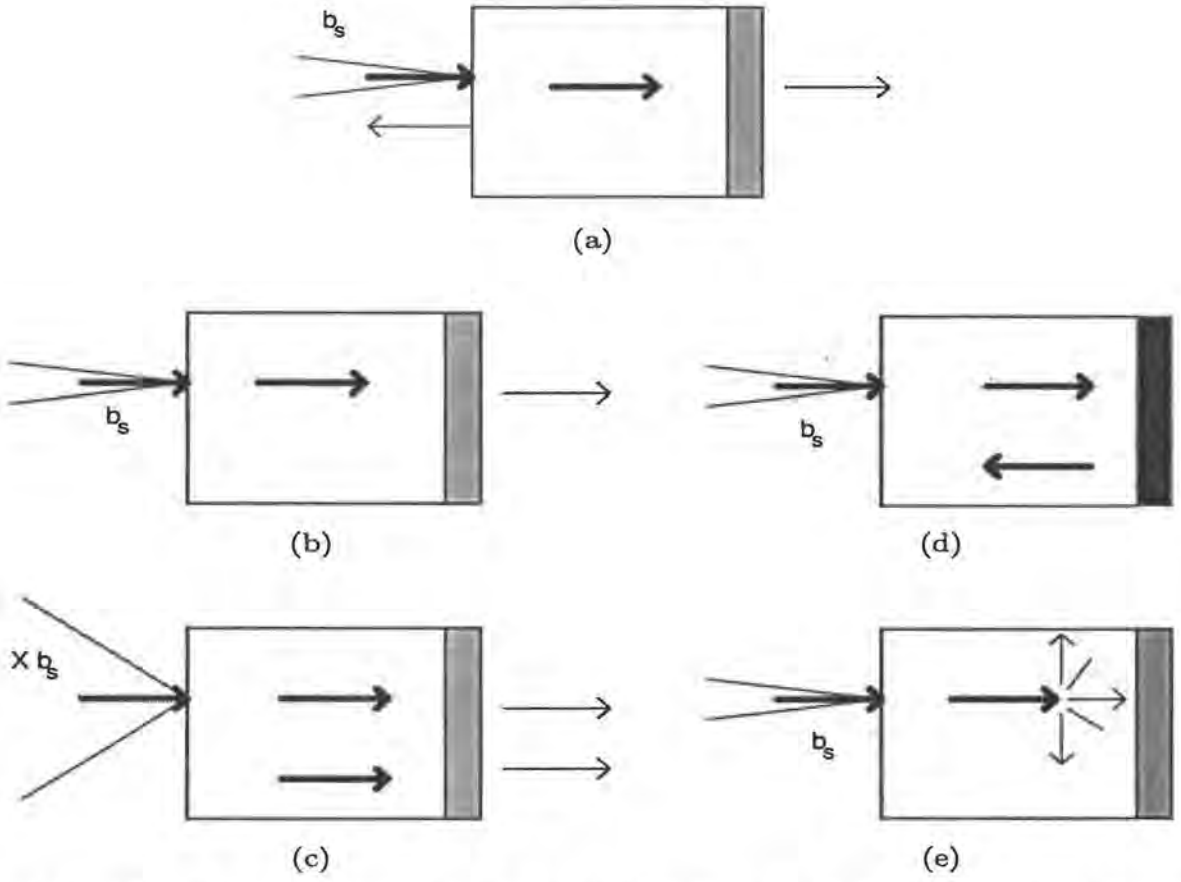


Fig. 9.2. (a) In a non optimised solar cell, light is incident at normal (one sun) concentration and light is lost by reflection and transmission. Photogeneration can be increased by (b) reducing reflection with an anti-reflection coating; (c) increasing the incident photon flux density by concentrating the light. Note that this increases the angle subtended by the light source; (d) confining the light by means of internal reflection (or light trapping) strategies; (e) reabsorption of photons produced by radiative recombination. Some of these strategies would normally be used together, *e.g.*, an AR coat would be used under concentration, and light trapping structures would be used to enhance photon recycling.

In this chapter we will look at the basic physics of each of the processes, how they are or might be achieved in practice, and how they affect the physics and performance of the photovoltaic cell.

9.3. Minimising Reflection

9.3.1. *Optical properties of semiconductors*

The optical properties of a solid are described by the dielectric constant, ε_s. ε_s is a complex quantity and obeys

$$\sqrt{\varepsilon_s} = n_s - i\kappa_s \tag{9.13}$$

where n_s is the refractive index of the material and the imaginary part κ_s is related to the absorption coefficient of the material through

$$\alpha = \frac{4\pi\kappa_s}{\lambda}. \tag{9.14}$$

In general, ε_s, n_s and κ_s are wavelength dependent, and may be direction dependent.

Differences in refractive index determine the reflection and transmission of light at the interface between two materials. For thin films, light should be treated as coherent, and Maxwell's equations can be solved to find the relative amplitudes of transmitted and reflected waves. In the simplest case, that of a plane boundary between materials of refractive index n_0 and n_s, light striking the interface at normal incidence is reflected with probability

$$R = \left(\frac{n_0 - n_s}{n_0 + n_s}\right)^2. \tag{9.15}$$

For a semiconductor, n_s is typically 3–4 at visible wavelengths (Table 9.1), so that some 30–40% of light normally incident on the surface from free space will be reflected.

At oblique incidence the reflectivity is angle and polarisation dependent. If the incident ray makes an angle θ_0 with the surface normal, and the transmitted ray an angle θ_s with the surface normal *inside* the semiconductor then

$$R = \left(\frac{\eta_0 - \eta_s}{\eta_0 + \eta_s}\right)^2 \tag{9.16}$$

Table 9.1. Refractive indices of selected semiconductors and AR coat materials. From various sources including Pankove [Pankove, 1971].

Material	Refractive index (at ca. 1.5 eV)
Si	3.44
Ge	4.00
GaAs	3.6
$Al_{0.8}Ga_{0.2}As$	3.2
AlAs	3.0
Tl_2O_5	2.1
ZnO	2.02
$Si_3\ N_4$	1.97
SiO_2	1.46

where

$$\eta_s = n_s \sec \theta_s \tag{9.17}$$

for p polarised light (with the electric field vector *in* the plane of incidence) and

$$\eta_s = n_s \cos \theta_s \tag{9.18}$$

and for s-polarised light (electric field vector *normal* to the plane of incidence). θ_0 and θ_s are related through Snell's law

$$n_0 \sin \theta_0 = n_s \sin \theta_s \, .$$

For unpolarised light, considered as an equal mix of s and p polarisations, the net reflectivity generally increases with angle (for the s polarised component it increases while for the p component it has a minimum), approaching one at large angles. The minimum reflectivity of a semiconductor surface to unpolarised light occurs at normal incidence and is given by Eq. 9.15. At 30–40%, this is unacceptably high for efficient photovoltaic energy conversion.

Snell's law also means that light travelling *within* the semiconductor towards the surface at an angle greater than the critical angle, θ_c, is internally reflected. For an air-semiconductor interface the critical angle is given by

$$\theta_c = \sin^{-1}\left(\frac{n_0}{n_s}\right) . \tag{9.19}$$

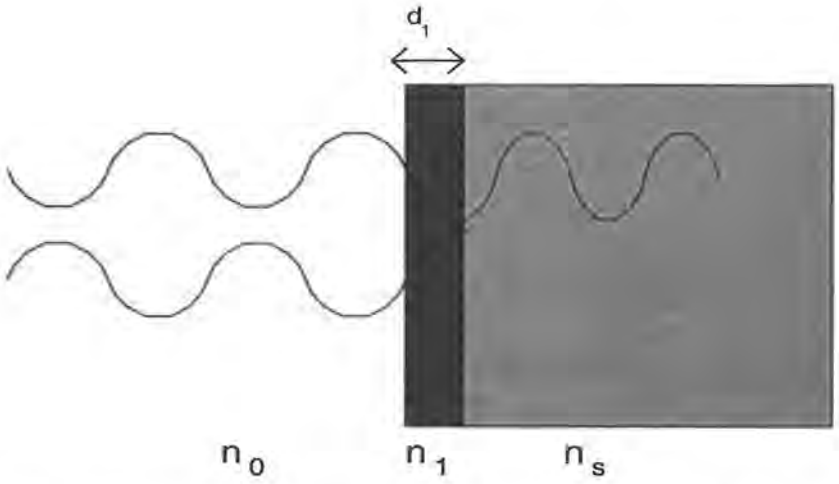


Fig. 9.3. Single layer anti-reflection coating. When the thickness of the dielectric layer is equal to one quarter wavelength, the incident and reflected waves interfere destructively to cancel out the reflected ray at the outer surface.

For the air-silicon interface $\theta_c = 16.9°$. In the presence of an optical coating n_0 in Eq. 9.19 is replaced by the refractive index of the coating and θ_c is increased. This phenomenon of total internal reflection is exploited in light trapping structures, which are discussed in Sec. 9.6.

9.3.2. *Anti-reflection coatings*

Reflectivity of the air-semiconductor interface can be reduced using an *anti-reflection (AR) coating*. An AR coat is a thin film of a dielectric with refractive index n_1 intermediate between those of the semiconductor (n_s) and free space (n_0). By considering forward and backward travelling waves in each medium it can be shown [Macleod, 2001] that the reflectivity of the film for light of wavelength λ is given by

$$R = \frac{(\eta_0 - \eta_s)^2 + (\eta_0\eta_s/\eta_1 - \eta_1)^2 \tan^2 \delta_1}{(\eta_0 + \eta_s)^2 + (\eta_0\eta_s/\eta_1 + \eta_1) \tan^2 \delta_1} \tag{9.20}$$

where δ_1 is the phase shift in the film,

$$\delta_1 = 2\pi\eta_1 d_1 \cos\theta_1/\lambda\,, \tag{9.21}$$

θ_1 is the angle between the light ray and the normal within the film, and d_1 is the film thickness. R clearly has its minimum value when $\delta_1 = \pi/2$. For normal incidence, this first happens when d_1 is equal to a quarter wavelength in the thin film material. (In those conditions the waves reflected

from the front and rear interface of the thin film are out of phase and interfere destructively.) R *vanishes* when it is also true that

$$n_1 = \sqrt{n_0 n_s}\,. \tag{9.22}$$

So by coating our semiconductor with a thin layer of a medium with refractive index $\sqrt{n_s}$, we can reduce the reflectivity to zero at some particular wavelength, λ_0. Since the solar spectrum is broad, this wavelength should be chosen to lie towards the middle of the range of wavelengths which can be usefully absorbed for that semiconductor, and this determines the AR coat thickness. Close to λ_0, R increases with wavelength approximately like $(\frac{\Delta\lambda}{\lambda_0})^2$, and at wavelengths where the phase shift becomes a multiple of π, R reaches its maximum value equal to the natural reflectivity of the uncoated interface. This means that an AR coat optimised at one visible wavelength may be quite highly reflecting at others. AR coats on silicon solar cells are usually optimised for red light, where solar irradiance is strong, and become reflective in the blue. For this reason silicon solar cells often appear violet or blue. Similarly, the reflectivity of the AR coat depends on angle, and increases at wide angles. This means that AR coats are of limited use in diffuse light or for non-planar surfaces. Figure 9.4 shows how an AR coat which is tuned for normal incidence at a certain wavelength becomes reflecting at other wavelengths and wide angles.

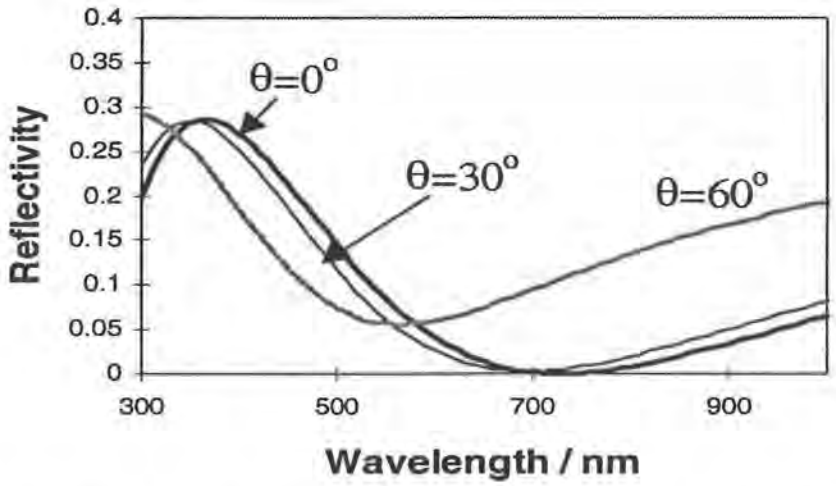


Fig. 9.4. Calculated reflectivity of single layer AR coat as function of wavelength at different angles of incidence $\theta = 0°$, 30° and 60°. Notice how the minimum of reflectivity lifts and shifts to shorter wavelengths at wider angles of incidence. Calculated for a 100 nm layer of refractive index $n_1 = \sqrt{3.3}$ on top of a semiconductor of refractive index $n_s = 3.3$.

In the above treatment of AR coats, we neglect the absorption of the optical materials. This is usually a good approximation since, for most semiconductors, the absorption visible wavelength is sufficiently weak that $\kappa_s \ll n_s$. For reflection from an absorbing medium, n_s should properly be replaced by the complex refractive index $n_s - i\kappa_s$ in Eqs. 9.17–9.20 and R calculated with complex arithmetic. The *incident* medium should of course be non-absorbing.

According to Eq. 9.22, the ideal AR coat material for silicon should have a refractive index of around 1.84, and for GaAs, 1.90. Good materials are silicon nitride which has $n_1 = 1.97$ and tantalum oxide with $n_1 = 2.1$. These give relectivities of less than 1% at normal incidence at the optimum wavelength.

Improved reflectivity over a band of wavelengths can be achieved with two or more thin films. The greater the number of layers, the greater the range of wavelengths over which the reflectivity can be minimised. Multiple layers are not usually practical for solar cells, given their cost and the sensitivity to angle of incidence, but double layers are used on some high efficiency cells. Layers should be deposited so that refractive index increases consecutively from air (n_0) to first coat (n_1) to second coat (n_2) to semiconductor (n_s). Reflectivity vanishes when both films have quarter-wave

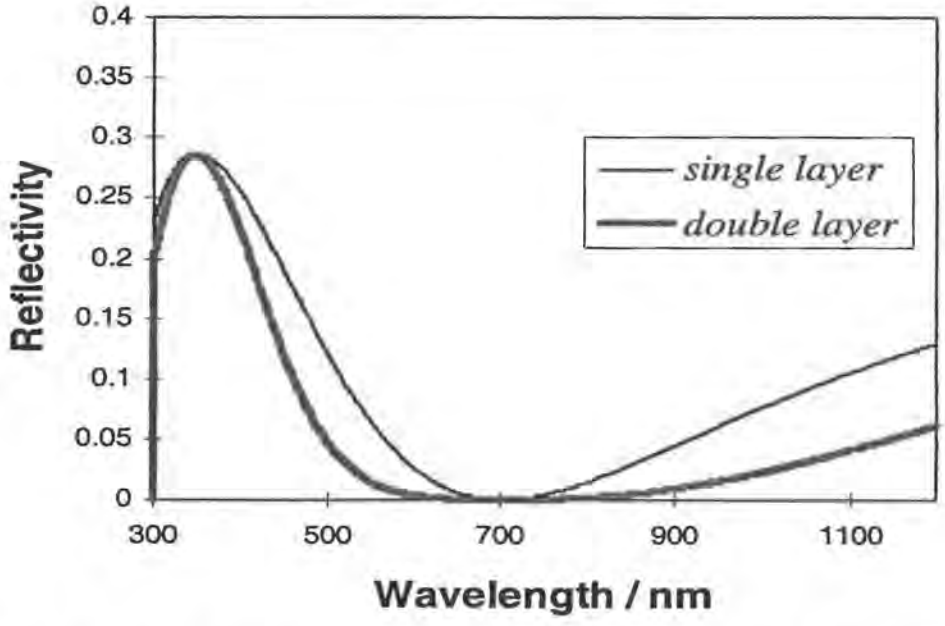


Fig. 9.5. Calculated reflectivity of double layer AR coat, optimised for 700 nm. The reflectivity of a single layer is presented for comparison.

thickness and

$$\left(\frac{n_2}{n_1}\right)^2 = \frac{n_s}{n_0}. \tag{9.23}$$

This condition allows a little more flexibility in choice of materials than the single layer case. In some cell designs, where a wide band gap window layer is present for improved carrier collection, a double layer AR coat can be made simply by choosing an optical coating with refractive index n_1 satisfying Eq. 9.23 where n_2 is the refractive index of the window layer, and choosing appropriate thicknesses. This has been done with GaAs devices having a high aluminium content AlGaAs window.

Regarding solar cell device physics, the effect of anti-reflection coatings is simply to reduce R, and increase the flux and generation rates, as defined in Eqs. 9.1 and 9.2.

9.4. Concentration

9.4.1. *Limits to concentration*

In photovoltaic *concentration* systems, light is collected over a large area and focused on to a cell surface of smaller area. This increases the incident flux density at the cell surface by a concentration factor, X, which is approximately the ratio of the collection and cell areas. The main objective of this strategy is to reduce the cost of the photovoltaic conversion system. Concentrating optics are much cheaper per unit area than photovoltaic cells and the cells convert the concentrated and unconcentrated light with approximately the same efficiency. To a first approximation, the concentrator cell delivers the power of X identical cells in unconcentrated light.

According to geometrical optics, concentrating the incident light is equivalent to expanding the angular range subtended by the source. Concentrating sunlight by a factor X using a spherical concentrator increases the angular range to θ_X such and enhances the total (angle-integrated) incident flux by a factor

$$X = \frac{\int_0^{2\pi}\int_0^{\theta_X} d\Omega}{\int_0^{2\pi}\int_0^{\theta_s} d\Omega} = \frac{\sin^2\theta_X}{\sin^2\theta_{sun}} \tag{9.24}$$

where θ_{sun} is the half angle subtended by the sun, 0.26°. It is assumed that the intensity of the concentrated light is uniform over the angular range. Because of refraction of light, most of the light entering the semiconductor

is directed close to the normal, and so the flux normal to the surface is enhanced approximately by the factor X relative to the unconcentrated case. For normal incidence,

$$b(E,x) = (1 - R(E))Xb_s(E)e^{-\int_0^x \alpha(E,x')dx'} . \tag{9.25}$$

To a first approximation, the generation rate, as given in Eq. 9.2 increases by the same factor.

The geometrical factor F_s, introduced in Chapter 2 to relate the absorbed light to the angular range of the source, changes to

$$F_X = \pi \sin^2 \theta_X = X \sin^2 \theta_{sun} . \tag{9.26}$$

The maximum theoretical concentration occurs when $\theta_X = 90°$, *i.e.*, $X = 46050$. For a concentrator with cylindrical rather than spherical symmetry the maximum concentration is lower: $X = 1/\sin\theta_{sun}$.

9.4.2. *Practical concentrators*

In practice, concentration is limited by the quality of the optics, by tracking errors in working photovoltaic systems, and by the practical difficulty of coupling very highly concentrated light, with its very wide angular range, into the cell. Practical concentration ratios of tens to thousands of suns have been achieved using reflective optics or Fresnel lenses.

Note that imaging concentration is only possible when incident light rays are parallel and the concentrator is pointing at the sun. This means that only *direct* sunlight can be concentrated. *Diffuse* sunlight, which has been scattered by the atmosphere and composes some 15% of the standard

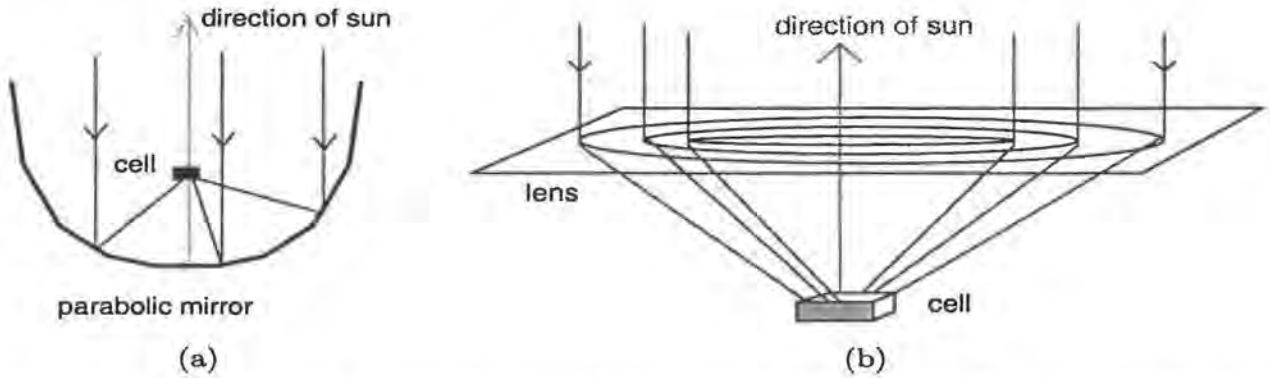


Fig. 9.6. Concentrator system based on (a) parabolic reflector (b) Fresnel lens. The cell is placed at the focus in either case.

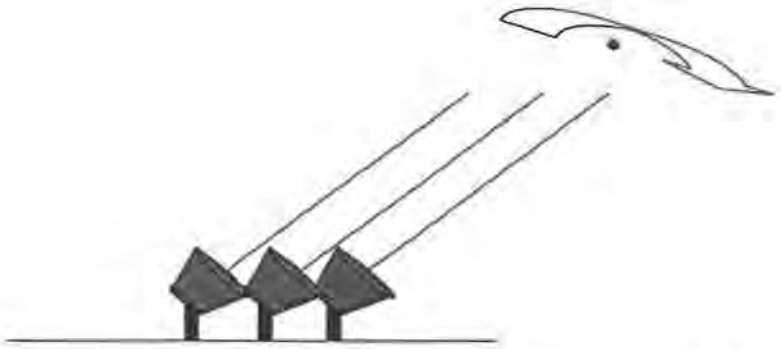


Fig. 9.7. Tracking system. The concentrator must be mounted to follow the direction of the sun on a daily and seasonal cycle.

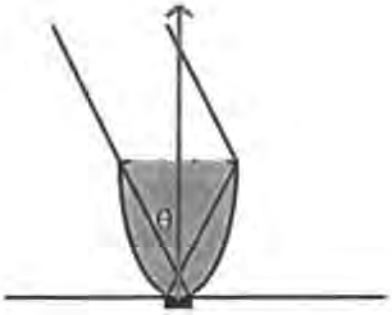


Fig. 9.8. Compound parabolic concentrator.

AM1.5 spectrum, is incident on the concentrator from wide angles and cannot be coupled into the cell. Even in direct sunlight, the concentrator must be made to follow the sun during its daily and seasonal trajectories, and this is achieved with *tracking* systems. Concentrator systems are most cost effective with high quality cells where the saving on photovoltaic materials is greatest, and in sunny climates where the direct solar irradiation is largest. Tracking systems are essential for high efficiency but add to the cost and complexity of the system.

An alternative to imaging concentrators are non-imaging, low concentration systems which do not require tracking. An example is the compound parabolic concentrator [Welford, 1990]. The cell receives light with the angular range θ, where θ is the half angle of the paraboloid, and achieves concentration ratios of $4/\sin^2\theta$, or $4n^2/\sin^2\theta$ if filled with a medium of refractive index n.

Another approach to non-imaging concentration is the use of fluorescent concentrators. A refractive material filled with fluorescent particles re-emits

photons on irradiation. The emitted light can be directed on to the cell exploiting total internal reflection, and the spectral range of the re-emitted light can be controlled [Goetzberger, 1977].

9.5. Effects of Concentration on Device Physics

We will distinguish cells which operate under concentration in *low injection* conditions, where photogenerated carrier density is smaller than the equilibrium majority carrier density, from those which operate in *high injection* conditions, where the photogenerated density is greater.

9.5.1. *Low injection*

In low injection, a p–n concentrator cell can be treated essentially with the standard theory for a p–n junction as developed in Chapter 6. In low injection conditions, minority carrier recombination rates are linear and the current densities resolve into a dark component $J_{\rm dark}$ which depends on V and a bias independent, light generated component $J_{\rm sc}$ which depends on flux density and acts in the opposite direction. Thus, from Eq. 1.4

$$J(V) = J_{\rm sc} - J_{\rm dark}(V)$$

where the dark current typically has the diode-like form so that

$$J(V) = J_{\rm sc} - J_0(e^{{\rm qV}/{\rm mk_B T}} - 1)\,. \tag{9.27}$$

Now if the incident flux $b_{\rm s}$ is increased by a factor X through concentration of the light, $J_{\rm sc}$ will increase to a first approximation by the same factor, *i.e.*, $J_{\rm sc}(Xb_{\rm s}) \approx XJ_{\rm sc}(b_{\rm s})$, but leave the dark current, for a given bias, unaffected. This means that at any given bias below $V_{\rm oc}$ the *net* current increases with X. Moreover, a higher bias is needed to achieve the open-circuit condition where dark current cancels photocurrent. Replacing $J_{\rm sc}$ with $XJ_{\rm sc}$ in Eq. 9.27 we see that for a diode of ideality factor m, the open circuit voltage becomes

$$V_{\rm oc}(X) = \frac{mk_{\rm B}T}{q} \ln\left(\frac{XJ_{\rm sc}}{J_0} + 1\right) \approx V_{\rm oc}(1) + \frac{mk_{\rm B}T}{q} \ln X \tag{9.28}$$

where $V_{\rm oc}(1)$ represents the open circuit voltage under unconcentrated light (1 sun). Thus $V_{\rm oc}$ increases logarithmically with X, and if the cell fill factor remains the same then the power delivered by the cell should increase by a

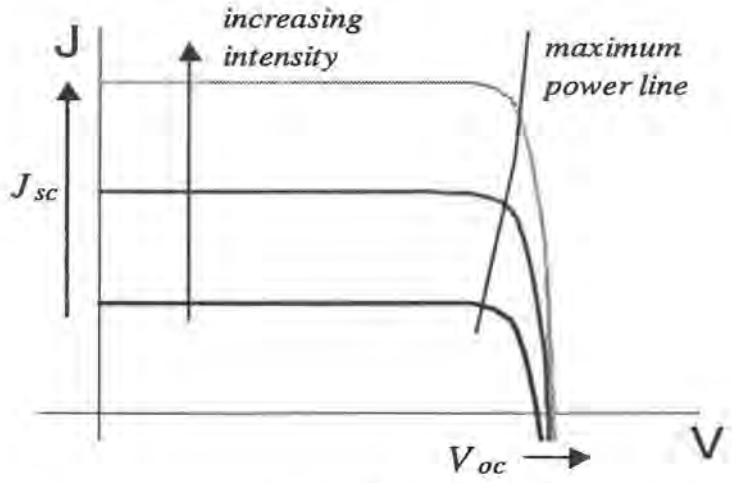


Fig. 9.9. Current–voltage curve at different intensities. J_{sc} increases linearly with light intensity and V_{oc} increases like the logarithm of intensity. The result is that the maximum power point moves to higher voltages, and the efficiency increases.

factor

$$X\left(1+\frac{mk_{\mathrm{B}}T}{qV_{\mathrm{oc}}(1)}\ln X\right),$$

and the efficiency by a factor

$$\left(1+\frac{mk_{\mathrm{B}}T}{qV_{\mathrm{oc}}(1)}\ln X\right).$$

Figure 9.9 shows how the operating point of the cell moves to higher voltage as X increases.

In practice, however, the dark current *is* affected by concentration through (i) increasing carrier densities leading to high injection conditions, and (ii) increasing temperature, both leading to deviations from this behaviour. These effects are discussed in the following sections.

9.5.2. *High injection*

Concentrator cells may be designed to operate in *high injection* conditions. That means that the photogenerated carrier densities are large compared to the doping density in the base, and both carrier types must be included in the transport equations. Generally n and p will be different and electric fields due to carrier density gradients must be taken into account. In the limit where $n = p$ we showed in Sec. 8.4, for p–i–n structures, that the transport equation for either carrier reduces to the diffusion equation-like

form,

$$\frac{d^2n}{dx^2} - \frac{(n-n_0)}{L_a^2} + \frac{g(E,x)}{D_a} = 0 \qquad (8.9^*)$$

where $D_a = \frac{D_n D_p}{D_n + D_p}$, $L_a^2 = D_a \tau_a$, and τ_a is the ambipolar carrier lifetime, given by Eq. 8.12.

If the dominant recombination process is *linear* with excess carrier density, then L_n is independent of n, the differential equation (8.9) is linear in n and the solution can be separated into independent light dependent and bias dependent terms, as in the treatment in Sec. 6.5, *i.e.*, superposition is valid. This is the case for Shockley Read Hall recombination, even at high injection. Letting $n = p$ in Eq. 4.79 for the SRH recombination rate, and assuming $n \gg n_t$, and $p \gg p_t$, we find that

$$U_{SRH} \approx \frac{n}{\tau_n + \tau_p} . \qquad (9.29)$$

U_{SRH} is still proportional to n but the carrier lifetime has been increased. Physically this is due to the relative scarcity of majority carriers. In low injection, the capture of a minority carrier in a trap state is immediately followed by the capture of a majority carrier, simply because of the abundance of majority carriers, and the rate determining step was the minority carrier capture. Now both hole capture *and* electron capture are rate determining.

If SRH recombination is dominant at all injection levels, superposition should hold and the arguments given above for the variation of V_{oc} and cell power with concentration should be valid. Indeed, J_{sc} may increase even faster than X, on account of the longer carrier lifetime at high injection.

Radiative and Auger recombination, however, become non linear with n in high injection. From Eq. 4.57 it is clear that the radiative recombination rate will vary like

$$U_{rad} = Bn^2$$

when $n = p$, implying a lifetime which varies like $\tau \propto n^{-1}$. Similarly, band-to-band Auger recombination (Eq. 4.66) varies like

$$U_{Aug} = An^3$$

implying that $\tau \propto n^{-2}$. Therefore at injection levels where radiative and Auger processes are dominant, superposition becomes invalid, the dark current term increases with generation rate and V_{oc} rises more slowly than predicted by Eq. 9.28.

The different recombination regimes may be distinguished as a change in the intensity dependence of V_{oc}. If the separation in electron-hole Fermi levels is equal to qV and $n = p$, then from the law of mass action we have

$$n = n_i e^{qV/2k_BT}$$

and so we expect recombination to vary like $\sim e^{qV/2k_BT}$ for SRH process, like $\sim e^{qV/k_BT}$ for radiative processes and like $\sim e^{3qV/2k_BT}$ for Auger. Consequently we expect V_{oc} to vary like $\frac{2k_BT}{q} \ln X$, $\frac{k_BT}{q} \ln X$ and $\frac{2k_BT}{3q} \ln X$, respectively, in the three cases. In the case of silicon, Auger recombination places a limit on the cell performance under concentration.

9.5.3. *Limits to efficiency under concentration*

In Sec. 2.5 we showed how to calculate the efficiency of a solar cell in the radiative limit, where the only loss process is radiative recombination. The net flux absorbed in the cell is the difference between an incident flux term which is proportional to F_s, and an emitted flux term which is proportional to F_e, the geometrical factor for emitted light. For a flat plate cell in a standard sun $F_s/F_e = 2.16 \times 10^{-5}$, because of the small angular width of the sun. Increasing F_s to F_X by concentrating the light (Eq. 9.25) increases the *absorbed* relative to *emitted* fluxes at any value of the chemical potential (see Eq. 2.16). This increases the product of net flux and chemical potential, and hence the limiting efficiency of the solar cell. Maximum concentration is achieved when the angle of acceptance θ_X reaches 90°, and $F_X = F_e = \pi$ radians, so that the cell accepts and emits light over the same angular range. This leads to a maximum efficiency of over 40% if the temperature is unaffected. Another way of achieving the condition that $F_e = F_s$ is to *restrict* the angular range for *emission*. This idea is developed in Chapter 10.

In practice, concentration systems increase the maximum efficiency of high quality Si and GaAs *cells* by a few percent. However, when the optical losses in the concentrating system and the wastage of diffuse light are taken into account, concentrator modules actually perform no more efficiently than one sun modules in AM1.5. Concentrators are of interest primarily for the reduced cost through reduced cell area, and not for improved efficiency.

Changing spectral intensity also changes the optimum band gap. Figure 9.10 shows the limiting efficiency of an ideal photoconverter, calculated as a function of band gap according to Sec. 2.4–2.5, for an AM1.5 spectrum with $X = 1$ and $X = 1000$ ($F_X = 1000F_s$). It is clear that the maximum efficiency point shifts to lower band gap as concentration increases.

Note that this calculation does not include Auger recombination, which is, like radiative recombination, an unavoidable process, nor the effects of temperature.

9.5.4. *Temperature*

High concentration raises the *temperature* of the cell. At higher temperature the intrinsic carrier population

$$n_i^2 \sim e^{-E_g/k_B T}$$

increases, an effect which is enhanced by the shrinkage of the band gap E_g with increasing temperature. Increasing n_i increases the dark current. The reduced band gap increases the photocurrent, but only slightly compared to the effect on J_{dark}, so the increasing dark current dominates. This opposes the effect of increasing efficiency under concentration with the result that optimum performance is achieved at some intermediate concentration level. GaAs has a slightly better temperature coefficient of efficiency than silicon and finds its optimum at a higher concentration level.

9.5.5. *Series resistance*

By increasing the current density, concentration increases the potential drop, JAR_s, due to *series resistance* in the cell. This is a problem particularly in the emitter layer where lateral current densities are high. Concentrator cells must therefore be designed differently, with different metallisation patterns and ideally with higher emitter doping than one-sun cells. Emitter doping levels are limited by recombination and cannot be increased indefinitely, nor can the contacted area become too large, so compromises must be found. Minimisation of series resistance losses is a major objective in concentrator cell design.

9.5.6. *Concentrator cell design*

Concentrator cells should be of very high purity, since high efficiency is a priority, and because the cell is a smaller fraction of the system cost. The concentrator cell should have very low bulk recombination losses, so high quality crystals should be used, and good surfaces. Light trapping is helpful in reducing the cell thickness and thereby reducing resistive losses due to carrier diffusion in the high injection limit. More lightly doped material can

be used for the active layers than for a one sun cell because high carrier densities mean higher conductivity, and also because lower doping produces better quality material. Examples of concentrator cells are the PERL cell, developed at UNSW, ($\eta = 26\%$ in 20 suns) which operates in low injection conditions, and the rear point contact cell, developed at Stanford, which is based on an undoped substrate and operates in high injection [Sinton, 1986; Swanson, 1990]. Buried contacts which combine low resistivity with low shading offer to improve series resistance in future designs.

As noted above, higher emitter current densities mean that series resistance minimisation through contact and emitter design is important. Contact grids should have thicker, denser fingers and often are designed with symmetry to match the image of the concentrated light. They are normally designed to give maximum performance under certain concentration levels, and do not as a rule give the best performance under one-sun conditions. The higher degree of metallisation needed increases losses by reflectivity and shading (to 10–20%), but this can be overcome by encapsulating the cell beneath a prismatic lens which refracts the light onto the bare cell surface and away from the contact fingers.

9.5.7. *Concentrator cell materials*

Figure 9.10 shows that the optimum band gap for photovoltaic energy conversion is reduced under concentration. This would suggest that silicon, with $E_g = 1.1$ eV, is a better concentrator cell material than GaAs with $E_g = 1.4$ eV. However, GaAs has two important advantages: (i) GaAs enjoys a better temperature coefficient of efficiency than silicon and (ii) Auger recombination is stronger in silicon because of its lower band gap and limits the theoretical efficiency to 36% at a concentration of 100 suns with perfect light trapping. Auger recombination is much less important in GaAs and efficiency continues to rise with concentration up to several hundred suns. At 1000 suns the theoretical efficiency is 36–37% [Sinton, 1995; Klausmeier-Brown, 1995].

No simple direct semiconductors exist with band gaps in the range 1.1–1.3 eV, where the theoretical efficiency under concentration is high. Some ternary alloys, such as $In_xGa_{1-x}As$, have suitable band gaps but their material quality is inferior. To date, GaAs remains the best material for high concentration with a record efficiency of 27.6% in AM 1.5 under 255 suns [Vernon, 1991]. The best silicon cell is 26.8% at 96 suns [Verlinden, 1995].

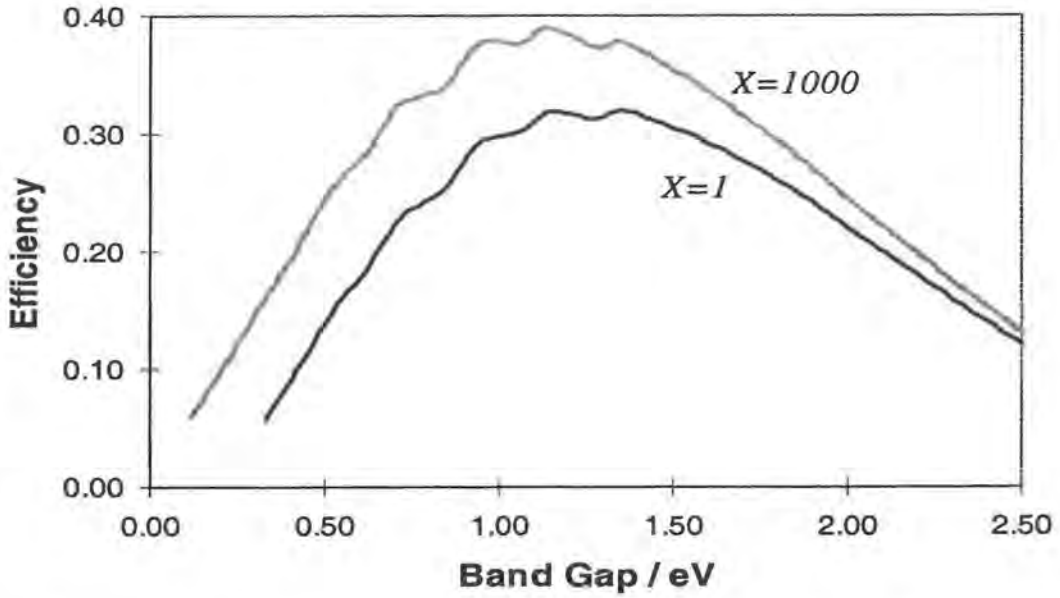


Fig. 9.10. Limiting efficiency versus band gap for an ideal solar cell in AM 1.5 with $X = 1000$ and $X = 1$.

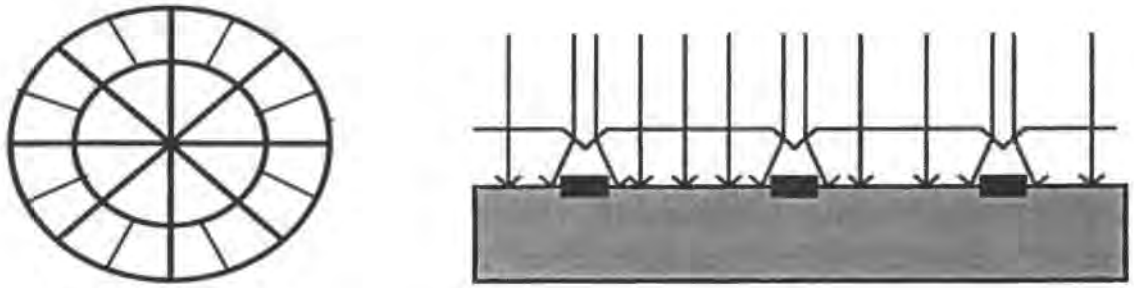


Fig. 9.11. (a) Contact grid for concentrator cell with circular symmetry; (b) Prismatic lens refracts light on to cell surface away from contacts.

9.6. Light Confinement

9.6.1. *Light paths and ray tracing*

In contrast to anti-reflection coats, which increase the fraction of photons admitted to the cell, and concentration which increases the incident flux, light confinement techniques increase the path length of photons *inside* the cell, once admitted. Increasing the path length increases the probability of photogeneration per incident photon, particularly when the absorption coefficient is small, increasing the absorbed fraction. Light trapping is normally considered in the regime of geometrical optics where structures are

large compared to the coherence length of the light, and light rays with different history do not interfere. This is a good approximation in silicon where cells are hundreds of microns thick. In micron scale structures, light should be treated as coherent and interference becomes important. In such systems, classical ray tracing approaches are not valid, and the photogeneration rate must be found from the gradient of the Poynting vector.

Light trapping involves the scattering and reflection of light rays within the cell so that they travel at wider angles to the surface normal. The problem can be considered in terms of the *paths* taken by light rays inside the cell. An admitted ray may travel in one of a number of paths, P_i, which are defined by the cell geometry. At some point within the cell volume, the total flux is given the sum of the contributions along the different paths. The relevant quantity for photogeneration is the light absorption rate, from Eq. 9.7,

$$u_{\rm ph}(E,\mathbf{r}) = \alpha(E,\mathbf{r}) \sum_{P_i} \beta_i(E,\mathbf{r},\theta_i,\phi) \tag{9.30}$$

where the integral over solid angle is converted to a sum over the paths, and β_i represents the photon flux at $\mathbf{r}$ from the direction (θ_i, ϕ_i). Equation 9.30 provides the photogeneration rate at $\mathbf{r}$ and should replace Eq. 9.2 in the calculation of photogeneration. However, because of the different history of the rays it is difficult to determine the β_i at a general point, and it is easier to find the total absorbed flux by considering absorption along the length of each ray path, using an approach known as 'ray tracing'. Thus,

$$f_{\rm abs} = \frac{(1-R(E))}{b_s(E)} \left\{ 1 - R(E) - \sum_i \beta_i(E,0) \exp\left(-\alpha(E) \int_{P_i} dl_i \right) \right\} \tag{9.31}$$

where l_i is a length co-ordinate along the path P_i, and the integration is taken over the ray path, from entry to exit. $\beta_i(E,0)$ is the flux of the ith ray at the start of its path inside the cell — effectively a weighting factor for the path — and satisfies

$$\sum_{P_i} \beta_i(E,0) = (1-R(E)) b_s(E)\,. \tag{9.32}$$

For clarity, we have assumed a position independent absorption coefficient.

Clearly a longer ray path will lead to a higher attenuation of the ray and a higher $f_{\rm abs}$. Different light trapping schemes can be evaluated by

comparing the average path length

$$\langle l \rangle = \sum_{P_i} \beta_i(E,0) \int_i dl_i \Big/ \sum_{P_i} \beta_i(E,0) \tag{9.33}$$

with the width of the cell, w. Note that the path length is dependent only on geometry and reflection and transmission probabilities, and does not contain the probability of photon absorption.

9.6.2. *Mirrors*

The simplest light trapping scheme is to introduce an optical mirror at the rear surface of the cell, either by metallising the rear cell surface or by growing the active layers on top of a Bragg stack. The mirror typically reflects over 95% of rays striking the rear surface. Rays which subsequently reach the front, semiconductor-air, surface are likely to pass through since the reflectivity of that interface must be small for efficient light capture. So the rear mirror effectively doubles the path length of the light. For an ideal mirror (with $R = 1$) and ideal front surface (with $R = 0$) the path length is $2w$.

Greater path length enhancements can be achieved by exploiting total internal reflection at the front surface. This is not possible for a cell with parallel planar surfaces since, by symmetry, reflected rays must approach the front surface at the same angle at which they were admitted, and consequently encounter the same reflectivity, which should be small. For light trapping, rays should approach the front surface at a wider angle than the angle at which they were admitted. This can be achieved by tilting the front or the rear surface, or by scattering the light within the cell. If light rays can be manipulated so that they approach the surface at an angle $\theta_s > \theta_c$

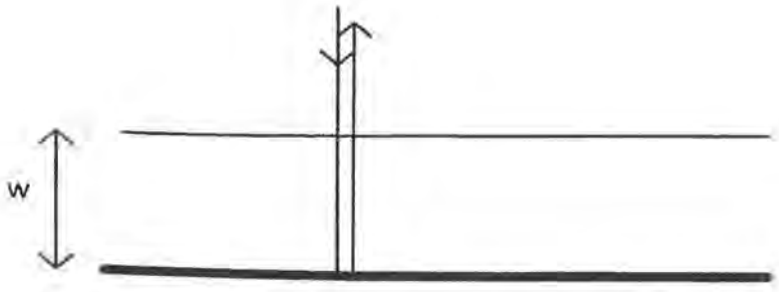


Fig. 9.12. Double path length in metallised cell.

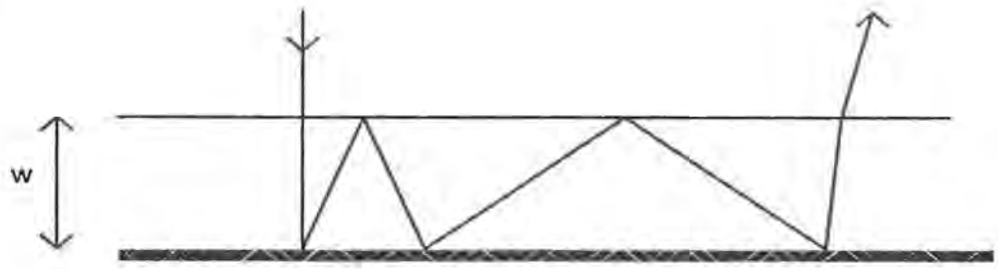


Fig. 9.13. Ideal randomising rear surface. Light rays striking the rear surface are reflected in a random direction. On the first two reflections, the illustrated ray is reflected at an angle larger than θ_c. On the third bounce, it is reflected with $\theta < \theta_c$ and escapes.

they will be unable to escape from the cell and be reflected back for another double pass within the cell.

9.6.3. *Randomising surfaces*

One approach is to use a roughened surface which scatters, or *randomises*, the light upon reflection, changing the value of θ_s. Here we consider a randomising rear surface although in principle either the front or the rear surface, or both, can be randomising. A perfectly randomising surface, called a *Lambertian* surface, scatters reflected rays with spherical symmetry so that all angles in the hemisphere pointing into the cell are equally likely. Assuming normal incidence, a light ray enters and crosses the cell and is scattered at the rear surface. The mean path length for the reflected ray to cross back to the front surface is twice the cell width:

$$\langle l \rangle = \frac{w}{\langle \cos \theta_s \rangle} = 2w\,. \tag{9.34}$$

This doubling of the optical depth of the cell illustrates the advantage of randomising the light even before light confinement has occurred. Now, since only those rays with $\theta_s > \theta_c$ can escape, only $\frac{1}{n_s^2}$ of the rays striking the front surface will leave the cell; the rest will be reflected to make a second pass, of length $2w$, across the cell and be scattered again at the rear surface. Adding up the path lengths of the rays according to the number of times they are scattered before leaving the cell we find

$$\langle l \rangle = \frac{1}{n_s^2} \times 2w + \frac{1}{n_s^2} \times \left(1 - \frac{1}{n_s^2}\right) \times 6w + \frac{1}{n_s^2} \times \left(1 - \frac{1}{n_s^2}\right)^2 \times 10w + \cdots$$

where it is assumed that the path length is w for the first pass (normal incidence) and the final pass (only small θ_s can escape) but $2w$ for all

intermediate passes. Summing the series to infinite order,

$$\langle l \rangle = (4n_s^2 - 2)w \approx 4n_s^2 w \,. \tag{9.35}$$

This ideal path length enhancement, $4n_s^2$, is around 50 for silicon and GaAs, which shows the potential of light trapping systems [Lush, 1991].

If the rear surface is imperfectly reflecting then a fraction of the light is lost through the rear surface at each scattering event, and the mean path length is reduced. The effect can be considerable because of the large number of reflection events (around n_s^2) at that surface.

Although useful for indicating the potential of light trapping systems, perfectly randomising surfaces are not easily realised. Practical techniques usually depend on the geometrical-optical properties of more regular, textured surfaces.

9.6.4. *Textured surfaces*

The simplest system is where one surface is tilted relative to the other. Consider a rear surface tilted at an angle θ_{tilt} relative to the planar front surface, as in Fig. 9.14. When $\theta_{\text{tilt}} > \frac{1}{2}\theta_c$, normally incident rays will be reflected from the rear surface at an angle greater than θ_c, and be totally reflected at the front. If a ray strikes the same portion of the rear surface on the second pass, it will be reflected at an even wider angle, and trapped again. For uniform cell width, both positive and negative tilt angles must be present, so that trapped rays will eventually be reflected at narrower angles and escape. If the positive and negative tilt angles are equal, then each ray makes a multiple of four passes across the cell. The worst case, four passes (a 'double bounce'), is illustrated in Fig. 9.14. The overall mean path length depends upon the size and tilt angles of the texture, but it can

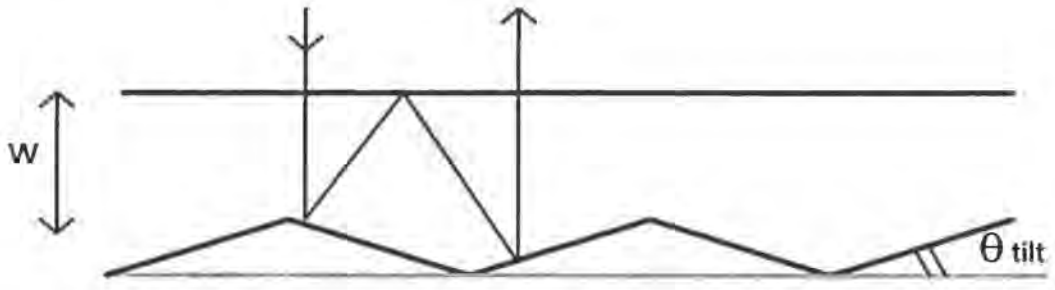


Fig. 9.14. Light trapping structure with a tilted rear surface, illustrating a 'double bounce' light path.

immediately be stated that

$$\langle l \rangle \geq 2(1 + \sec 2\theta_{\text{tilt}})w \gtrsim 4w\,. \tag{9.36}$$

Although higher tilt angles lead to higher path length enhancements for the shortest pass, shallower angles allow longer textured faces, and a higher probability of multiple passes.

Texturing the front surface in the same way leads to similar path length enhancements. Normally incident rays striking the textured surface are refracted into the cell at an angle

$$\gamma = \theta_{\text{tilt}} - \sin^{-1}\left(\frac{\sin\theta_{\text{tilt}}}{n_{\text{s}}}\right) \tag{9.37}$$

which increases the path length from side to side of the cell. If the texturing is symmetrical, then most of those rays will strike the front surface again on a face which has opposite tilt to the face of entry, and leave the cell. So, at the minimum, only a double pass is available and

$$\langle l \rangle \geq 2w\sec\gamma > 2w\,. \tag{9.38}$$

However, if the texturing is asymmetrical, then by choosing the tilt angles carefully, four passes of the light can be achieved, as shown in Fig. 9.15.

As a general trend, the lower the degree of symmetry, the greater the degree of light trapping. If the surface is textured in two directions, with pyramids rather than grooves, then rays are scattered in three dimensions rather than two and light trapping is improved. Calculation of the mean path length for any texturing pattern is complicated and requires the average of ray paths over the different points of incidence by *ray tracing*. However, it is easy to see that for a surface textured with pyramids the probability of a light ray striking the surface at the same angle at which it was admitted — the condition for escape — is smaller than for a surface

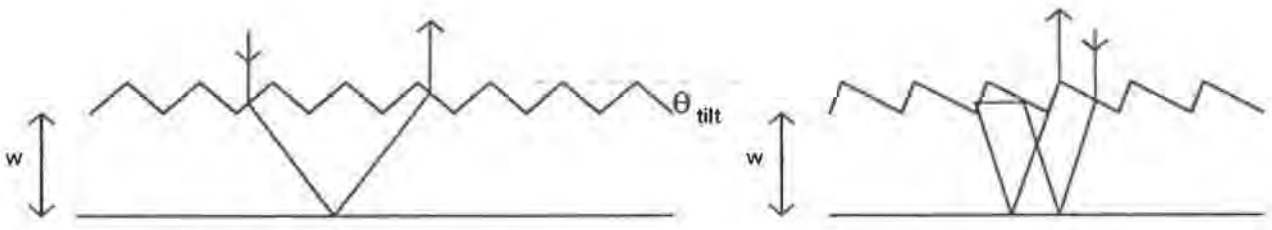


Fig. 9.15. (a) A simple reflection in a structure with symmetrical surface texture. Notice how the surface texture extends the path length from front to rear of the cell. (b) A four pass path in a cell with asymmetric surface texture.

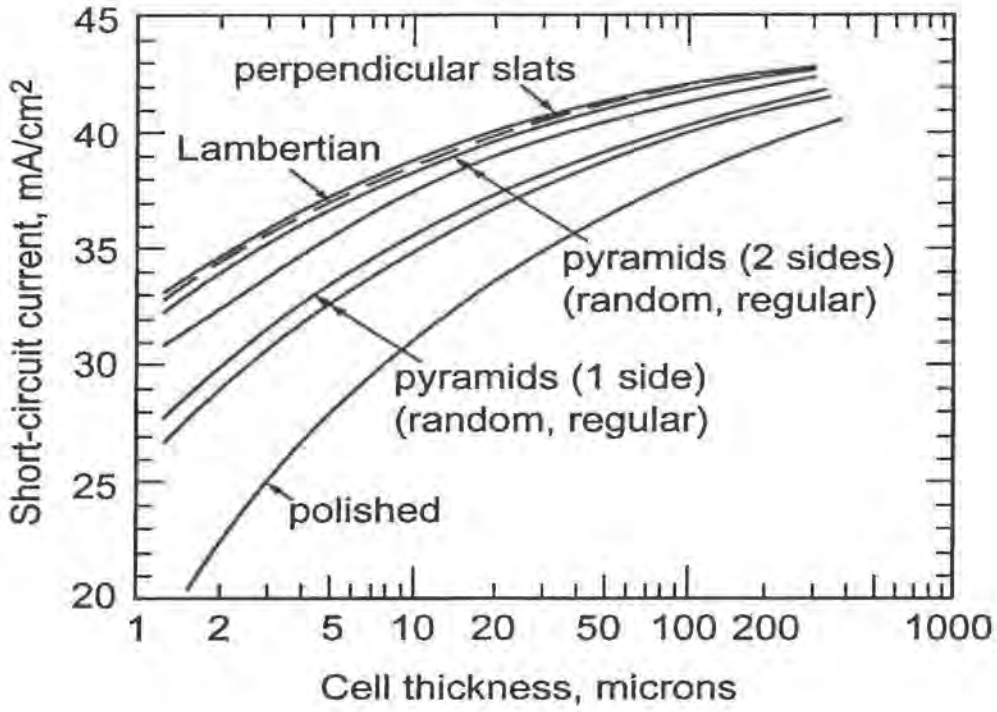


Fig. 9.16. J_{sc} as a function of cell thickness, for different light trapping schemes. [Swanson, 1990.]

textured only with grooves. For a *regular* array of pyramids, light confinement is better if the pyramids are offset from each other. *Random* arrays of pyramids, which have a very low degree of symmetry, trap light very well and are comparable with the ideal randomising surface. Inverted regular pyramids, as used in the PERL cell (Sec. 7.4) perform slightly better than upright ones because of an increased probability of the escaping light re-entering the cell. Figure 9.16 compares the light trapping performance of a range of different surfaces.

9.6.5. *Practical schemes*

Front surface texturing in monocrystalline materials may be achieved by using a selective etch which acts in preferred directions to expose particular crystal planes. In silicon a basic etch can be used on the (100) surface to to expose the (111) crystal faces. The size of surface pyramids can be defined before etching by using a mask.

Surface texturing is less important for polycrystalline materials. Depending on the method of film deposition, the surface is already likely to

be rough and therefore have a lower reflectivity than single crystal. Chemical etches have been used in polycrystalline silicon to create 'honeycomb' and other structures which enhance light trapping.

Textured rear surfaces can be prepared by growing the active layers of the cell on patterened substrates. These can be prepared by etching; or by epitaxial methods, and the active layers then grown epitaxially on top. The size of features on textured surfaces are typically in the μm range, which is acceptable for silicon where cells are at least tens of microns thick, but problematic for highly absorbing materials like GaAs. An alternative is to grow the active layers over a blazed grating [Heine, 1995] which achieves the light scattering without increasing cell width.

Approximately randomising front surfaces have been prepared on silicon using random pyramidal surface texture [Green, 1995], and alternatively with a thin layer of porous silicon on top of the active layers. The porous silicon coating can be produced by electrochemically etching the silicon surface. Nanometre sized crystallites and pores in the porous silicon scatter the light to achieve wider angles of entry [Stalmans, 1998].

Reflective rear surfaces are prepared by metallising the rear of the cell with aluminium or gold. Such semiconductor-metal interfaces have reflectivities of over 95%. Reflectivity at that surface can be improved if a thin layer of dielectric — typically SiO_2 — is inserted between semiconductor and metal. Rays incident at wide angles are then reflected by total internal reflection. For epitaxially grown crystals, an alternative is to grow the active layers on top of a Bragg stack, which has very high reflectivity over a band of wavelengths.

Actual path length enhancements of over 10, rather smaller than the maximum of $4n_s^2$, have been achieved in silicon cells with roughened surfaces and with pyramidal texturing. Note that this is considerably smaller than the flux enhancements available through concentration; unlike concentrators, light trapping cells operate in relatively low injection conditions.

We have not discussed the application of resonant cavities and photonic band gap to photovoltaics. These are useful for strong confinement of particular wavelengths or bands of wavelengths, and may be of interest for photovoltaic effects with narrow band light sources, but are not expected to be beneficial for broad band sources.



(a) (b)

Fig. 9.17. (a) Highly confining cell structure with mirrored surface; (b) A reflective cavity containing arrays of cells (the 'photovoltaic eye').

9.6.6. *Light confining structures: restricted acceptance areas and external cavities*

The degree of light trapping can be increased if the cell is enclosed in a reflective cavity with restricted aperture for acceptance of light. Two examples are illustrated in Fig. 9.17. In the first, the entire surface of the cell is coated with a reflective material, apart from a small aperture which is open to incident light. Rays admitted through this aperture will be reflected many times before they escape and have a path length many times the width of the cell. In the second example, many cells are mounted in an reflective cavity which has a small aperture for acceptance of light. In this case, the path length of rays inside the cavity is long, and rays which are not absorbed in one cell are likely to be absorbed by another, so the path length of rays in the assembly of cells is long compared to the width of a cell [Minano, 1990].

In both cases the aperture area, A_{aperture}, is smaller than the geometrical surface area of the cell or cells, A_{cell}, and although the optical path length is increased, say by $\frac{\langle l \rangle}{w}$, the photon flux inside the cell material is increased by a much smaller factor, approximately $\frac{\langle l \rangle}{w} \times \frac{A_{\text{aperture}}}{A_{\text{cell}}}$, which is comparable to the enhancement available with a rear mirror. Such cavity structures are really only useful in conjunction with concentrating systems, where light is concentrated on to the aperture to increase the photon flux admitted and consequently increase the photon flux density inside the cell material.

9.6.7. *Effects of light trapping on device physics*

Incorporating light trapping can be compared to increasing the cell thickness. To see this, compare a standard cell of width w without rear reflector with a light trapping cell of width $W = \frac{w}{X}$ and a mean optical path length of $\langle l \rangle = WX$. The extended path length compensates for the reduced cell thickness so that, to a first approximation, the fraction of light absorbed is the same in either case.

Now we can make the following observations.

- Bulk recombination losses are reduced. The thickness of the light trapping cell is smaller compared to the minority carrier diffusion length, so a smaller fraction of the photogenerated carriers should be lost by minority carrier recombination than in the standard cell.
- Rear surface recombination losses may be increased. The point generation rate is made up of contributions from reflected and incident rays and is more uniform than for the standard cell. This means that generation will be significant near the rear surface, and the fraction of carriers available for recombination there is higher. Very good quality rear surfaces are therefore needed, with low recombination velocities as well as very high optical reflectivities.
- Bulk resistivity losses are reduced. Because the same amount of generation is occurring in a smaller volume, the photogenerated carrier densities are higher, and the conductivity is increased.
- The probability Auger recombination is increased, again because of higher carrier densities.

The procedure for calculating the carrier and current densities in the light trapping structure is similar to that used for a standard cell, such as the p–n junction example in Chapter 6. The main modification is that the generation rate, given by Eq. 9.2, must be replaced by a function, such as Eq. 9.30, which incorporates the contributions from rays which have been reflected from the front and rear surfaces. Provided that the carrier densities are still low enough for the Shockley Read Hall recombination rate to vary linearly with carrier density in the neutral regions, *i.e.*, to satisfy low injection conditions, then the net current still decouples into a bias dependent dark current and a light dependent photocurrent. Comparing the same two cells as above, we see that the photocurrent should be similar in the two cases, although it may be slightly higher in the light trapping cell because of reduced bulk recombination losses. The dark current at any

given applied bias will be smaller for the light trapping structure because it is thinner. Consequently the net photocurrent, $J(V) = J_{sc} - J_{dark}(V)$, will be increased.

The low injection approximation is likely to be valid for *practical* light trapping structures where path length enhancements are seldom greater than 10. In some very lightly doped device structures light trapping may push the device into the high injection regime, where recombination becomes non-linear. This regime is discussed above in Sec. 9.5.2.

Comparison of light confinement with concentration

Let's compare a concentrator cell of width w and concentration factor X with a light trapping cell of the same width and path length enhancement $\langle l \rangle = wX$. The cells have the same volume and the same average photon flux density. If only bulk recombination is important, then the cells should perform equivalently since there will be the same probability of a recombination event per generation event in either case. However, if surface recombination *is* important, then the light trapping cell will perform less well. Because of the higher number of reflections per generation event, more carriers are likely to be generated near the rear surface, and surface recombination will therefore be greater.

The maximum theoretical power density enhancement available with light trapping is the same as that available with concentration: $4n^2/\sin^2\theta_{sun}$ where θ_{sun} is the half angle subtended by the sun. In practice, however, it is harder to achieve high enhancement ratios with light trapping than to achieve high concentration factors.

9.7. Photon Recycling

9.7.1. *Theory of photon recycling*

Photon recycling (PR) is the reabsorption of photons emitted by radiative recombination inside the cell. Until now, it has been assumed that the photon generation term in Eq. 9.6 is negligible. This is a good approximation in materials where non-radiative processes dominate. Nevertheless, we know from Chapters 2 and 4 that in a semiconductor away from equilibrium, radiative recombination of electrons and hole is unavoidable, and so there is in fact a non-zero photon generation term in the photon continuity equation, though it may be extremely small. Including this term, Eq. 9.9 should

be used

$$\frac{d\beta}{dl} = \alpha\beta_e - \alpha\beta \tag{9.39}$$

where we have used the detailed balance result, that $\varepsilon = \alpha$. $\beta(E, \mathbf{r}, \theta, \phi)$ is, as usual, the component of the flux along (θ, ϕ), and β_e is the magnitude of emitted flux per solid angle, given by Eq. 2.11

$$\beta_e(E, \mathbf{r}, \theta, \phi) = \frac{2n_s^2}{h^3c^2} \frac{E^2}{e^{(E-\Delta\mu)/kT_c} - 1} \tag{9.40}$$

where $\Delta\mu$ is the splitting in the quasi Fermi levels at $\mathbf{r}$ and n_s the refractive index. Thus, when photon recycling is included, the photon flux density at any point within the cell is increased by the spontaneous emission of photons from other points.

Let's look at the solution to the inhomogenous differential equation, Eq. 9.39. If we consider a system with planar symmetry, then β depends only on the depth x into the cell and we may convert the co-ordinate l to x. For light normally incident on to a planar cell, we define

$$l = x \sec\theta \tag{9.41}$$

where θ is the angle between the ray direction and normal to the cell surface.

In the absence of photon recycling, the first term on the right-hand side of Eq. 9.39 is zero and the solution for β is the Beer–Lambert form. This is the *externally generated* flux, and it has the form

$$\beta(x) = (1-R)\beta_s e^{-\alpha x \sec\theta} = \beta(0)e^{-\alpha x \sec\theta} \quad \text{for } 0 \le \theta \le \theta_c \tag{9.42}$$

where β_s is the angularly resolved emission from the sun (Eq. 2.1) and α is assumed uniform. In the following, the parameter E has been dropped for clarity and β_e is expressed as a function of x, indicating the dependence on $\Delta\mu$ at x.

Now, when photon recycling is included, Eq. 9.39 has the general solution

$$\begin{aligned} \beta(x) &= \beta(0)e^{-\alpha x \sec\theta} + \Delta\beta(x) \\ \Delta\beta(x) &= \int_0^x \alpha\beta_e(x')e^{-\alpha(x-x')\sec\theta} \sec\theta dx' \end{aligned} \tag{9.43}$$

which shows that the flux at any point inside the cell is increased by contributions emitted from all other points. To find the carrier generation rate

at x we sum the flux over all ray directions, (Eq. 9.8)

$$g(E,x) = \int \alpha\beta(x)d\Omega = 2\pi \int \alpha\beta(x) \quad d\cos\theta\,, \tag{9.44}$$

whence

$$g(E,x) = (1-R)\alpha b_s e^{-\alpha x} + \alpha \int_{-1}^{1}\int_{0}^{x} \alpha\beta_e(x')e^{-\alpha(x-x')\sec\theta}\sec\theta dx' d\cos\theta \tag{9.45}$$

where we have used the fact that $\cos\theta_c \sim 1$ at a semiconductor-air interface, and the relationship between $\beta(0)$ and b_s. Renaming the terms we have

$$g(E,x) = g_{ext}(E,x) + g_{in}(E,x) \tag{9.46}$$

where the first term in Eq. 9.46 is solely due to *external* generation, and the second to *internal* generation. It is clear from Eqs. 9.40 and 9.45 that $g_{in}(E,x)$ depends on the electron and hole densities at x' through the quasi Fermi level separation at that point, $\Delta\mu(x')$, and hence is coupled to the differential equations for electron density, hole density and Poisson's equation. In general, the integral Eq. 9.44 for $g(E,x)$ needs to be added to the differential system and solved numerically. The relationship between $g_{in}(E,x)$ and $\Delta\mu(x)$ depends upon the geometry of the cell and degree of light trapping, as is illustrated in the example below.

Some approximations can be made in the calculation of photon recycling effects [Parrott, 1993; Balenzategui, 1995; Marti, 1997].

(i) Most simply, $\Delta\mu(x)$ is set to a constant, whence $\beta_e(x)$ is constant. This is a good approximation for very thin cells, where $g_{ext}(E,x)$ is roughly constant;

(ii) $\Delta\mu(x)$ is set to the value it would have if photon recycling was ignored, and $\beta_e(x)$ is worked out from that;

(iii) Iterating (ii), $\Delta\mu(x)$ is recalculated from the approximate result for $\beta_e(x)$, and that is used to improve the solution for $\Delta\mu(x)$. Successive iterations should yield a solution for $\Delta\mu(x)$ which is stable to some desired tolerance. Using this method, it was estimated that photon recycling could increase the efficiency of a GaAs cell with a rear mirror and high radiative efficiency by 1.3%, and the V_{oc} by tens of mV [Balenzategui, 1995]. Another way to look at photon recycling is an extension of the radiative *lifetime* leading to reduced radiative recombination rate.

9.7.2. *Practical schemes*

Unlike other strategies discussed above, photon recycling is an intrinsic process which happens in all illuminated semiconductors. It is relevant only in materials with high radiative efficiency — direct band gap semiconductors such as GaAs — and can be exploited only in devices which are already limited by radiative recombination. Since all practical solar cells are limited by non-radiative processes, managing photon recycling to achieve higher efficiency is still a theoretical objective.

Photon recycling clearly increases with average ray path length and therefore with the degree of light trapping. As shown in the example below, the effect is stronger in a cell with good rear reflectivity. As material quality improves, photon recycling will need to be included as a factor in the design of highly efficient solar cells. Note that the effect is included by default in the calculation of limiting efficiency in Chapters 2 and 10.

Box 9.2. Photon recycling in a cell with rear mirror

$g_{\text{in}}(x)$ depends upon the geometry of the cell and degree of light trapping. Here we consider the case of a planar cell which has a perfectly reflecting rear surface and a front surface with zero reflectivity within the critical angle. The additional flux at a depth x is given by the second term in Eq. 9.45. The angular integral contains contributions from the following three types of trajectory:

- $\theta < \theta_c$ Ray path direct from front surface. No reflections. (Left-hand ray in Fig. 9.18.)
- $\theta_c < \theta < (\pi - \theta_c)$ Ray reflected more than once from front or rear surfaces. (Centre rays in Fig. 9.18.)

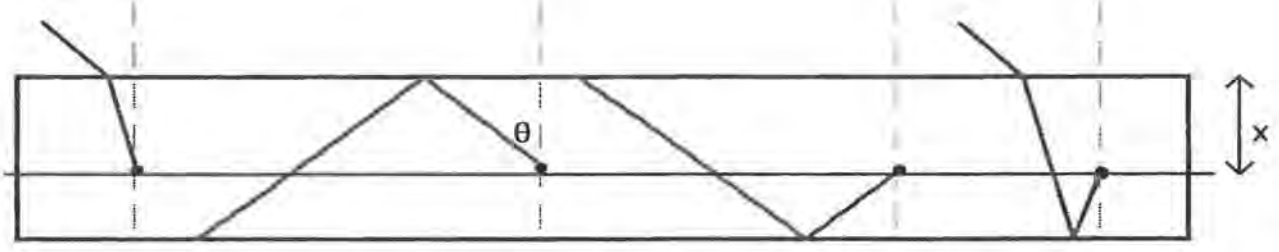


Fig. 9.18. Photon recycling in a cell with a perfect rear reflector. The figure illustrates the four types of rays which need to be included.

- $\pi - \theta_c < \theta < \pi$ Ray reflected once at rear surface. (Right hand ray in Fig. 9.18).

From Eq. 9.43 we have for $\theta < \theta_c$,

$$\Delta\beta(x) = \int_0^x \alpha\beta_e(x')e^{-\alpha(x-x')\sec\theta}\sec\theta dx' \tag{9.47a}$$

for $\theta_c < \theta < (\pi - \theta_c)$,

$$\Delta\beta(x) = \int_0^x \alpha\beta_e(x')e^{-\alpha(x-x')\sec\theta}\sec\theta dx' + \left(\frac{1}{1-e^{-2\alpha w\sec\theta}}\right)\int_0^{2w} \alpha\beta_e(x')e^{-\alpha(x-x')\sec\theta}\sec\theta dx' \tag{9.47b}$$

and for $\pi - \theta_c < \theta < \pi$,

$$\Delta\beta(x) = \int_x^w \alpha\beta_e(x')e^{-\alpha(x-x')\sec\theta}\sec\theta dx' + e^{-\alpha(w-x)\sec\theta}\int_0^w \alpha\beta_e(x')e^{-\alpha(w-x')\sec\theta}\sec\theta dx' . \tag{9.47c}$$

Equation 9.47 can be solved iteratively to yield a solution for $\Delta\beta(x)$ and corresponding $\Delta\mu(x)$. Notice that potentially the largest contribution comes from the second term of the second result, which accounts for multiple reflections of internally reflected rays. The higher the degree of light trapping, the greater the effect on $\Delta\beta$.

9.8. Summary

Photocurrent generation in a solar cell can be increased if the photon flux density within the cell can be increased. The total photon flux available for photogeneration consists of the flux density incident on the cell surface, less the fractions which are reflected and transmistted without being absorbed. The incident flux density can be increased by concentrating the light; reflectivity can be reduced by surface treatments; and the transmitted component reduced using light trapping structures.

Concentration of solar radiation using reflective optics or Fresnel lenses increases the flux density incident on the cell surface. The main motivation is reduction in system cost due to reductions in the amount of photovoltaic material needed. In theory, concentration ratios of over 46,000 are possible but in practice the detrimental effects of increased temperature and series

resistance losses limit concentration ratios to a few hundred suns. Exploiting concentration effectively requires mechanical tracking systems which follow the path of the sun, and this adds to the cost. Lower levels of concentration can be achieved with passive, non-tracking optical concentrators. Concentration increases the limiting efficiency to over 40%, and efficiencies of up to 28% have been achieved in practice. It is primarily of interest for high efficiency solar cells. Under high concentration levels, cells work in the 'high injection' limit where carrier densities are high, recombination through defects is suppressed and Auger and radiative recombination processes become more important.

The reflectivity of a semiconductor at visible wavelengths is typically over 30%. This can be reduced to a few percent using an anti-reflection coating, which is a thin film of dielectric material with refractive index smaller than the cell material. Anti-reflection coats are normally optimised for a certain wavelength, and are less effective at different wavelengths and off-normal incidence. They are routinely used in crystalline solar cells. Reflectivity may also be reduced by texturing the surface, and this technique is used in crystalline and polycrystalline materials.

Light trapping structures confine the light by exploiting internal reflection and scattering. Strategies include reflective mirrors at the rear of the cell, textured front and rear surfaces. The effective path length of a light ray within the cell is increased by multiple reflections or by refraction at wide angles to the normal. The photon flux density at any point within the cell is increased by contributions from internally reflected rays, resulting in higher effective optical depth. Light trapping is usually treated with geometrical optics and quantified by the enhancement of the optical path length compared to the cell without light trapping. Path length enhancements of around 50 are theoretically possible, and enhancements of around 10 have been reported in practice. Light trapping is routinely used in silicon and is of strong interest for thin film silicon photovoltaics.

In materials with high radiative efficiencies, the useful flux density can be increased by reabsorption of emitted photons or 'photon recycling'. The photon flux at a point is increased by the emitted flux from other points due to the light-induced quasi Fermi level separation, and the quasi Fermi level separation is increased by the enhanced light absorption. Photon recycling effects are useful only in materials where radiative recombination is an important loss mechanism and they are larger within light trapping structures. They are not yet exploited within any practical cell structures.

References

J.L. Balenzategui, A. Marti and G.L. Araujo, "Numerical modelling of photon recycling effects in solar cells using PC1D", *Proc. 13th European Photovoltaic Solar Energy Conf.*, 1223–1226 (1995).

S. Chandrasekhar, *Radiative Transfer* (Oxford University Press, 1950).

A. Goetzberger and W. Greubel, "Solar energy conversion with fluorescent concentrators", *Appl. Phys. Letts.* **14**, 123 (1977).

M.A. Green, *Silicon Solar Cells: Advanced Principles and Practice* (Sydney: Centre for Photovoltaic Engineering, 1995).

C. Heine and R.H. Morf, "Submicrometer gratings for solar-energy applications", *Appl. Opt.* **34**, 2476 (1995).

M.E. Klausmeier-Brown, "Status, prospects and economics of terrestrial single junction GaAs concentrator cells", in *Solar Cells and their Applications*, ed. L.D. Partain (Wiley, 1995).

G. Lush and M. Lundstrom, "Thin film approaches to high efficiency III-V cells", *Solar Cells* **30**, 337–344 (1991).

H.A. Macleod, *Thin Film Optical Filters* (Bristol: Institite of Physics, 2001).

A. Marti, J.L. Balenzategui and R.F. Reyna, "Photon recycling and Shockley's diode equation", *J. Appl. Phys.* **82**, 4067–4075 (1997).

J.C. Minano, "Optical confinement in photovoltaics", in *Physical Limitations to Photovoltaic Energy Conversion*, eds. A. Luque and G.L. Araujo (Adam Hilger, 1990).

J.I. Pankove, *Optical Processes in Semiconductors* (Englewood Cliffs: Prentice Hall, 1971).

J.E. Parrott, "Radiative recombination and photon recycling in photovoltaic solar cells", *Sol. En. Materials Sol. Cells* **30**, 221–231 (1993).

R. Sinton *et al.*, "27.5% silicon concentrator solar cells", *Electron Device Letts.* **EDL-7**, 567 (1986).

R.A. Sinton, "Terrestrial silicon concentrator cells", in *Solar Cells and their Applications*, ed. L.D. Partain (Wiley, 1995).

L. Stalmans *et al.*, "Porous silicon in crystalline silicon solar cells: A review and the effect on the internal quantum efficiency", *Progr. Photovoltaics* **6**, 233–246 (1998).

R.M. Swanson and R.A. Sinton, "High efficiency silicon solar cells", *Advances in Solar Energy* **6**, 427–484 (1990).

P.J. Verlinden *et al.*, "A 26.8% efficient concentrator point-contact solar cell", *Proc. 13th European Photovoltaic Solar Energy Conf.*, 1582–1585 (1995).

S.M. Vernon *et al.*, "High-efficiency concentrator cells from GaAs on Si", *Conf. Record, 22nd IEEE Photovoltaic Specialists Conf.*, 353–357 (1991).

W.T. Welford, "Limits to concentration by passive optical systems", in *Physical Limitations to Photovoltaic Energy Conversion*, eds. A. Luque and G.L. Araujo (Adam Hilger, 1990).

Chapter 10

Over the Limit: Strategies for High Efficiency

10.1. Introduction

A photovoltaic cell operates by converting light energy into electricity. It is a quantum energy conversion process, whereby packets of light energy, *photons*, are consumed to deliver units of electrical charge (*electrons*) to an external circuit where they do electrical work. The power conversion *efficiency* of a solar cell is a measure of the amount of work done per photon. Increasing the efficiency is essentially a matter of increasing the ratio of work extracted to photon energy supplied. This chapter is concerned with ways — both practical and theoretical — of increasing that ratio.

In Chapter 2 we saw how the limiting efficiency of a simple, single junction solar cell is calculated by considering detailed balance. This limit, of 33% in a standard air mass 1.5 (AM 1.5) spectrum, follows from a number of assumptions. Important amongst these are the assumptions that exactly one electron-hole pair is generated by each absorbed photon, that electrons and holes relax to form populations in thermal equilibrium with the lattice, and that all the light is absorbed in a junction of single band gap. Given these constraints, the majority of the sun's energy is lost by (i) failure to capture the energy of photons with energy (E) smaller than the band gap (E_g), and (ii) thermal dissipation of kinetic energy of carriers generated with $E \gg E_g$, as shown in Fig. 10.1. Processes (i) and (ii) account for about 23% and 33% of the sun's energy, respectively. The ratio of work out to photon energy in is greatest for photons with energy close to the band gap; elsewhere it is zero (for the low energy photons where $E < E_g$), or poor (for $E \gg E_g$). Figure 2.9 compares the power available at different photon energies with the incident power spectrum. Losses result from the poor match of the single band gap of the solar cell with the broad spectrum of solar radiation.

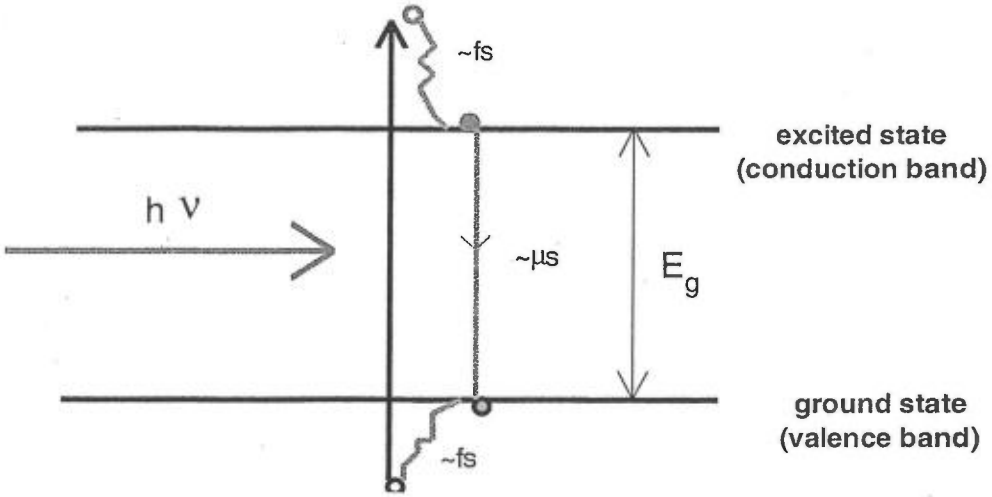


Fig. 10.1. Schematic energy band diagram for a single band gap solar cell. Photons of energy $h\nu$ greater than the band gap, E_g, are capable of promoting electrons to the excited state (the conduction band in a semiconductor). The excess energy is quickly lost as heat as the excited carriers thermalise to the band edges. Relaxation across the band gap is much slower, and in a photovoltaic device the thermalised carriers can be extracted before they finally recombine.

Chapters 3–9 were all concerned with devices and materials which operate within the limits presented in Chapter 2. In this chapter we discuss ways in which the theoretical limiting efficiency may be increased. A separate route to increasing the limiting efficiency, by concentration of solar radiation to expand the acceptance solid angle, has already been discussed in Chapter 9.

In the following sections we first review the calculation of the detailed balance limiting efficiency and underlying assumptions. Then we consider a number of strategies for increasing the amount of work done per photon by relaxing the assumptions underlying the detailed balance limit. Briefly, these strategies are concerned with:

- increasing the number of band gaps, to utilise different photon energies more efficiently (tandem and other multi-band solar cells);
- reducing the dissipation of thermal energy by photogenerated carriers ('Hot carrier' solar cells); and
- increasing the number of electron-hole pairs per photon (impact ionisation solar cells).

Apart from tandem solar cells, the strategies are mainly theoretical and would require materials with properties which have not yet been realised.

However, all are based on physical observations which suggest that the proposed phenomena *may* be exploited in practice.

10.2. How Much is Out There? Thermodynamic Limits to Efficiency

We start by calculating the maximum work available from a photoconverter considering only thermodynamic arguments. Essentially, the solar cell is a cool body which is radiatively coupled to its environment and operates by absorbing short wavelength radiation from a hot body (the sun) at temperature T_s and allowing some of the absorbed energy to be extracted as work. To satisfy detailed balance the device must also emit longer wavelength radiation characteristic of its own temperature T_c. In the simplest picture, both sun and cell absorb and emit light like black bodies at their respective temperatures. A black body at temperature T emits an amount of energy given by σT^4 per unit time and surface area, where σ_S is Stefan's constant. If the incident radiation is fully concentrated and the only loss process is spontaneous emission by the cell, then the net energy flux density received by the solar cell is given by

$$\sigma_S T_s^4 - \sigma_S T_c^4 .$$

Work can be extracted from the hot cell by means of a heat engine. Least energy is lost if the work is extracted isoentropically by means of a Carnot engine. Then the work available is

$$W = (\sigma_S T_s^4 - \sigma_S T_c^4)\left(1 - \frac{T_a}{T_c}\right) \tag{10.1}$$

where the final factor represents the Carnot factor for extraction of work from the cell at constant entropy. The power conversion efficiency is the ratio of extracted to incident power,

$$\eta = \frac{W}{\sigma_S T_s^4} = \left(1 - \left(\frac{T_c}{T_a}\right)^4\right)\left(1 - \frac{T_a}{T_c}\right). \tag{10.2}$$

For a sun temperature of 5760K and cell temperature of 300K, this has a maximum of 85% at a cell temperature of 2470K. In this limit, all photons are absorbed and the maximum amount of work is extracted per photon. Thermal dissipation is not considered. This limit is identical to the limiting efficiency of a solar thermal converter.

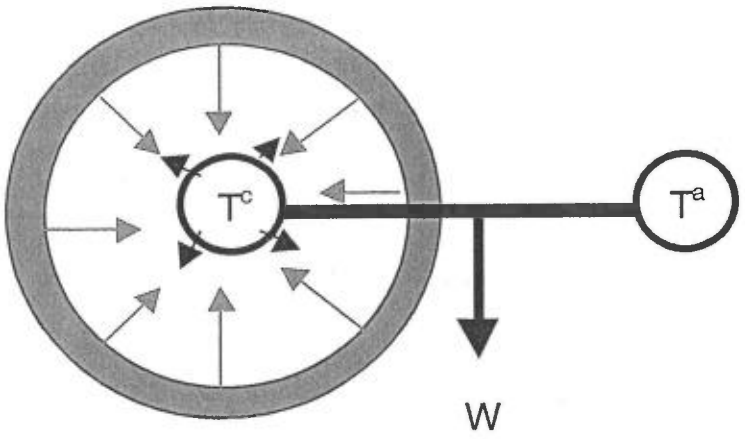


Fig. 10.2. Schematic of an ideal solar energy converter. The incident radiation is fully concentrated so that the cell absorbs and emits radiation in all directions. The cell, at T_c, is coupled to the environment, at T_a, through a heat engine which extracts work, W.

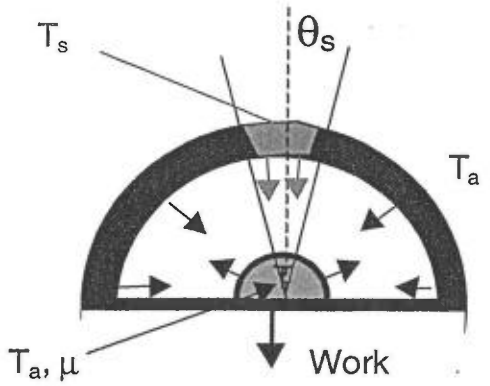
10.3. Detailed Balance Limit to Efficiency, Reviewed

In Chapter 2, we calculated the limiting efficiency of a two level, or single band gap, photoconverter. Here we review that calculation and and introduce some of the concepts which will be used later in this chapter. A more detailed treatment of detailed balance as applied to solar cells may be found elsewhere [Araujo, 1994; de Vos, 1990].

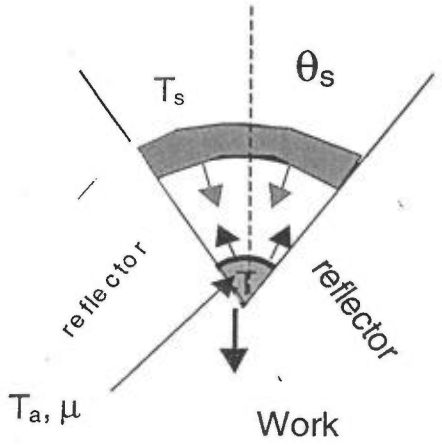
The photoconverter has a band gap E_g, below which photons are not absorbed, and above which photons are absorbed to generate electron-hole pairs in thermal equilibrium with the lattice. It is usually assumed that the cell has planar geometry, and is able to receive and emit radiation into a hemishpere. For unconcentrated sunlight, the cell receives solar radiation through a small azimuthal range $0 < \theta < \theta_{sun}$ and ambient radiation through the remaining range $\theta_{sun} < \theta < \pi/2$. Assuming that the only loss process is radiative recombination, then the current output is equal to q times the difference between the photon fluxes which are absorbed and emitted through the cell surface, where q is the charge on the electron. Formally,

$$I = q \left\{ \begin{array}{l} \iint a(E,\mathrm{s},\theta,\phi)\beta_\mathrm{s}(E,\mathrm{s},\theta,\phi)dE\, d\mathbf{S}\cdot d\Omega \\ \quad + \iint a(E,\mathrm{s},\theta,\phi)\beta_\mathrm{a}(E,\mathrm{s},\theta,\phi)dE\, d\mathbf{S}\cdot d\Omega \\ \quad - \iint a(E,\mathrm{s},\theta,\phi)\beta_\mathrm{e}(E,\mathrm{s},\theta,\phi)dE\, d\mathbf{S}\cdot d\Omega \end{array} \right\} \qquad (10.3)$$

where $a(E, \mathrm{s}, \theta, \phi)$ is the *absorptivity*, or probability of photon absorption, for light of energy E approaching along the direction (θ, ϕ) at a point s on the surface. $\varepsilon(E, \mathrm{s}, \theta, \phi)$ is the *emissivity*, or probability of photon emission, at that point. β, β_a and β_e are the photon fluxes incident from the sun, incident from the ambient, and emitted from the cell, respectively. Photon flux is defined as the number of photons per unit solid angle passing through unit area in unit time. The angular integrals for β_s and β_a should be taken over the appropriate range, as shown in Fig. 10.3.



(a)



(b)

Fig. 10.3. (a) In the calculation of limiting efficiency, a flat plate cell at T_a receives light from sun within an angular range $\theta < \theta_\mathrm{sun}$. The cell emits into a hemisphere. Emission into the rear hemisphere is prevented by a reflector. The chemical potential of the cell is maintained at a positive value, $\Delta\mu$, by the incident radiation. For concentrated light, the angular range of the sun is increased and with it the efficiency of the cell; (b) When the angular range of emission is restricted to be the same as the range of incidence, the performance is equivalent to the case of full concentration.

It is assumed for simplicity that both sun and ambient radiate like black bodies, producing isotropic photon fluxes described by the Planck distribution. So for the black body sun at T_s, we have

$$\beta_s = \frac{2}{h^3c^2}\frac{E^2}{e^{E/k_B T_s} - 1} \tag{10.4}$$

over the range $0 < \theta < \theta_{sun}$. For the ambient environment at temperature T_a we have

$$\beta_a = \frac{2}{h^3c^2}\frac{E^2}{e^{E/k_B T_s} - 1} \tag{10.5}$$

over the range $\theta_{sun} < \theta < \pi/2$.

The cell is assumed to possess a uniform chemical potential, $\Delta\mu$, on account of the light, and to be in thermal equilibrium with the ambient at T_a. This leads to spontaneous emission and produces an isotropic emitted flux of

$$\beta_e = \frac{2}{h^3c^2}\frac{E^2}{e^{(E-\Delta_\mu)/k_B T_s} - 1} \tag{10.6}$$

for $0 < \theta < \pi/2$.

For maximum performance the cell should absorb all incident photons above band gap. Therefore we assume

$$a(E) = \left\{ \begin{array}{ll} 1 & E \geq E_g \\ 0 & E < E_g \end{array} \right\}. \tag{10.7}$$

Two further simplifications result from the following. First, we assume that the chemical potential is constant and equal to q times the applied bias,

$$\Delta\mu = qV\,. \tag{10.8}$$

This implies that no potential is lost between light absorption and charge collection, *i.e.*, charge transport is lossless and mobilities are infinite. Second, we use the generalised detailed balance result that $\varepsilon(E) = a(E)$. This has been shown to be the case when when $\Delta\mu$ is uniform and equal to qV [Araujo, 1994].

For the geometry of a flat plate photoconverter of area A with a perfect rear reflector, the angular integrals can now be evaluated easily. Using the

limits for θ given above we find for the current, I, at a bias V

$$I(V) = qA\left\{\frac{2F_s}{h^3c^2}\int \frac{E^2}{e^{E/k_BT_s}-1}dE + \frac{2(F_a-F_s)}{h^3c^2}\int \frac{E^2}{e^{E/k_BT_a}-1}dE - \frac{2F_a}{h^3c^2}\int \frac{E^2}{e^{(E-qV)/k_BT_a}-1}dE\right\} \tag{10.9}$$

where the constants

$$F_a = \pi \tag{10.10}$$

and

$$F_s = \pi f_s \tag{10.11}$$

where $f_s = \sin^2\theta_s = 2.16 \times 10^{-5}$ represents the angular range of the sun. If the light were concentrated by a factor X then

$$F_s = \pi X f_s\,. \tag{10.12}$$

Full concentration is reached when $X = 1/f_s$ and $F_s = \pi$.

It will be useful to define the function

$$N(E_{min}, E_{max}, T, \Delta\mu) = \frac{2\pi}{h^3c^2}\int_{E_{min}}^{E_{max}} \frac{E^2}{e^{(E-\Delta\mu)/k_BT}-1}dE \tag{10.13}$$

representing the maximum absorbed or emitted *photon* flux density integrated over an energy band E_{min} to E_{max}, and the function

$$L(E_{min}, E_{max}, T, \Delta\mu) = \frac{2\pi}{h^3c^2}\int_{E_{min}}^{E_{max}} \frac{E^3}{e^{(E-\Delta\mu)/k_BT}-1}dE \tag{10.14}$$

representing the emitted *energy* flux density. Then we have for the current density,

$$J(V) = \frac{I(V)}{A} = q\{Xf_sN(E_g,\infty,T_s,0) + (1-Xf_s)N(E_g,\infty,T_a,0) - N(E_g,\infty,T_a,qV)\}\,. \tag{10.15}$$

The output *power* density, given by

$$P(V) = V \times J(V) \tag{10.16}$$

is now a function of E_g and V. For each value of E_g there exists some value of V, V_m, in the range 0 to E_g, for which $P(V)$ is a maximum, *i.e.*,

$$P_{max} = P(V_m) \tag{10.17}$$

where

$$\left.\frac{\partial P}{\partial V}\right|_{V=V_m} = 0\,. \tag{10.18}$$

The *maximum* efficiency is then a function only of E_g and X and is given by

$$\eta_{max} = \frac{P(V_m)}{P_s} \tag{10.19}$$

where

$$P_s = X f_s L(0, \infty, T_s, 0) \tag{10.20}$$

represents the power received from the (black body) sun. For unconcentrated sunlight ($X = 1$) η_{max} has a maximum of 31% at $E_g = 1.3$ eV [Araujo, 1994].

In Chapter 9 we saw how concentrating the light improves the balance between absorbed and emitted flux. At the maximum concentration, $X = 1/f_s$, and Eq. 10.15 becomes

$$J(V) = q\{N(E_g, \infty, T_s, 0) - N(E_g, \infty, T_a, qV)\} \tag{10.21}$$

leading to a maximum efficiency of around 41% at $E_g = 1.1$ eV. Calculated limiting efficiency as a function of band gap for a black body sun is shown in Fig. 10.4. The same analysis can be done using the standard AM1.5 solar spectrum, as shown in Fig. 9.10, and leads to very similar values for the limiting efficiency and optimum band gap.

It is straightforward to show that the same limit results if the angular range of emission is limited to the angular range of the sun. This may be achieved by inserting the cell inside a reflective cavity which admits solar and thermal radiation only through a single aperture, as shown in Fig. 10.3(a). In most of the examples considered in this chapter, we will consider only the case of full concentration, for clarity.

Summarising the assumptions made in the detailed balance analysis:

(i) All photons of energy greater than E_g, are absorbed to create one electron-hole pair. All photons with $E < E_g$ are not absorbed.
(ii) The electron and hole populations relax to form separate, distributions in quasi thermal equilibrium with the lattice at temperature T_a and with quasi Fermi levels separated by $\Delta\mu$.

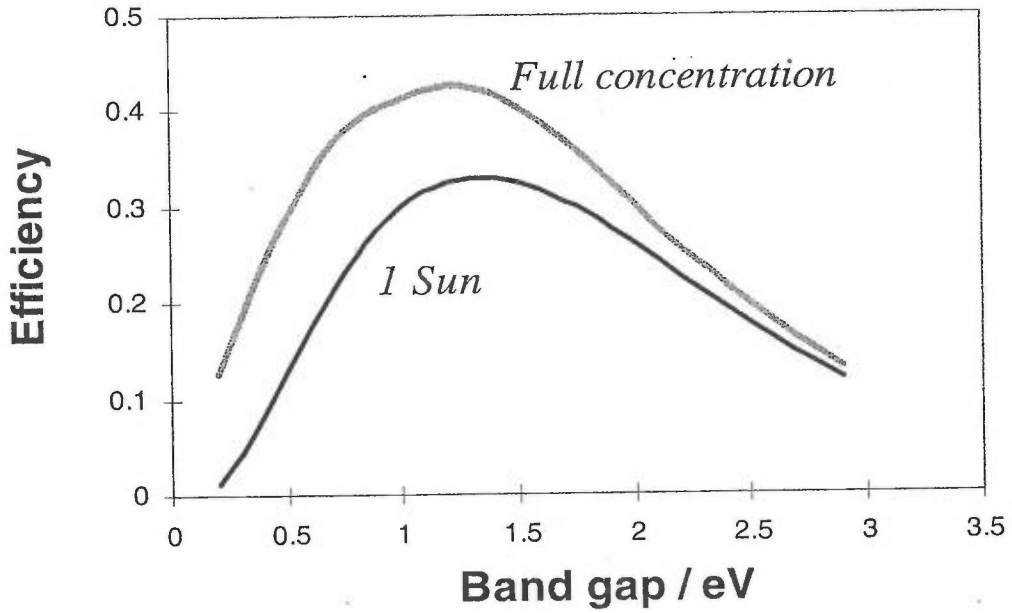


Fig. 10.4. Detailed balance limiting efficiency as a function of band gap, for a single band gap solar cell in a 5760K black body spectrum, at full concentration and at one-sun concentration levels. The maximum is similar to that obtained with the AM1.5 spectrum (Fig. 2.8).

(iii) Each electron is extracted with chemical potential energy μ such that $qV = \Delta\mu$. *I.e.*, the quasi Fermi level separation is constant throughout the device and equal to qV, where V is the potential difference at the terminals. This requires that carriers have infinite mobility.

(iv) The only loss process is spontaneous emission (also known as radiative recombination), which is required by detailed balance.

10.4. Multiple Band Gaps

The amount of work done per photon could clearly be increased if photons of different energies could be absorbed preferentially in cells of different band gap. A single band gap photoconverter functions most efficiently with monochromatic light which is tuned to the band gap. Then all photons are absorbed, and because the photon energy is close to the band gap, almost no electron kinetic energy is lost and the electrochemical potential extracted is close to the photon energy. If the solar spectrum could be split up and channelled into photoconverters of different band gaps, then more of the solar spectrum could be harnessed, each electron could be extracted with a chemical potential closer to the original photon, and a higher power could

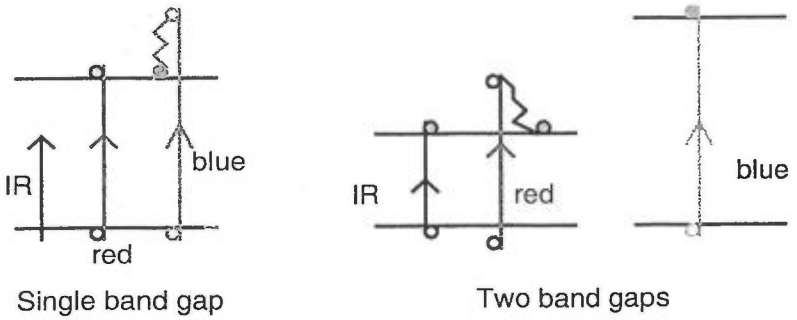


Fig. 10.5. Photon absorption in single gap and two gap system. The two gap system is capable of absorbing lower energy photons which are lost in the single band gap, and of extracting carriers generated by higher energy photons at an energy closer to the original photon energy.

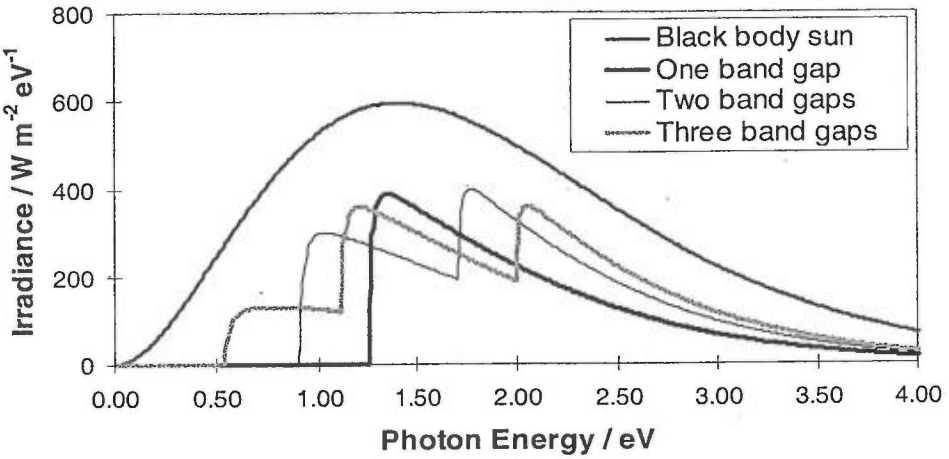


Fig. 10.6. Power available from optimised one, two and three band gap systems. (The different cells are assumed to be independently optimised.)

be extracted from the same spectrum. Figures 10.5 and 10.6 illustrate the point. Figure 10.7 illustrates one possible scheme for exploiting multiple band gaps, where sunlight is split up by means of dichroic mirrors and directed on to cells of different band gap.

10.5. Tandem Cells

10.5.1. *Principles of tandem cells*

In practice, efficient spectral splitting, as shown in Fig. 10.7, is hard to achieve. A more practical strategy is to stack different band gap junctions in optical series, and allow the wider band gap materials at the top to filter

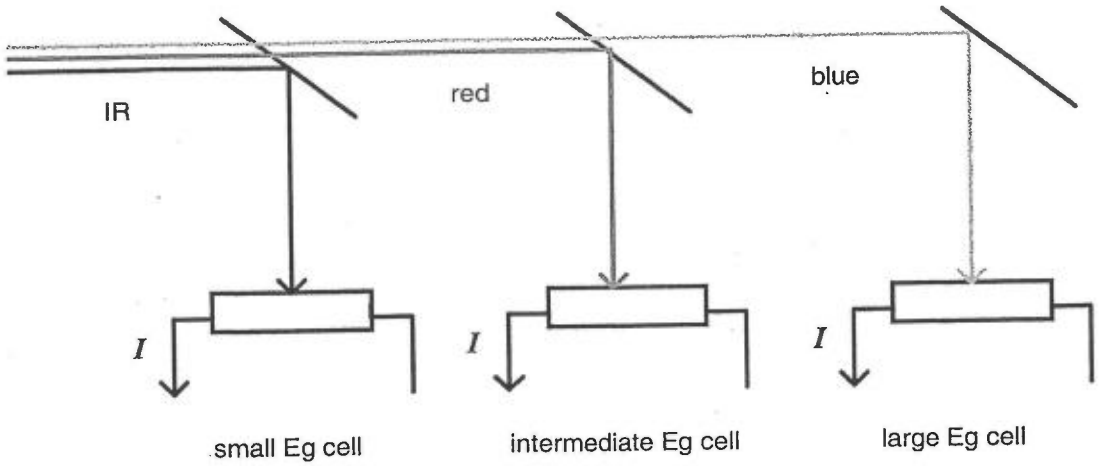


Fig. 10.7. Spectral splitting with dichroic mirrors allows different wavelength ranges to be absorbed preferentially in cells of different band gap.

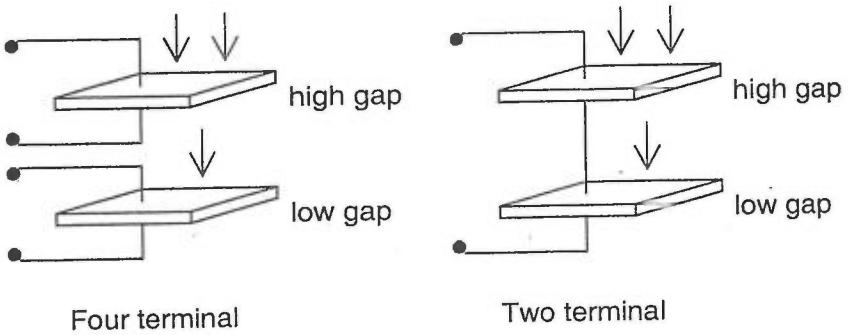


Fig. 10.8. Two and four terminal configurations for tandem cells. In either case, short wavelength light is preferentially absorbed in the top cell, and longer wavelength light in the bottom cell.

out most of the high energy photons, while less energetic photons pass through to smaller band gap materials below (Fig. 10.8). Greatest power is extracted if the output from the different junctions can be independently optimised. In the case of two band gaps, this is called a 'four terminal' tandem. The four terminal arrangement requires independent electrical contacts to top and bottom cell which is hard to achieve in practice. A more elegant arrangement is to connect the cells directly in series, so that a single current passes and voltages from the two cells are added. This 'two terminal' arrangement requires that currents from each cell be matched and constrains the performance so that the maximum output power is slightly less. Moreover, since current matching cannot be satisfied under all illumination conditions, the design is subject to additional, practical losses.

10.5.2. *Analysis*

We now calculate the power available using two band gaps, E_{g1} and E_{g2}, in a four terminal and a two terminal configuration.

In the four terminal configuration, the maximum output power is the sum of the maximum output powers from the two independent junctions. We assume that the spectral splitting is perfect, so that all photons with $E > E_{g2}$ are absorbed in the top cell, and all with $E_{g1} < E < E_{g2}$ in the bottom cell, where E_{g1} and E_{g2} are the band gaps of the bottom and top cell, respectively. Then the output power under maximum concentration is

$$P_{max} = qV_{m1}\{N(E_{g1}, E_{g2}, T_s, 0) - N(E_{g1}, E_{g2}, T_a, qV_{m1})\} + qV_{m2}\{N(E_{g2}, \infty, T_s, 0) - N(E_{g2}, \infty, T_a, qV_{m2})\} \quad (10.22)$$

where V_{m2} is the voltage at which the top cell, and V_{m1} the voltage at which the bottom cell, produces maximum power. Provided that the two biases can be independently optimised, P_{max} is a function only of the two band gaps. Under full concentration, a maximum efficiency of over 55% is reached with band gaps of 0.75 and 1.65 eV, as shown in Fig. 10.9.

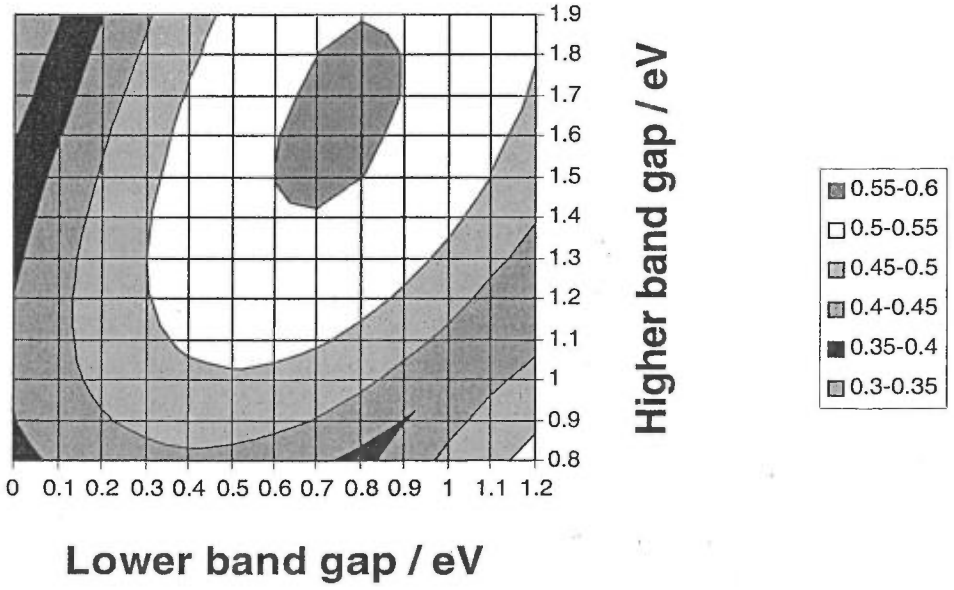


Fig. 10.9. Calculated efficiency of a four terminal tandem under a 5760K black body sun in full concentration, as a function of the band gaps. A maximum of around 56% occurs in the central shaded region.

In the two terminal configuration the maximum power is given by the maximum of

$$P_{\max} = q(V_1 + V_2)\{N(E_{g1}, E_{g2}, T_s, 0) - N(E_{g1}, E_{g2}, T_a, qV_1)\} \quad (10.23)$$

where V_1 and V_2 no longer represent the optimum voltages for the individual cells, but they satisfy the constraint of equal current

$$\{N(E_{g1}, E_{g2}, T_s, 0) - N(E_{g1}, E_{g2}, T_a, qV_1)\} = \{N(E_{g2}, \infty, T_s, 0) - N(E_{g2}, \infty, T_a, qV_2)\}, \quad (10.24)$$

as well as the condition that $P_{\max}$ is stationary. The limiting efficiency in this case is slightly smaller, and more sensitive to the band gaps.

For either configuration, increasing the number of band gaps increases the limiting efficiency. In the limit of an infinite number of band gaps, the smallest band gap falls to zero and the limiting efficiency under one sun reaches 69%. Under full concentration, the limit approaches the thermodynamic limits of 86%.

10.5.3. *Practical tandem systems*

Tandem cells have been developed in a variety of materials combinations. Two terminal designs have been more widely studied than four terminal designs, because of the technologically appealing possibility of integrating different junctions in a single multilayer device by using 'tunnel junctions' to connect different p–n junctions. The tunnel junction is a heavily doped n–p junction which is generally assumed to introduce an Ohmic contact between the p terminal of one cell and the n terminal of the next. III–V materials are preferred on account of the high absorption coefficient and the possibility of tuning the band gap by compositional variation of ternary and higher alloys. In practice, the well understood binary semiconductor gallium arsenide (GaAs), is chosen for one of the cells in most cases. Although its band gap at $E_g = 1.42$ eV is different from the ideal for the top or bottom cell in a two junction tandem, the good material quality and carrier transport properties lead to better performance than poorer quality ternary materials with more suitable band gaps. Another important consideration is the compatibility of the different semiconductor materials in terms of their lattice constant, and thermal expansivity. Mismatched lattice constants lead to defective interfaces between the cells and enhance

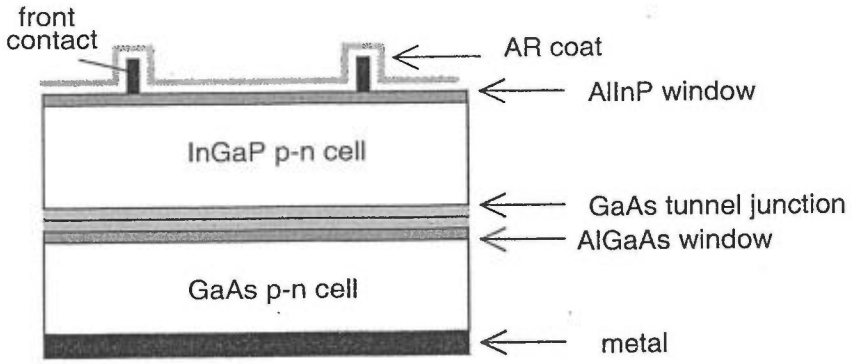


Fig. 10.10. Layer structure of an indium gallium phosphide/gallium arsenide two terminal tandem cell.

loss through recombination. Different coefficients of expansion lead to strain when the cell experiences changes in temperature.

The most efficient tandem cell produced to date is a 30.3% efficient indium gallium phosphide/gallium arsenide two terminal stack, developed by Japan Energy in 1996 (Fig. 10.10) [Green *et al.*, 2001]. Practical triple junction cells have been developed using the same two materials as the wider band gap components and the narrow band gap semiconductor, germanium, as the bottom cell. These have not yet surpassed the efficiency of the best double junction tandem. Tandem cells are expensive to manufacture, and are mainly being developed for use in space where efficiency is a premium.

10.6. Intermediate Band and Multiple Band Cells

10.6.1. *Principles of intermediate and multiple band cells*

The advantage of the tandem cell lies in the fact that the quasi Fermi level separation of the photogenerated electron and hole populations is closer, through the use of multiple band gaps, to the chemical potential of the exciting photon. A system which supports different quasi Fermi level separations delivers more work per incident photon. It would clearly be attractive if this could be achieved, not in several junctions made from different materials, but in a *single* junction using a *single* material. In a single band gap photoconverter, this is not possible because all of the conduction band states are coupled through phonon interactions and consequently all photogenerated electrons relax into a thermodynamic equilibrium with a single electrochemical potential, μ_n. All holes likewise relax into quasi thermal equilibrium with a single electrochemical potential, μ_p. Although it is

possible to design a junction of variable band gap, preferentially absorbing photons of different energy at different spatial locations, phonon coupling of the spatially continuous band of electron states brings all photogenerated electrons into quasi thermal equilibrium with a single μ, and likewise for holes. In, the limit of an ideal, radiatively dominated, solar cell, the difference in μ's is dominated by the smallest band gap in the device [Araujo, 1994].

A hypothetical solution is to use a material which contains more than two bands. Figure 10.11 illustrates the case of an intermediate band system, containing two separate conduction bands. Under illumination, electrons are excited from the valence band into both the upper and intermediate conduction bands, and from the intermediate into the upper conduction band. Provided that the intermediate and conduction bands are not thermally coupled, then the electron populations in the different bands each form a local quasi thermal equilibrium with its own quasi Fermi level, $\mu_{\mathrm{n,i}}$. The intermediate band may be introduced via impurities or quantum heterostructures which introduce electronic levels into the band gap, or it may be a result of the band structure.

Intermediate band solar cells have been proposed as hypothetical devices by several authors, [Kettemann, 1995; Luque, 1997; Green, 2000, 2001]. The idea of exploiting radiative transitions between intermediate levels is also a central concept of practical 'quantum well' solar cells [Barnham, 1991, 1997].

10.6.2. *Conditions*

We first explain the conditions for such a device to work, and then calculate the limiting efficiency.

(i) A condition for carriers to achieve independent quasi thermal equilibrium in a band is that collisions or scattering events *within* the band should be much more frequent than events *between* bands (discussed in Sec. 3.6). *This requires that there be a gap in the band structure which is large compared to the maximum phonon energy.* Otherwise, for a density of states which is continuous in energy, carriers can always be scattered into lower energy states by means of collisions with phonons.

(ii) *The feature which gives rise to the intermediate level should be periodic in space.* Periodicity imparts definite momentum to carrier states permitting spatial delocalisation, which aids transport. An imperfect

array of isoenergetic impurities will lead to carrier localisation and poor transport. Moreover, the arrangement of levels into a band helps to suppress phonon scattering. For an intermediate band which lies within k_BT of the conduction or valence band, inter band scattering events are energetically allowed but are restricted by the requirement for momentum conservation. The restriction is lifted for an *isolated* impurity level because all momentum states are represented in the stationary wavefunction of the defect. Even in the case of a deep level which lies beyond k_BT from the nearest band edge, it is advantageous to form a band. Although single phonon scattering events to the ground state of the impurity are forbidden, in certain conditions *multiple* phonon scattering events may be allowed, whereby an electron is trapped by the impurity and then relaxes to the ground state by a series of phonon emissions, as the electron environment is successively altered by the presence of the electron. This successive relaxation and distortion would be symmetry forbidden in a periodic structure, and so multiple phonon emissions would not provide a route to the trapping of an electron by a band of deep levels.

(iii) For the intermediate band to be thermally isolated from the conduction band, it is necessary that electrons are extracted from only one of the bands. A selective contact should be made to the conduction band and not to the intermediate band. Otherwise the electron populations would be brought into thermal equilibrium through the contact. With this satisfied, the intermediate band is coupled to the valence and conduction bands only through optical transitions.

Analysis of the intermediate band cell

Let's assume that all the conditions for a system supporting separate quasi Fermi levels are satisfied, and calculate the limiting efficiency for the three band cell illustrated in Fig. 10.11. We will assume planar geometry and full concentration. There are three types of photon absorption event: photons of $E > E_{g2}$ which promote an electron from the valence band to **C1**; photons of $E > E_g$ which promote an electron from the valence band to **C2** and those of $E > E_{g1}$ which are absorbed to promote an electron from band **C1** to **C2**. Note that because **C1** and **C2** are both conduction bands and normally empty, the threshold for (strong) absorption is from the *bottom* of **C1** to the *bottom* of **C2**, and therefore $E_{g2} = E_g - E_{g1}$.) For the maximum amount of work per photon, a photon should only be absorbed across the

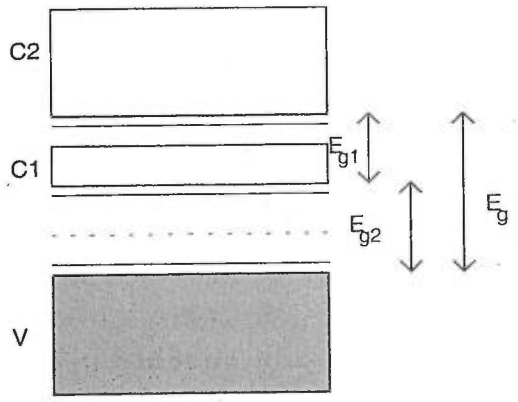


Fig. 10.11. Energy band diagram of an intermediate band material with a single valence band and two conduction bands, C1 and C2.

highest band gap which it is capable of crossing. This is the case if the absorption coefficient for the transition **V–C1** is much greater than for **C1–C2**, and that for **V–C2** much greater than for **V–C1**. Then each of the three types of absorption event occurs over a limited energy range: E_{g1} to E_{g2}; E_{g2} to E_g; and E_g to ∞. Assuming, as before, unit absorbance in each energy range, the output current density becomes

$$\begin{aligned} J(V) = q\{&N(E_g, \infty, T_s, 0) - N(E_g, \infty, T_a, \mu_{c2} - \mu_v)\} \\ &+ q\{N(E_{g2}, E_g, T_s, 0) - N(E_{g2}, E_g, T_a, \mu_{c1} - \mu_v)\} \end{aligned} \quad (10.25)$$

where μ_v, μ_{c1} and μ_{c2} are the quasi Fermi levels of the valence, intermediate and conduction bands, respectively. In the steady state the net current from **V** to **C1** must be matched by the current from **C1** to **C2**, which yields the constraint,

$$\begin{aligned} &q\{N(E_{g1}, E_{g2}, T_s, 0) - N(E_{g1}, E_{g2}, T_a, \mu_{c2} - \mu_{c1})\} \\ &\quad = q\{N(E_{g2}, E_g, T_s, 0) - N(E_{g2}, E_g, T_a, \mu_{c1} - \mu_v)\} \end{aligned} \quad (10.26)$$

and the quasi Fermi levels must satisfy

$$(\mu_{c1} - \mu_v) + (\mu_{c2} - \mu_{c1}) = \mu_{c2} - \mu_v = qV\,. \quad (10.27)$$

For any combination of E_g and E_{g1}, there is some value of V which makes the power density, $V \times J(V)$, a maximum, subject to the above constraints. Optimum efficiency, the ratio of maximum output power density to input solar power density, is then a function only of the band gaps. For a 6000K

black body sun at full concentration, this efficiency has a maximum of 63.1% when $E_g = 1.93$ eV and $E_{g1} = 0.7$ eV [Luque, 1997].

This result can be compared to the two junction tandem cells where η_{max} is likewise a function of two band gaps but the limiting value is much smaller — around 55% for the two terminal configuration under maximum concentration. There are different ways to regard the comparison. One is that the multi band cell actually presents *three* and not two band gaps to the incoming photons, albeit with some limitations on the band gaps and the Fermi level separations, and should be compared with the three junction tandem cell. Another consideration is that with a series connected tandem cell, two photon absorption events are always required to deliver one electron to the external circuit, while with the intermediate band cell either two or one photons are needed, depending on photon energy, so the photocurrent quantum efficiency is higher. In any case, the intermediate band cell is based upon a technically more challenging materials situation, on account of the requirements for a thermally isolated intermediate band and selective contacts.

10.6.3. *Practical strategies*

A variety of practical routes to the intermediate band solar cell have been proposed. The simplest is the idea that intra band states, introduced by ambivalent impurities in the semiconductor, can take part in optical transitions, and absorb photons of lower energy than the band gap to accept an electron from the valence band (if empty) or to deliver an electron to the conduction band (if full), as shown in Fig. 10.12. Enhanced photocurrent at sub band gap energies has in fact been observed, in silicon doped with indium [Keevers, 1994]. However, it has not been shown, and it is unlikely, that such a system could support independent quasi Fermi levels, since the isolated impurities may act as centres for non radiative recombination via multiple phonon events.

Another proposal is to use materials which naturally possess multiple bands of narrow width. Examples are polar semiconductors, such as II–VI materials or semiconducting oxides, and molecular semiconductors, which have conduction band widths typically of some tenths of eV. Narrow bandwidth materials are also attractive for tandem devices, on account of their finite band widths, which enhance the photon selectivity in different materials. Finite band widths are offered by dye molecules, which can be implemented as sensitisers on non absorbing conducting substrates.

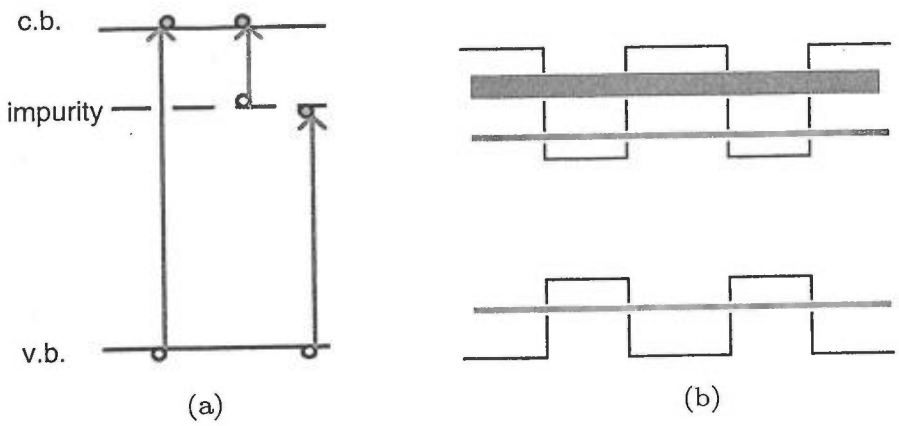


Fig. 10.12. (a) Impurity band semiconductor; (b) Quantum dot minibands.

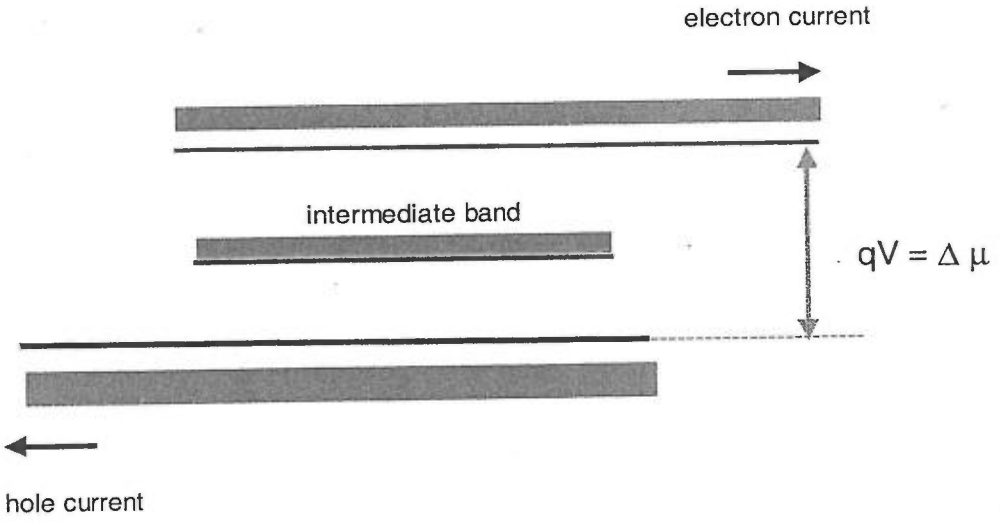


Fig. 10.13. Schematic energy-space diagram of intermediate band cell. Selective contacts are needed so that the intermediate band can be isolated from the external circuit, and the extracted voltage determined by the splitting of conduction and valence band Fermi levels.

A more elegant but ambitious proposal is to engineer materials with the desired band structure using low dimensional semiconductor heterostructures. Complete discretisation of the electronic energy levels can be achieved using quantum dots. In these nanometre-sized semiconductor particles the carriers are confined to three dimensions, resulting in an atomic like band structure of discrete energy levels. If the quantum dot is grown within a second, wider band gap semiconductor material, then the dot introduces a discrete level within the band gap of the host semiconductor. A *regular* array or *superlattice* of identical quantum dots leads to a set of minibands

within the band gap of the host semiconductor. Momentum conservation for band — band transitions reduces the probability of phonon scattering events between bands, making separate quasi Fermi level for electrons in the minibands possible. If the size and spacing of the dots can be controlled, the energy and width of the minibands can be tuned.

Box 10.1. Quantum semiconductor heterostructures for multiple band photovoltaics

Building semiconductor heterostructures on the nanometre scale leads to confinement of electrons and holes in one, two or three dimensions, and leads to densities of states which reflect the symmetry of the confining structure. For a quasi two-dimensional system, such as a 'quantum well', the density of states is quantised in one direction, so that the electron has kinetic energy

$$E = E_{\mathrm{n}} + \frac{\hbar^2(k_{\mathrm{x}}^2 + k_{\mathrm{y}}^2)}{2m^*}$$

where E_{n} is the energy due to confinement in the z direction, and the second term is the kinetic energy due to fully delocalised motion in the x and y directions. Following the approach presented in Sec. 3.3 for a three-dimensional geometry, we find that the density of states $g(E)$ is a staircase like function, with each step corresponding to a new confined level or sub-band E_{n}. The positions of these steps can be tuned by the choice of 'well' and 'barrier' materials and the width of the well, allowing engineering of the band structure. When carriers are confined in three directions, a quantum dot results, with discrete set of energies like a super atom. A regular array of such dots allows the discrete QD spectra to merge into bands. The dot dimensions determine the positions of the bands and the array spacing determines the band width, so that the band structure can be completely designed. Such systems are the basis of ideas for multiple band solar cells.

Quantum well structures are less suitable for intermediate band structures because the QW density of states is continuous in energy. A carrier in one sub-band can always be scattered into a lower band, (with different in-plane momentum) by phonon interaction. Nevertheless, QWs have already been exploited for multiple band photovoltaic devices by embedding a set of quantum wells within the depletion region of a p–i–n structure. The QWs introduce a new band gap so that photons of low energy may be absorbed, increasing the current, while still allowing high energy photons to

be efficiently absorbed in the host material. QW devices appear to benefit from a difference in quasi Fermi level separation between the wide band gap and low band gap regions, in contrast to the assumptions of detailed balance theory that $qV = \Delta\mu$ at all points. The reasons for this and the implications have yet to be clarified.

In addition, the loss of symmetry in electron and phonon density of states in QWs may make quasi 2D systems useful for hot carrier solar cells, discussed below.

10.7. Increasing the Work Per Photon using 'Hot' Carriers

Multiple band gap approaches are based on capturing photons of different energy in materials of different band gap, and extracting the photogenerated carriers with a chemical potential related to the band gap of the absorbing material used. An alternative approach is to increase the work done per photon by harnessing some of the excess kinetic energy of the photogenerated carriers before they relax. This could be done if electron phonon interactions could be slowed down so that the photogenerated carriers can be collected while still 'hot', or if the excess kinetic energy of hot carriers can be exploited to generate more carrier pairs by a process known as *impact ionisation*. The first results in an increased voltage and the second in an increased photocurrent. Both rely on similar physics and lead to identical limiting efficiencies, but we will treat them separately here, since the routes have been proposed on account of different physical observations.

10.7.1. *Principles of cooling and 'hot' carriers*

Let's first consider the sequence of events following photon absorption. At the instant of excitation, the photogenerated carriers have a kinetic energy distribution which reflects the distribution of photon energies from the sun and the absorption spectrum of the material. Elastic carrier–carrier scattering ('equilibriation') over the next few hundred femtoseconds brings each carrier population into a self equilibrium which can be described by a temperature, T_{H}, and a chemical potential, μ_{H}. T_{H} is greater than the ambient temperature and in certain cases can be equal to T_{s}.

Over the next few picoseconds, carriers in the hot plasmas lose kinetic energy through collisions with phonons until the temperature of the distribution is reduced to the lattice temperature, T_{a} ('cooling'). Then each carrier distribution is in quasi thermal equilibrium with the lattice, and

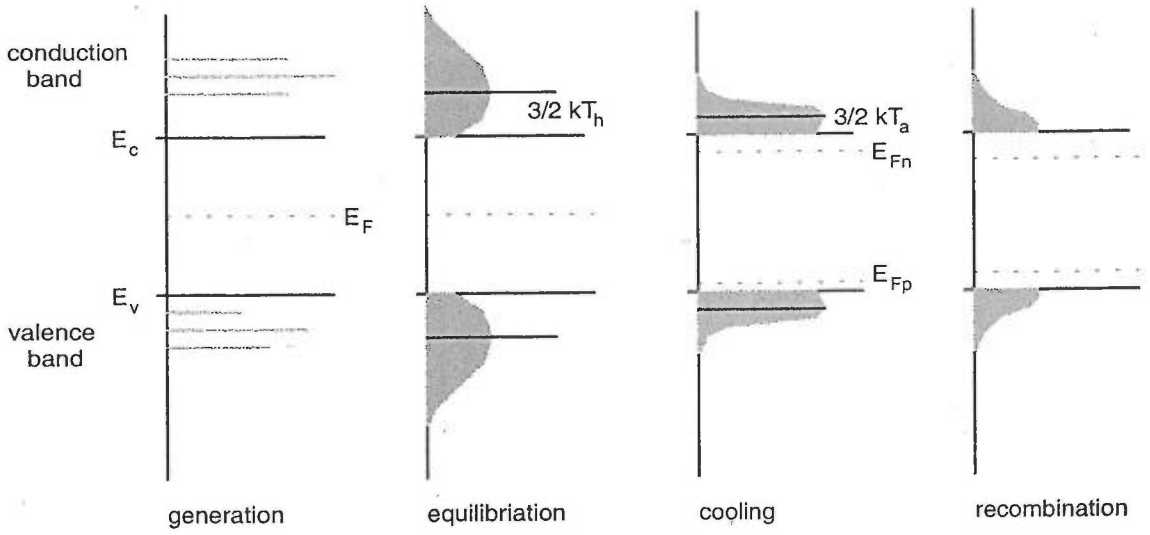


Fig. 10.14. Evolution of the carrier energy distributions following photogeneration.

the phonon absorption rate is balanced by phonon emission. However, the electron and hole populations are not yet in thermal equilibrium with each other. The populations of electrons in the conduction band, and of holes in the valence band, are increased over the equilibrium populations in the semiconductor in the dark, as though the electron Fermi level had moved towards the conduction band, and the hole Fermi level towards the valence band, as shown in Fig. 10.14. This Fermi level *splitting* is equal to the chemical potential of the radiation, and increases with light intensity. During cooling, the carriers give up kinetic energy as heat to the lattice, and the entropy of the system increases.

Over a longer time scale the carriers *recombine* by radiative and non radiative routes. In the current context we ignore nonradiative processes which are assumed to be avoidable. (Auger processes are considered in the next section.) Radiative recombination occurs over nanoseconds to microseconds, depending on the absorption coefficient of the material. The effect of recombination is to reduce the electron and hole populations and move the quasi Fermi levels back towards the equilibrium position. Under constant illumination, populations achieve a steady state where the photogeneration rate is equal to the sum of the recombination rate and the rate of carrier extraction (as electric current). This balance determines the electrochemical potential and the power output of the cell. For efficient energy conversion the rate of carrier extraction should be comparable with or faster

than the rate of recombination. In the conventional picture, the first two processes of equilibration and cooling are assumed to be complete for any carrier before it is extracted, and only recombination needs be considered in determining the power output of the cell.

The central concept of the hot carrier solar cell is that cooling *need not* be complete before carriers are extracted. If extraction can be speeded up, or cooling slowed down, then some of the carriers can be extracted while still hot and deliver some of their excess kinetic energy to the electrical output of the cell. These conditions could arise in systems of unusual electronic and phononic structure. (Cooling is influenced by the availability of both phonons of appropriate energy and momentum to absorb electronic kinetic energy and of lower energy states for the carriers to be scattered into.) A further requirement is that the contacts are designed to allow adiabatic cooling of hot carriers during extraction. Some proposed practical schemes are discussed below.

10.7.2. *Analysis of the hot carrier solar cell*

To calculate the maximum efficiency of the hot carrier solar cell we consider the limit where there is *no* carrier–lattice scattering. We consider a material of band gap E_g and make the usual assumptions for the limiting photovoltaic efficiency: that all photons of $E > E_g$ are absorbed; that each photon generates one electron-hole pair; and that radiative recombination is the only loss process. As usual, illumination is provided by a black body sun at temperature T_s and geometric factor F_s. (The formulation of this section is based on [Wuerfel, 1997] and [Ross, 1982].)

After photogeneration, each carrier population equilibrates through carrier–carrier scattering. At equilibrium, the Gibbs free energy is minimised and therefore the quantity

$$\sum_i \eta_i dn_i$$

is conserved through any scattering event, where η_i is the electrochemical potential of carrier i and dn_i the particle number. For the electron scattering event $e_1 + e_2 \rightarrow e_3 + e_4$, it follows that

$$\eta_{e1} + \eta_{e2} = \eta_{e3} + \eta_{e4} \tag{10.28}$$

since $dn_i = 1$ for each carrier. Since carrier scattering is elastic, the kinetic energies obey

$$E_{e1} + E_{e2} = E_{e3} + E_{e4} \tag{10.29}$$

These equations are satisfied for all carrier pairs when η_i depends on kinetic energy through

$$\eta_i = \eta_0 + \gamma E_i \tag{10.30}$$

where γ and η_0 are constants. If $\gamma = 0$ we have the familiar case that all carriers at equilibrium have the same electrochemical potential. In the absence of electron-lattice scattering, the solution $\gamma > 0$ is also valid [Wuerfel, 1997]. Assuming a similar relation for the holes (in principle the value of γ may be different for holes), we have for the chemical potential of the electron-hole pair

$$\Delta\mu = \eta_{e0} + \eta_{h0} + \gamma(E_e + E_h) \tag{10.31}$$

Hence the chemical potential of an electron-hole pair generated by a photon of energy E is given by

$$\Delta\mu = \mu_0 + \gamma E \tag{10.32}$$

where

$$\mu_0 = \eta_{e0} + \eta_{h0} - \gamma E_g \,. \tag{10.33}$$

The constants η_{e0}, η_{h0} depend upon the extraction conditions.

Now, by considering interaction with photons of energy E, the carriers will be distributed according to the function

$$f = \frac{1}{e^{(E-\Delta\mu)/k_B T_a} + 1} \tag{10.34}$$

where T_a is the (ambient) temperature and $\Delta\mu$ the energy dependent chemical potential, Eq. 10.32 above. By rewriting f as

$$f = \frac{1}{e^{E(1-\gamma)/k_B T_a - \mu_0/k_B T_a} + 1} \tag{10.35}$$

it is evident that the same distribution can be described by a *single* chemical potential, μ_H, if the temperature is defined as

$$T_H = \frac{T_a}{1-\gamma} \,. \tag{10.36}$$

Then

$$\Delta\mu = \mu_{\mathrm{H}}\frac{T_{\mathrm{a}}}{T_{\mathrm{H}}} + E\left(1 - \frac{T_{\mathrm{a}}}{T_{\mathrm{H}}}\right). \tag{10.37}$$

(The same relations follow by considering the entropy change per absorbed photon.) This means that the luminescence from the carrier distribution is exactly equivalent to that from a distribution at temperature T_{H} and chemical potential μ_{H}, and so it is meaningful to consider the carriers as being at a hot temperature T_{H}.

This enables us to calculate the current output. As usual in the radiative limit, J is given by the difference between the net absorbed and emitted photon flux

$$\begin{aligned} J(V) = q\{&Xf_{\mathrm{s}}N(E_{\mathrm{g}}, \infty, T_{\mathrm{s}}, 0) - N(E_{\mathrm{g}}, \infty, T_{\mathrm{H}}, \mu_{\mathrm{H}}) \\ &+ (1 - Xf_{\mathrm{s}})N(E_{\mathrm{g}}, \infty, T_{\mathrm{a}}, 0)\}. \end{aligned} \tag{10.38}$$

Note that in the case of full concentration, Xf_{s} becomes equal to 1 and the final term in Eq. 10.38 vanishes.

For the hot carrier device, the dependence of V on μ_{H} is no longer trivial. On extraction, the carriers must cool to the ambient temperature T_{a} and in so doing the chemical potential of the extracted carriers will increase. For maximum work, this cooling should be done at constant entropy. This is possible if electrons and holes are withdrawn through separate contacts of extremely narrow energy range, $\Delta E \ll k_{\mathrm{B}}T$, so that no kinetic energy is lost during cooling. At constant entropy Eq. 10.37 is valid, and the chemical potential of the extracted carriers is given by

$$\mu_{\mathrm{out}} = \mu_{\mathrm{H}}\frac{T_{\mathrm{a}}}{T_{\mathrm{H}}} + E_{\mathrm{out}}\left(1 - \frac{T_{\mathrm{a}}}{T_{\mathrm{H}}}\right) \tag{10.39}$$

where E_{out} is the energy separation of the electron- and hole-selective contacts. Thus the voltage output is $V = \mu_{\mathrm{out}}/q$ and the extracted power is

$$P(V) = \frac{\mu_{\mathrm{out}}}{q} \times J. \tag{10.40}$$

This description is valid so long as carrier extraction is slow compared to carrier–carrier scattering. It is assumed that the equilibrium distribution of carriers is unaffected by the extraction of carrier pairs. Fast carrier scattering replenishes the population of carrier pairs with energy E_{out} as the carriers are extracted.

One more condition is required to calculate the power output. This comes from the conservation of energy within the hot device. Since energy is introduced only through photon absorption and is lost only through power extraction and hot carrier emission, we can write

$$JE_{\text{out}} = q\{Xf_{\text{s}}L(E_{\text{g}},\infty,T_{\text{s}},0) - L(E_{\text{g}},\infty,T_{\text{H}},\mu_{\text{H}}) + (1-Xf_{\text{s}})L(E_{\text{g}},\infty,T_{\text{a}},0)\}\,. \tag{10.41}$$

We can now calculate the current–voltage characteristics of the system. The performance of the hot carrier systems depends upon *three* externally determined, or design, parameters: E_{out} as well as the usual parameters of E_{g} and V. The value of E_{out} importantly influences the performance of the device as a solar cell.

For a given E_{g} and E_{out} the device operates as a solar cell over a range of V from 0 (short circuit) to V_{oc}, where $J = 0$. Varying the output voltage controls the temperature of the carriers. At short circuit where all photogenerated carriers are collected, $\mu_{\text{out}} = 0$, and $T = T_{\text{a}}$ in all cases; *i.e.*, there is no hot carrier effect. As V is increased towards V_{oc}, J decreases due to increased radiative recombination, and T_{H} increases towards a maximum value which is close to T_{s} at open circuit. During this cycle, the behaviour of μ_{H} is rather counterintuitive: μ_{H} becomes more negative as the output voltage is increased, unlike the usual case where $\Delta\mu$ increases exactly as V.

It may be helpful to think of this behaviour in terms of energy balance, since the photogenerated carriers introduce excess kinetic energy which is conserved during elastic scattering and not lost by cooling. At short circuit, all carriers introduced by the light are extracted, and so contribute no excess kinetic energy to the carrier distribution in the solar cell. Therefore $T_{\text{H}} = T_{\text{a}}$. As output voltage is increased, the carrier density in the cell increases, the kinetic energy per carrier increases on account of the more energetic photogenerated carriers, and so the temperature of the carrier distribution increases. At open circuit where no carriers are extracted, the photogenerated carriers contribute the maximum kinetic energy and T_{H} has its maximum value. In the limit of full concentration and small band gap, T_{H} at open circuit is equal to T_{s}.

Once E_{g} and E_{out} are determined, the current–voltage characteristic for the hot carrier device can be found by varying V between 0 and V_{oc}. The maximum power point will correspond to certain values of T_{H} and μ_{H}. A family of J–V curves are illustrated in Fig. 10.15. Notice that the curve is

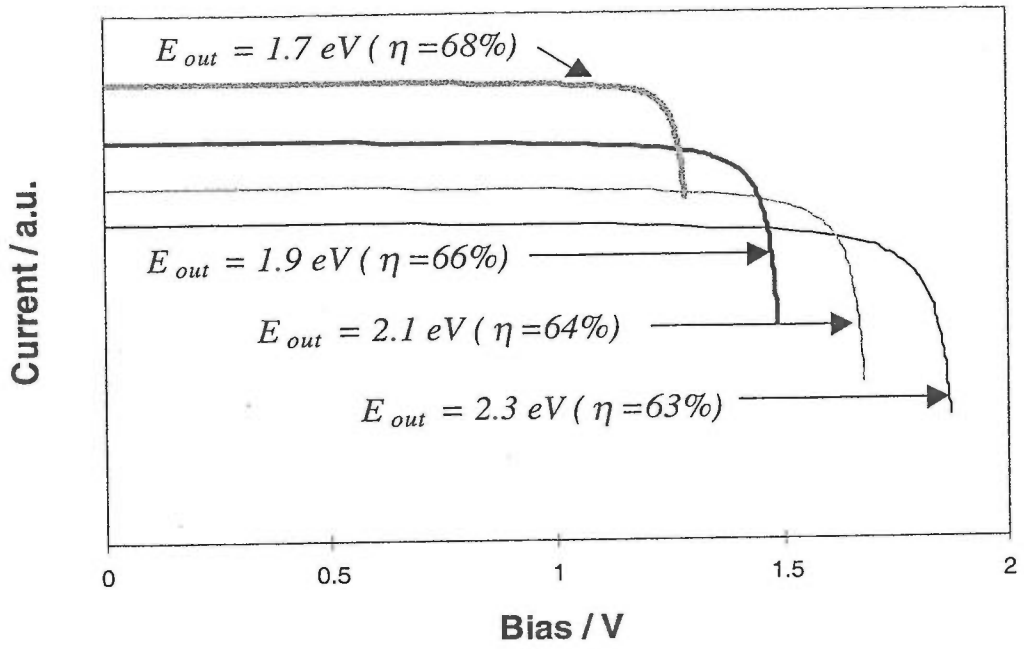


Fig. 10.15. Calculated current–voltage characteristics for a hot carrier solar cell in an unconcentrated black body sun. Curves for different values of the extraction energy, E_{out}, are shown, and the maximum efficiencies for each curve are given. Notice how the knee of the curve is almost square in the highest efficiency case.

much squarer than those for the radiative limit, consistent with a higher efficiency. In the radiative limit, the curvature of the J–V curve is due to the radiative current rising like $\exp(V/k_{\mathrm{B}}T)$ at constant T. For the hot carrier cell, although losses are still radiative, the temperature is increasing as V increases, and the shape no longer described by a single T.

Efficiency can be calculated as a function of the design parameters, E_{g} and E_{out}. For comparison with the radiative limit, it is convenient to make E_{out} a variable parameter and find the values of V and E_{out} which lead to the highest output power for each E_{g}. Those values will correspond to particular values of T_{H} and μ_{H}. (Indeed, the calculation simplifies if we consider T_{H} and μ_{H} as optimisable parameters rather than the design and operation variables of E_{out} and V.) Figure 10.16 shows the maximum efficiency as a function of E_{g}, together with the corresponding values of E_{out}, T_{H}, and μ_{H}. For full concentration, η has a maximum of $\sim$ 85% at a temperature of 3600K and a band gap of a few tenths of eV. For an unconcentrated black body sun η has a maximum of $\sim$ 65%. The temperature for maximum power is unaffected by concentration, in contrast to the case of impact ionisation, discussed below.

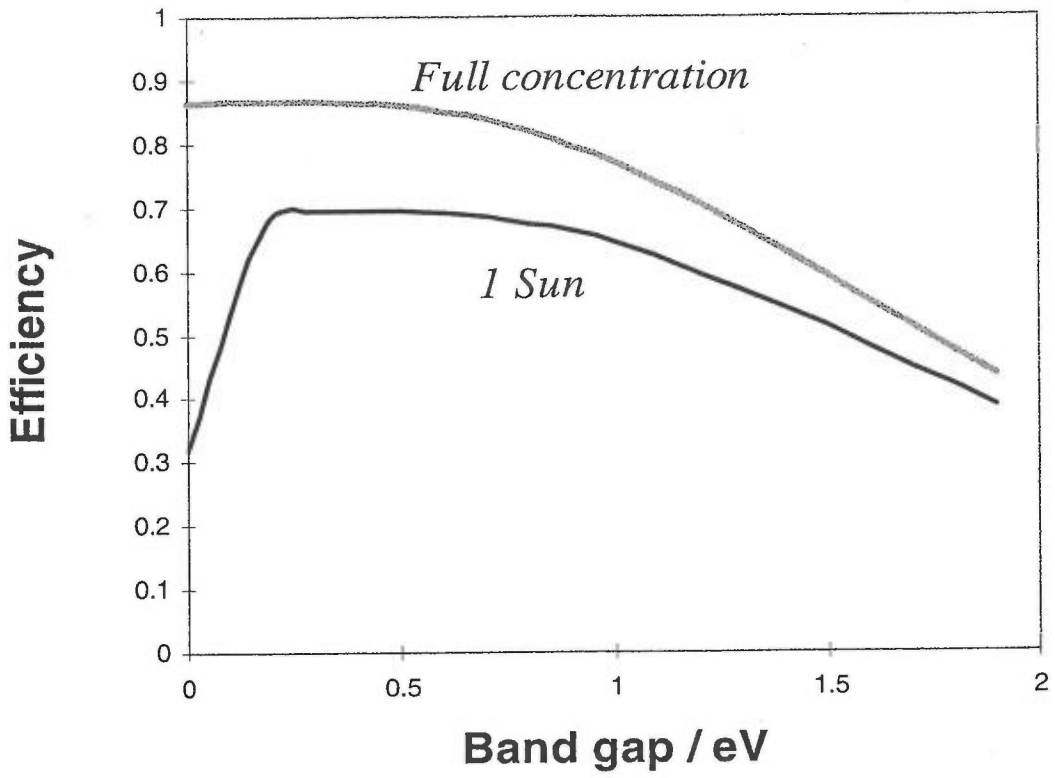


Fig. 10.16. Calculated efficiency as a function of band gap for hot carrier solar cell under full concentration and 1-sun illumination by a 5760K black body sun.

10.7.3. *Practical strategies*

There are two materials requirements for hot carrier devices: a photoactive material where cooling is slower than transport to the contacts; and contact materials which permit selective electron or hole extraction through a narrow energy band.

In typical bulk semiconductor devices, cooling occurs in less than 10 ps, while carrier extraction may take nanoseconds or longer. The rate of cooling depends upon the availability of phonons of appropriate energy and momentum to scatter the excited carriers into lower energy states. Phonons fall into two groups: acoustic phonons, which typically have energy of a few meV at room temperature and longitudinal optical (LO) phonons which have a higher energy — typically 30–40 meV in III–V semiconductors at room temperature. LO phonons are most likely to be involved in cooling hot carriers. The phonon momentum distribution in a bulk material is approximately isotropic but related to the crystal band structure. Since the density of electron states in a bulk semiconductor is continuous, and isotropic, a

carrier in an excited state has a high probability of being scattered into an unoccupied, lower energy state by a suitable phonon. This can be prevented if (i) there is a shortage of phonons or (ii) there are no available electron states within a phonon energy. The first situation arises under very intense illumination. At high enough injection levels, there may not be enough phonons for all the photogenerated carriers to relax promptly, and in the steady state the energy distribution of the carriers is hot. In addition, the strong interactions of higher energy phonons with the hot carrier distribution disturbs the equilibrium of the phonon distribution, so that the phonons themselves may become hot, and less effective at cooling. However, this requires very high light intensities, orders of magnitude more intense than the sun.

Slowed cooling may be observed more easily in semiconductor structures of reduced dimension. In quantum wells, carriers are confined into two dimensional slabs and the electronic structure is quantised into sub-bands. Although the density of states is still continuous, it is highly anisotropic and an excited carrier may only be scattered into a lower energy level by a phonon of the correct momentum (for intra sub-band transitions, the phonon wavevector must lie in the plane of the QW). Greatly increased cooling times and hot electron spectra have been observed in QWs of GaAs in $Al_xGa_{1-x}As$, with cooling times of up to 1000 ps compared to less than 10 ps for bulk GaAs, at light intensities of 10^4 suns. (The carrier temperature can be observed experimentally from the energy distribution of the photoluminescence spectrum.) For a review, see Nozik [Nozik, 2000].

Quantum *dots* (QDs) are potentially more useful. Carriers in a QD are confined in all dimensions and the electronic structure is quantised into discrete levels. If the spacing of these levels is larger than the energy of the most energetic phonon, then scattering of a carrier from an excited to a lower energy state through a single phonon interaction is impossible. Carriers could only relax via multi phonon interactions which are expected to be relatively very slow. This is the phonon bottleneck effect, mentioned above in the context of intermediate band solar cells. Slowed cooling has been observed in QD structures, again under high injection conditions, but not to a greater extent than in QWs, and the predicted phonon bottleneck is not seen. In practice, alternative relaxation mechanisms, such as Auger recombination and trap assisted relaxation, may operate.

An alternative strategy is to make use of structures where the charge separation is extremely fast. Interfacial charge transfer at semiconductor electrolyte junctions and in heterogeneous molecular systems can be very

fast: for instance, in dye sensitised solar cells, electrons are injected from a molecular sensitiser into a bulk semiconductor in less than a picosecond [Hagfeldt, 2000]. The injected carriers are inevitably hot, and this could be exploited if they could be collected before cooling. This would require very thin devices, to reduce the transit time of the injected carrier to the contacts. The challenge will be to combine high optical absorption with short transit times.

The second requirement is selective contacts with a narrow energy band. Usual contacting materials, such as metals or heavily doped semiconductors, are not suitable but low dimensional semiconductor structures may again be useful. One possibility is to a superlattice structure, where the density of states is quantised into bands, which may be made narrow and well separated by choosing the materials and periodicity carefully.

In the analysis of *practical* hot carrier systems, cooling by phonon interactions cannot be ignored. The drift-diffusion and carrier continuity equations which are used to analyse normal semiconductor devices are based upon *thermalised* carrier populations and do not apply. To allow for spatial variations in the carrier temperature, a further continuity equation must be introduced. This is the energy balance equation. In the steady state the electron kinetic energy flux S_n obeys

$$\nabla S_n = J_n F + W_{relax} + W_{gen} + W_{rec} \tag{10.42}$$

where the first term represents the gain in kinetic energy through electron acceleration by the electric field, W_{relax} is the loss of kinetic energy by relaxation, W_{gen} the loss by electron recombination and W_{gen} the generation of kinetic energy by optical generation. A similar equation applies for hot holes. These should be solved along with the current and continuity equations and Poisson's equation, leading to profiles for the carrier kinetic energy (which is effectively the carrier temperature, since kinetic energy $= 3/2k_BT$) as well as for carrier density, current density and electric field. Note that in defining the current, the contribution due to a gradient in carrier temperature (the Seebeck effect) must be included. For further details see [Lunsdtrom, 1990; Hanna, 1997.]

10.8. Impact Ionisation Solar Cells

The final route, to be discussed here, to increasing the work done per photon is impact ionisation. This is a scheme where a relaxation process is

introduced which competes with cooling and leads to further carrier pair generation. In the following analysis we will see that the impact ionisation solar cell is closely related to the hot carrier solar cell.

Impact ionisation, or Auger generation, is the reverse of Auger recombination. In Chapter 4 we saw that Auger recombination is a three body process, where an electron collides with a second electron, or with an impurity, recombines with an available hole and gives up its electrochemical potential energy as kinetic energy to the second electron. In the reverse process, an energetic electron collides with the lattice and gives up its kinetic energy to excite a further electron across the band gap. In the context of a photovoltaic device, this means that the quantum efficiency for light with $E > 2E_g$ can be greater than one. These high energy photons are capable of *multiple pair generation*. This effect has been observed, and attributed to Auger generation, in germanium photodiodes at photon energies $E > 2.5$ eV ($E_g = 0.7$ eV) and in silicon diodes for $E > 3.3$ eV ($E_g = 1.1$ eV) [Werner, 1995].

In Auger processes energy and momentum must be conserved. Suppose an electron $\mathbf{e_1}$ with kinetic energy E_{e1} and momentum $\mathbf{k}_{e1}$ relative to the symmetry point k_0 causes generation of a pair $\mathbf{e}_2$, $\mathbf{h}_2$. Then

$$\mathbf{k}_{e1} = \mathbf{k}'_{e1} + \mathbf{k}_{e2} - \mathbf{k}_{h2} \tag{10.43}$$

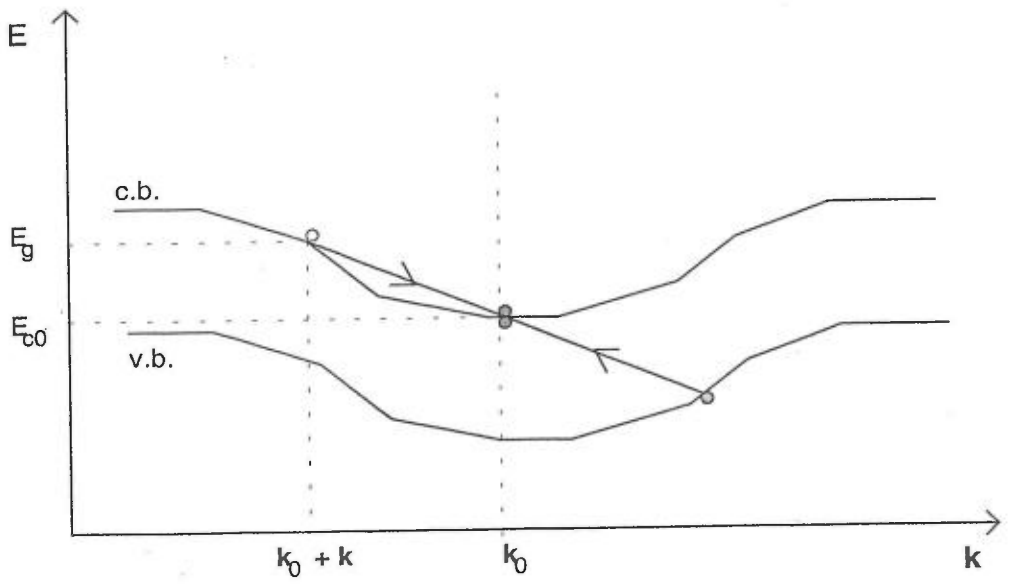


Fig. 10.17. Energy-momentum band structure diagram showing the conditions for Auger generation.

and

$$E_{e1} = E'_{e1} + E_{e2} - E_{h2} \,. \tag{10.44}$$

Auger generation events depend on availability of electron and hole states at correct $\mathbf{k}$ and E and therefore the band structure is most important in determining the probability of generation events. The conditions on the band structure for effective Auger generation have been discussed by Werner *et al.* [Werner, 1994] and are summarised at the end of this section.

10.8.1. *Analysis of impact ionisation solar cell*

In the following paragraphs we develop the theory for the limiting efficiency of a cell admitting multiple pair generation, as developed by Wuerfel [Wuerfel, 1997]. The different analysis proposed by the inventors, Werner, Brendel, Kodolinski and Queisser, has subsequently been shown to be flawed, in a thermodynamic respect. This original analysis is appended at the end of this section.

To calculate the maximum efficiency for an impact ionisation solar cell, we will assume that all the conditions for multiple pair generation are satisfied. The reverse process of impact ionisation, Auger recombination, must be included to satisfy detailed balance. We will assume that carrier relaxation by phonon interactions is sufficiently slow that it can be ignored. Therefore, the photogenerated carriers are allowed to equilibriate, but not to cool, and they may recombine only by radiative or Auger recombination. We know from Eq. 10.31 that the electrochemical potential of an equilibriated carrier pair depends upon the carriers' energy as

$$\Delta\mu = \eta_{e0} + \eta_{h0} + \gamma(E_e + E_h) \,. \tag{10.45}$$

When Auger recombination is allowed, an equilibrium will be established between Auger generation and recombination, where a number dn_1 of electron-hole pairs of energy E_{eh1} and chemical potential $\Delta\mu_{eh1}$ are transformed into a number dn_2 pairs of energy E_{eh2} and chemical potential $\Delta\mu_{eh2}$. The conditions of energy conservation and minimisation of the free energy $\sum_i \mu_i dn_i$ mean that the electrochemical potential must be *proportional* to the carrier energy, hence

$$\Delta\mu = \gamma(E_e + E_h) \,.$$

Considering the luminescence from such a distribution we find, as before (Eq. 10.36), that the distribution can be described by a raised temperature

$$T_{\mathrm{H}} = \frac{T}{1-\gamma}$$

and a chemical potential $\mu_{\mathrm{H}} = 0$. Zero chemical potential is expected for a thermal distribution where particle number is not conserved.

In calculating the power conversion efficiency we proceed as in the hot carrier case, except that the current can no longer be related to the difference of solar generation and radiative recombination (because additional current is generated through impact ionisation). Instead, the output current is calculated from the energy balance

$$\begin{aligned} JE_{\mathrm{out}} = q\{&Xf_{\mathrm{s}}L(E_{\mathrm{g}},\infty,T_{\mathrm{s}},0) - L(E_{\mathrm{g}},\infty,T_{\mathrm{H}},0) \\ &+ (1-Xf_{\mathrm{s}})L(E_{\mathrm{g}},\infty,T_{\mathrm{a}},0)\} \end{aligned} \tag{10.46}$$

and the power density is given by

$$\begin{aligned} P = JV = q\left(1-\frac{T_{\mathrm{a}}}{T_{\mathrm{e}}}\right)\{&Xf_{\mathrm{s}}L(E_{\mathrm{g}},\infty,T_{\mathrm{s}},0) - L(E_{\mathrm{g}},\infty,T_{\mathrm{H}},0) \\ &+ (1-Xf_{\mathrm{s}})L(E_{\mathrm{g}},\infty,T_{\mathrm{a}},0)\} \end{aligned} \tag{10.47}$$

where the output voltage, V, is related to the extraction energy E_{out} through

$$V = \frac{1}{q}\mu_{\mathrm{out}} = E_{\mathrm{out}}\left(1-\frac{T_{\mathrm{a}}}{T_{\mathrm{H}}}\right). \tag{10.48}$$

The energy balance, expressed by Eq. 10.47, is equivalent to that used in the hot carrier solar cell. It is still valid because Auger processes conserve energy, although they do not conserve particle number.

Although three design parameters (E_{g}, E_{out} and V) appear to be needed to find the J–V characteristic of an impact ionisation cell, only two of these are independent. Notice that P depends only on two parameters, E_{g} and T_{H}, the second of which can be controlled by varying either V or E_{out}. This means that for the impact ionisation solar cell, the extraction energy E_{out} makes no difference to the maximum efficiency. As with the hot carrier solar cell, T_{H} varies from its minimum value of T_{a} at short circuit, to its maximum value at open circuit as V is increased. The value of T_{H} at the maximum power point depends upon the level of illumination.

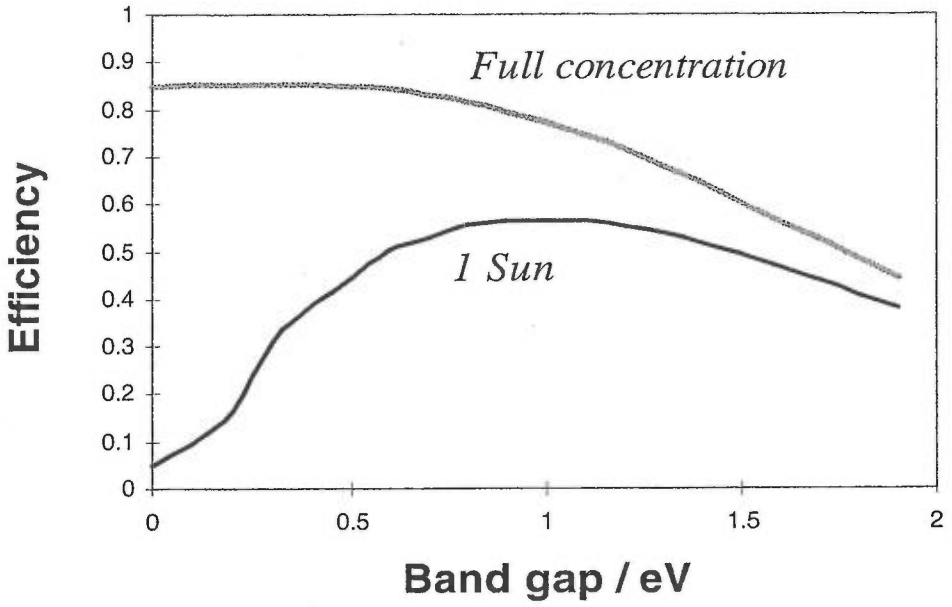


Fig. 10.18. Efficiency as a function of band gap for impact ionisation solar cell under 1-sun and full concentration.

Power conversion efficiency is plotted as a function of band gap in Fig. 10.18. For an unconcentrated black body sun, efficiency has a maximum of 55% at a band gap of around 1 eV. Under full concentration, the efficiency reaches a maximum of 85% at $T_{\mathrm{H}} = 2460$ and a small, vanishing band gap. The behaviour under full concentration is identical to that of a fully optimised hot carrier cell. Once again, this maximum performance under full concentration corresponds to the thermodynamic limit.

It is interesting to note that, except in the case of full concentration, impact ionisation leads to a lower limiting efficiency than simply preventing cooling, as in the case of the hot carrier cell. This suggests that a band structure conducive to impact ionisation may not be a desirable property of a high efficicency photovoltaic material. However it has the advantage of easier design conditions (E_{out} does not need to be optimised to obtain the maximum performance). Moreover, in realistic systems, impact ionisation needs to be considered in competition with relaxation by phonon interactions. In such conditions, the hot carrier device would not necessarily outperform the impact ionisation device, even at less than full concentration.

An alternative analysis of the impact ionisation solar cell was proposed by the inventors [Werner, 1994, 1995]. In their theory, it is assumed that the electron and hole populations are in thermal equilibrium with the lattice, at temperature T_{a}. This assumption is incompatible with suppressed carrier

cooling. This approach also arrives at a photon energy dependent chemical potential, but for a different reason, and leads to identical limiting efficiency of 85% under full concentration as $E_g \rightarrow 0$. The approach differs from the above in the models of recombination: above we consider that a *single* hot electron-hole pair with chemical potential $\mu = \mu E$ recombines to produce a photon of energy E. In the WKBQ model, *multiple* carrier pairs at $T = T_a$ (with $\Delta\mu = qV$) recombine to give a single photon. This multiple carrier recombination process is physically very unlikely.

Practical strategies

Since suppressed phonon interactions are a condition for impact ionisation solar cells, the practical strategies discussed above in the section on hot carrier devices apply, as does the requirement for contact materials with very narrow energy ranges, to allow isoentropic cooling.

In addition, the electronic band structure should favour impact ionisation. The conditions for this are discussed by Werner *et al.* [Werner, 1994]. They include the requirement for direct points in the band structure with band gap greater than twice the fundamental band gap; the requirement that valence band and conduction band be parallel over wide ranges of $\mathbf{k}$ (this allows carrier pairs to be generated so that only one carrier possesses the excess energy); and the requirement for strong light absorption at points which permit Auger generation. The first two of these imply that the material should have an indirect band gap, as do silicon and germanium. Finding the atomic or crystal configuration which generates the ideal band structure for multiple pair generation is an inverse problem, (inverse of solving the Schrödinger equation) and so far has not been attempted.

10.9. Summary

The limiting efficiency for a single junction solar cell, from considerations of the detailed balance between absorbed and emitted radiation, is around 31% in unconcentrated sunlight. In principle, this limit can be increased by relaxing one or more of the assumptions on which this calculation is based. In this chapter we have presented a number of routes to increased efficiency. These are based on: (1) the preferential absorption of photons of different energy in materials of different band gap, which is the basis of tandem and multiple band solar cells; (2) the exploitation of radiative transitions between the principal valence and conduction bands and an intermediate

band; (3) the rapid collection of photogenerated carriers to make use of their kinetic energy before they reach thermal equilibrium with the environment, and (4) the generation of multiple carrier pairs by absorbed photons with energy greater than the band gap.

All of these strategies place exceptional demands on the materials used. Ideally, high quality materials are needed so that non-radiative recombination can be eliminated. Materials of extraordinary electronic and optical properties are needed to engineer the required band structure. Nanostructured materials, including low dimensional semiconductor structures and quantum dots, are cited as possible ways of realising the hypothetical structures.

References

G.L. Araujo and A. Marti, "Absolute limiting efficiencies for photovoltaic energy-conversion", *Sol. Energy Mater. Solar Cells* **33**, 213 (1994).

K.W.J. Barnham, B. Braun, J. Nelson and M. Paxman, "Short-circuit current and energy efficiency enhancement in a low-dimensional structure photovoltaic device", *Appl. Phys. Letts.* **59**, 135–137 (1991).

K. Barnham *et al.*, "Quantum well solar cells", *Appl. Surf. Sci.* **114**, 722–733 (1997)

A. de Vos, *Endoreversible Thermodynamics of Solar Energy Conversion* (Oxford University Press, 1990).

A. Hagfeldt and M. Graetzel, "Molecular photovoltaics", *Acc. Chem. Res.* **33**, 269–277 (2000), and references therein.

M.A. Green, "Prospects for photovoltaic efficiency enhancement using low-dimensional structures", *Nanotechnology* **11**, 401–405 (2000).

M.A. Green, "Multiple band and impurity photovoltaic solar cells: General theory and comparison to tandem cells", *Progr. Photovoltaics* **9**, 137–144 (2001).

M.A. Green, K. Emery, D.L. King, S. Igari and W. Warta, "Solar cell efficiency tables", *Progr. Photovoltaics* **9**, 287 (2001).

M.C. Hanna, Z. Lu and A.J. Nozik, "Hot carrier solar cells", *Proc. 1st NREL Conference on Future Generation Photovoltaic Technologies*, AIP Conference Proceedings 404 (1997).

M.J. Keevers and M.A. Green, "Limiting efficiencies of ideal single multiple energy gap terrestrial solar cells", *J. Appl. Phys.* **75**, 4022 (1994).

S. Kettemann and J.-F. Guillemoles, "Limiting efficiency of LDS solar cells", *Proc. 13th European Photovoltaic Solar Energy Conference*, 119 (H.S. Stephens and Associates, Bedford, 1995).

M. Lundstrom, *Fundamentals of Carrier Transport* (Wokingham: Addison-Wesley, 1990).

A. Luque and A. Marti, "Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels", *Phys. Rev. Lett.* **78**, 5014 (1997).

A.J. Nozik, "Spectroscopy and hot electron relaxation dynamics in semiconductor. Quantum wells and quantum dots", *Ann. Rev. Phys. Chem.* (submitted) 2001.

R.T. Ross and A.J. Nozik, "Efficiency of hot carrier solar energy converters", *J. Appl. Phys.* **53**, 3813 (1982).

J.H. Werner, S. Kodolinski and H.J.Queisser, "Novel optimization principles and efficiency limits for semiconductor solar-cells", *Phys. Rev. Lett.* **72**, 3851 (1994).

J.H. Werner, R. Brendel and H.-J. Queisser, "Radiative efficiency limit of terrestrial solar-cells with internal carrier multiplication", *Appl. Phys. Letts.* **67**, 1028 (1995).

P. Wuerfel, "Solar energy conversion with hot electrons from impact ionisation", *Solar Energy Mater. Solar Cells* **46**, 43–52 (1997).