

Light-Emitting Diodes

E. Fred Schubert



CAMBRIDGE

CAMBRIDGE

more information - www.cambridge.org/9780521823302

This page intentionally left blank

LIGHT-EMITTING DIODES

This book covers all aspects of the technology and physics of infrared, visible-spectrum, and white-light-emitting diodes (LEDs) made from III–V semiconductors.

The book reviews elementary properties of LEDs such as the electrical and optical characteristics. The author also reviews advanced device physics including high-efficiency device designs, light extraction, radiative and non-radiative recombination dynamics, spontaneous recombination in resonant-cavity structures, and packaging. The reader is introduced to areas related to visible-spectrum and white LEDs such as human vision, photometry, colorimetry, and color rendering. Application of infrared and visible-spectrum LEDs in silica fiber, plastic fiber, and free-space communication is discussed. Extensive semiconductor material data, device design data, and analytic formulas governing the operation of LEDs are provided. Exercises and illustrative examples are used to reinforce the topics discussed. An introductory chapter reviews the historical developments and milestones of LED research and development.

This textbook will be of interest to scientists and engineers working on LEDs, notably in lighting, illumination and signage, and also to graduate students in electrical engineering, applied physics, and materials science.

E. FRED SCHUBERT received his MS degree in Electrical Engineering (Dipl.-Ing.) with honors from the University of Stuttgart, Germany, in 1981, and his Ph.D. degree (Dr.-Ing.) with honors in 1986, also in Electrical Engineering from the University of Stuttgart. From 1981 to 1985 he worked on compound semiconductor crystal growth at the Max Planck Institute for Solid State Research, Stuttgart, as a doctoral student. In 1985, he joined AT&T Bell Laboratories in Holmdel, NJ as a Postdoctoral Fellow. From 1988 to 1995, he was Principal Investigator in the Research Division of AT&T Bell Laboratories in Murray Hill, NJ. In 1995, he joined Boston University and was appointed tenured Full Professor in the Department of Electrical and Computer Engineering and Affiliated Member of the Photonics Center. At Boston University, he was responsible for GaN materials characterization and the fabrication and testing of compound semiconductor devices, in particular GaN-based devices.

In 2002 he was appointed Professor of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute in Troy, New York. At Rensselaer, he holds the Constellation Chair in Future Chips.

LIGHT-EMITTING DIODES

E. FRED SCHUBERT

Rensselaer Polytechnic Institute



CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 2RU, United Kingdom

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521823302

© E. Fred Schubert 2003

This book is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2003

ISBN-13 978-0-521-07704-3 eBook (EBL)

ISBN-10 0-521-07704-1 eBook (EBL)

ISBN-13 978-0-521-82330-2 hardback

ISBN-10 0-521-82330-7 hardback

ISBN-13 978-0-521-53351-5 paperback

ISBN-10 0-521-53351-1 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this book, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>Preface</i>	<i>page xi</i>
1 History of light-emitting diodes	1
1.1 History of SiC LEDs	1
1.2 History of GaAs and AlGaAs infrared and red LEDs	3
1.3 History of GaAsP LEDs	7
1.4 History of GaP and GaAsP LEDs doped with optically active impurities	9
1.5 History of GaN metal–semiconductor emitters	14
1.6 History of blue, green, and white LEDs based on GaInN p-n junctions	16
1.7 History of AlGaInP visible-spectrum LEDs	18
References	19
2 Radiative and non-radiative recombination	26
2.1 Radiative electron–hole recombination	26
2.2 Radiative recombination for low-level excitation	27
2.3 Radiative recombination for high-level excitation	31
2.4 Bimolecular rate equations for quantum well structures	32
2.5 Luminescence decay	33
2.6 Non-radiative recombination in the bulk	34
2.7 Non-radiative recombination at surfaces	40
2.8 Competition between radiative and non-radiative recombination	44
References	46
3 Theory of radiative recombination	48
3.1 The van Roosbroeck–Shockley model	48
3.2 The Einstein model	53
References	54

4	LED basics: electrical properties	55
4.1	Diode current–voltage characteristic	55
4.2	Deviations from the ideal I – V characteristic	59
4.3	Evaluation of diode parasitic resistances	60
4.4	Emission energy	62
4.5	Carrier distribution in p-n homojunctions	62
4.6	Carrier distribution in p-n heterojunctions	63
4.7	The effect of heterojunctions on device resistance	64
4.8	Carrier loss in double heterostructures	68
4.9	Carrier overflow in double heterostructures	71
4.10	Electron blocking layers	75
4.11	Diode voltage	77
4.12	Temperature dependence of diode voltage	79
4.13	Constant-current and constant-voltage DC drive circuits	81
	References	83
5	LED basics: optical properties	84
5.1	Internal, extraction, external, and power efficiency	84
5.2	Emission spectrum	85
5.3	The light escape cone	89
5.4	The lambertian emission pattern	92
5.5	Epoxy encapsulants	96
5.6	Temperature dependence of the emission intensity	97
	References	98
6	High internal efficiency LED designs	99
6.1	Double heterostructures	99
6.2	Doping of active region	102
6.3	P-n junction displacement	103
6.4	Doping of the confinement regions	105
6.5	Non-radiative recombination	108
6.6	Lattice matching	109
	References	112
7	High extraction efficiency structures	114
7.1	Absorption of below-band light in semiconductors	114
7.2	Double heterostructures	118
7.3	Shaping of LED dies	119
7.4	Current-spreading layer	123
7.5	Theory of current spreading	128

7.6	Current crowding in LEDs on insulating substrates	131
7.7	Cross-shaped contacts and other contact geometries	135
7.8	Transparent substrate technology	136
7.9	Anti-reflection optical coatings	138
7.10	Epoxy dome	139
7.11	Distributed Bragg reflectors	140
7.12	Current-blocking layers	147
7.13	Reflective and transparent contacts	148
7.14	Flip-chip packaging	149
	References	150
8	Visible-spectrum LEDs	155
8.1	The GaAsP, GaP, GaAsP:N, and GaP:N material systems	155
8.2	The AlGaAs/GaAs material system	160
8.3	The AlGaInP/GaAs material system	163
8.4	The GaInN material system	166
8.5	General characteristics of high-brightness LEDs	167
8.6	Optical characteristics of high-brightness LEDs	171
8.7	Electrical characteristics of high-brightness LEDs	173
	References	174
9	Spontaneous emission from resonant cavities	178
9.1	Modification of spontaneous emission	178
9.2	Fabry–Perot resonators	180
9.3	Reflectors	183
9.4	Optical mode density in a one-dimensional resonator	187
9.5	Spectral emission enhancement	191
9.6	Integrated emission enhancement	192
	References	194
10	Resonant cavity light-emitting diodes	198
10.1	Introduction and history	198
10.2	RCLED design rules	199
10.3	GaInAs/GaAs RCLEDs emitting at 930 nm	203
10.4	AlGaInP/GaAs RCLED emitting at 650 nm	209
10.5	Large-area photon recycling LEDs	211
10.6	Thresholdless lasers	213
10.7	Other RCLED devices	215
10.8	Other novel confined-photon emitters	215
	References	216

11	Human vision	219
11.1	Light receptors of the human eye	219
11.2	Basic radiometric and photometric units	220
11.3	Eye sensitivity function, luminous efficacy, and luminous efficiency	222
11.4	Color matching functions and chromaticity diagram	226
11.5	Color purity	229
11.6	LEDs in the chromaticity diagram	231
11.7	White illuminants and color temperature	232
11.8	Additive color mixing	234
11.9	Color-rendering index	236
	References	240
	Appendix 11.1	242
	Appendix 11.2	243
12	White-Light LEDs	245
12.1	Generation of white light	245
12.2	Wavelength converter materials	247
12.3	White LEDs based on phosphor converters	250
12.4	UV-pumped phosphor-based white LEDs	252
12.5	White LEDs based on semiconductor converters (PRS-LED)	253
12.6	Calculation of the power ratio of PRS-LED	254
12.7	Calculation of the luminous performance of PRS-LED	255
12.8	Spectrum of PRS-LED	257
12.9	White LEDs based on dye converters	259
	References	259
13	Optical communication	261
13.1	Types of optical fibers	261
13.2	Attenuation in silica and plastic optical fibers	263
13.3	Modal dispersion in fibers	265
13.4	Material dispersion in fibers	266
13.5	The numerical aperture of fibers	269
13.6	Coupling with lenses	271
13.7	Free-space optical communication	273
	References	276
14	Communication LEDs	277
14.1	LEDs for free-space communication	277
14.2	LEDs for fiber-optic communication	277

14.3	Surface-emitting Burrus-type communication LEDs emitting at 870 nm	278
14.4	Surface-emitting communication LEDs emitting at 1300 nm	279
14.5	Communication LEDs emitting at 650 nm	281
14.6	Edge-emitting superluminescent diodes (SLDs)	283
	References	287
15	LED modulation characteristics	288
15.1	Rise and fall times, 3 dB frequency, and bandwidth in linear circuit theory	288
15.2	Rise and fall time in the limit of large diode capacitance	290
15.3	Rise and fall time in the limit of small diode capacitance	291
15.4	Voltage dependence of the rise and fall times	293
15.5	Carrier sweep-out of the active region	294
15.6	Current shaping	295
15.7	dB frequency	297
15.8	Eye diagram	297
15.9	Carrier lifetime and 3 dB frequency	297
	References	299
Appendix 1	List of frequently used symbols	301
Appendix 2	Physical constants	305
Appendix 3	Periodic system of elements	306
Appendix 4	Room-temperature properties of semiconductors	307
	<i>Index</i>	309

Preface

Technical progress in the field of light-emitting diodes (LEDs) has been breathtaking during the last few decades. State-of-the art LEDs are small, rugged, reliable, bright, and efficient. In contrast to many other light sources, LEDs have the potential of converting electricity to light with near-unity efficiency.

The success story of LEDs has not ended but is still in full progress. Great technological advances will surely continue to be made. As a result, it is expected that LEDs will play an increasingly important role as light sources and will become the dominant light source in the future.

LEDs were discovered by accident early in the last century and the first LED results were published in 1907. LEDs became forgotten only to be rediscovered in the 1920s and again in the 1950s. In the 1960s, several groups pursued the demonstration of semiconductor lasers. The first viable LEDs were by-products in this pursuit. During the last 40 years, LEDs have become devices in their own right and today are versatile light sources with a bright future.

This book is dedicated to the technology and physics of LEDs. It reviews the electrical and optical fundamentals of LEDs as well as advanced device structures. Recent technological breakthroughs are also discussed. The book deals with LEDs made from III–V semiconductors. However, much of the science and technology discussed is relevant to other solid-state light emitters, including emitters made of polymers. Several application areas of LEDs are discussed as well, including illumination and communication applications.

Many colleagues have provided information not readily available and have given invaluable suggestions on the manuscript. In particular, I wish to thank Drs. M. George Craford (LumiLeds Corp.), Nick Holonyak Jr. (University of Illinois), Mike Krames (LumiLeds Corp.), Ralph Logan (retired, formerly with AT&T Bell Laboratories), Fred Long (Rutgers University), Paul Maruska (Crystal Photonics Corp.), Gerd Mueller (LumiLeds Corp.), Shuji Nakamura (University of California, Santa Barbara), Yoshihiro Ohno (National Institute of Standards and

Technology), Jacques Pankove (Astralux Corp.), Manfred Pilkuhn (University of Stuttgart, Germany), Hans Rupprecht (retired, formerly with IBM Corp.), Klaus Streubel (Osram Opto Semiconductors Corp., Germany), Li-Wei Tu (National Sun Yat-Sen University, Taiwan), Jerry Woodall (Yale University), and Walter Yao (Advanced Micro Devices Corp.).

E. F. Schubert

1

History of light-emitting diodes

1.1 History of SiC LEDs

Starting early in the 20th century, light emission from a solid material, caused by an electrical power source, was reported, a phenomenon termed *electroluminescence*. The first light-emitting diode (LED) had been born. At that time, the material properties were poorly controlled, and the emission process was not well understood. The light-emitting active material was SiC crystallites as used for sandpaper abrasive.

A Note on Carborundum.

To the Editors of Electrical World:

SIRs:—During an investigation of the unsymmetrical passage of current through a contact of carborundum and other substances a curious phenomenon was noted. On applying a potential of 10 volts between two points on a crystal of carborundum, the crystal gave out a yellowish light. Only one or two specimens could be found which gave a bright glow on such a low voltage, but with 110 volts a large number could be found to glow. In some crystals only edges gave the light and others gave instead of a yellow light green, orange or blue. In all cases tested the glow appears to come from the negative pole. a bright blue-green spark appearing at the positive pole. In a single crystal, if contact is made near the center with the negative pole, and the positive pole is put in contact at any other place, only one section of the crystal will glow and that the same section wherever the positive pole is placed.

There seems to be some connection between the above effect and the e.m.f. produced by a junction of carborundum and another conductor when heated by a direct or alternating current; but the connection may be only secondary as an obvious explanation of the e.m.f. effect is the thermoelectric one. The writer would be glad of references to any published account of an investigation of this or any allied phenomena.

NEW YORK, N. Y.

H. J. ROUND.

Fig. 1.1. Publication reporting on a “curious phenomenon”, namely the first observation of electroluminescence from a SiC (carborundum) light-emitting diode. The article indicates that the first LED was a Schottky diode rather than a p-n-junction diode (after H. J. Round, *Electrical World* **49**, 309, 1907).

The first report on light-emitting diodes, shown in Fig. 1.1, was published by Henry Joseph Round (1907). Round was a radio engineer and a prolific inventor who, by the end of his career, held 117 patents. These first light-emitting devices had rectifying current–voltage characteristics;

that is, these first devices were light-emitting *diodes* or LEDs. The light was produced by touching the SiC crystal with electrodes so that a rectifying Schottky contact was formed. Schottky diodes are usually majority carrier devices. However, minority carriers can be created either under strong forward-bias conditions, or by avalanche multiplication under reverse-bias conditions.

The mechanism of light emission in a forward-biased Schottky diode is shown in Fig. 1.2, which displays the band diagram of a metal–semiconductor junction under (a) equilibrium, (b) moderate forward bias, and (c) strong forward bias conditions. The semiconductor is assumed to be of n-type conductivity. Under strong forward bias conditions, minority carriers are injected into the semiconductor by tunneling through the surface potential barrier. Light is emitted upon recombination of the minority carriers with the n-type majority carriers. The voltage required for minority carrier injection in Schottky diodes is larger than typical p-n junction LED voltages. Round (1907) reported operating voltages ranging between 10 and 110 V.

Light can also be generated in a Schottky diode under reverse-bias conditions through the avalanche effect in which high-energy carriers impact-ionize atoms of the semiconductor. In this process, holes are created in the valence band as well as electrons in the conduction band, which will eventually recombine thereby creating light.

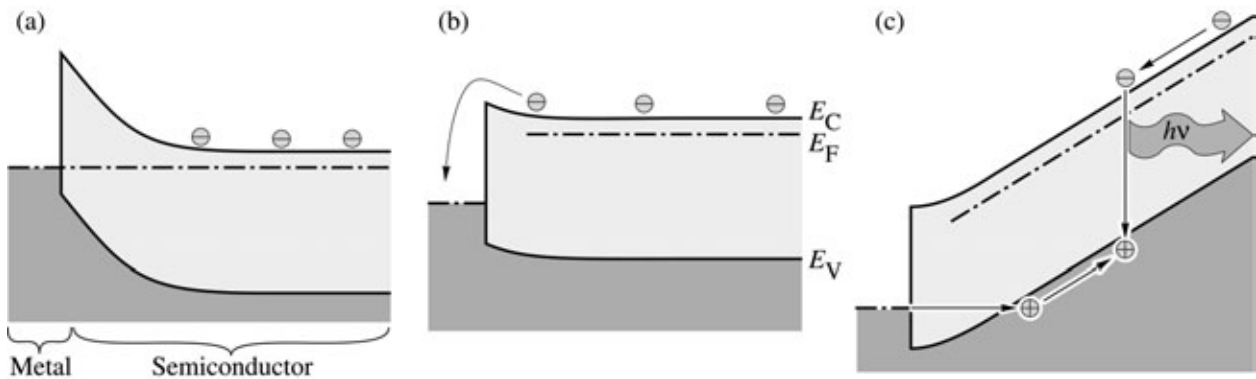


Fig. 1.2. Band diagram of a Schottky diode under (a) equilibrium conditions, (b) forward bias, and (c) strong forward bias. Under strong forward bias, minority carrier injection occurs making possible near-bandgap light emission.

Lossev (1928) reported detailed investigations of the luminescence phenomenon observed with SiC metal–semiconductor rectifiers. The main use of these rectifiers was in solid-state demodulation radio-circuits that did not employ vacuum tubes. Lossev found that luminescence occurred in some diodes when biased in the reverse direction and in some diodes when biased in

forward *and* reverse directions. The author was puzzled about the physical origin of the luminescence. He investigated whether light was generated by heat glow (incandescence) by testing the evaporation rate of a droplet of liquid benzene on the luminous sample surface. He found, however, that the benzene evaporated very slowly and correctly concluded that the luminescence was not caused by incandescence. He postulated that the process by which light was produced is “very similar to cold electronic discharge”. The author also found that the light could be switched on and off very rapidly, making the device suitable for what he called a “light relay”. The pre-1960 history of LEDs was further reviewed by Loebner (1976).

By the late 1960s, SiC films had been prepared by more careful processes (Violin *et al.*, 1969), and p-n junction devices were fabricated, leading to blue light-emitting diodes. Electrical to optical conversion efficiencies were only 0.005 % (Potter *et al.*, 1969). In the ensuing decades, blue SiC LEDs were never substantially improved, because SiC has an indirect bandgap. Although many blue SiC LEDs were actually sold commercially in the early 1990s, they are no longer a viable product. In the end, the best SiC LEDs, emitting blue light at 470 nm, had an efficiency of only 0.03 % (Edmond *et al.*, 1993). SiC, the material of the very first LED, could no longer compete with III–V semiconductors.

1.2 History of GaAs and AlGaAs infrared and red LEDs

Prior to the 1950s, SiC and II–VI semiconductors had been well-known materials. Both materials occur in nature. The very first LEDs had been made using SiC and there had been one publication by Destriau (1936) reporting LEDs made of zincblende (ZnS).

The era of III–V compound semiconductors started in the early 1950s when this class of materials was postulated and demonstrated by Welker (1952, 1953). The class of III–V compounds had been an unknown substance prior to the 1950s that does not occur naturally. The novel man-made III–V compounds proved to be optically very active and thus instrumental to modern LED technology.

Bulk growth of the III–V compound GaAs commenced in 1954. In the mid-1950s, large single-crystal boules of GaAs were pulled from the melt. The sliced and polished wafers were used as substrates for the epitaxial growth of p-n junction diode structures, either by vapor phase epitaxy (VPE) or liquid phase epitaxy (LPE). Infrared (870 – 980 nm) LEDs and lasers based on GaAs were first reported in 1962 by groups working at RCA, GE, IBM, and MIT (Pankove and Berkeyheiser, 1962; Pankove and Massoulie, 1962; Hall *et al.*, 1962; Nathan *et al.*, 1962; Quist *et al.*, 1962).

A sustained research effort on GaAs and AlGaAs/GaAs devices started in the early 1960s at the IBM Thomas J. Watson Research Center in Yorktown Heights, located about an hour's drive north of New York City. The IBM team consisted of well-known researchers such as Jerry Woodall, Hans Rupprecht, Manfred Pilkuhn, Marshall Nathan, and others.

Woodall (2000) recalls that his work centered on the bulk crystal growth of GaAs used to fabricate semi-insulating substrates for Ge device epitaxy, and n-type substrates to fabricate injection lasers via Zn diffusion. At that time the GaAs-based injection laser had already been demonstrated at IBM, GE, and MIT Lincoln Laboratories. Rupprecht's interests were in impurity-diffusion theory and experiment along with experimental investigations into the newly discovered injection laser. Rupprecht was associated with a laser device physics group headed by Marshall Nathan, a co-inventor of the first injection laser (Nathan *et al.*, 1962).

As Woodall developed a technique that lead to state-of-the-art horizontal Bridgman GaAs crystals, Rupprecht fabricated the materials into lasers and characterized them. This collaboration paid off immediately and continuous-wave (cw) operation of GaAs lasers at 77 K was attained (Rupprecht *et al.*, 1963). They then learned of the liquid phase epitaxy (LPE) technique pioneered by Herb Nelson at the RCA Labs in Princeton. The employment of LPE to grow GaAs lasers resulted in the achievement of 300 K lasers with lower threshold current densities than for Zn diffused lasers. Stimulated by papers found in a literature search, Woodall set out to grow GaAs p-n junction diodes by using Si as an amphoteric dopant, i.e. Si atoms on Ga sites acting as donors and Si atoms on As sites acting as acceptors. This was an interesting idea, as hitherto LPE had been used to grow epilayers with only a single conductivity type.

The LPE conditions to form Si-doped p-n junctions were found very quickly. Si-doped GaAs p-n junctions were formed by cooling a Ga-As-Si melt from 900 to 850 °C to form Si donors and Si acceptors at the two temperatures, respectively. By examining the cross section of the chemically stained epitaxial layer, the lower layer, grown at 900 °C, was identified as being an n-type layer and the upper layer, grown at 850 °C, as a p-type layer. No loss in crystal quality was found in the regions of lower temperature growth. Furthermore, owing to band tailing effects caused by the highly doped, compensated region of the p-n junction, the LED emission occurred at 900 – 980 nm, far enough below the GaAs band edge (870 nm) that the bulk GaAs substrate and the GaAs epilayer did not absorb much of the emitted light but acted as a transparent “window layer”. LED external quantum efficiencies as high as 6 % were attained, a major breakthrough in LED technology (Rupprecht *et al.*, 1966). Rupprecht (2000) stated: “Our demonstration of the highly efficient GaAs LED is a typical example of a discovery made by

serendipity”. The quantum efficiency of the amphoterically doped GaAs LEDs was five times greater than that of GaAs p-n junctions fabricated by Zn diffusion. Si acceptor levels are deeper than Zn acceptor levels so that the emission from the compensated Si-doped active region occurs at longer wavelengths where GaAs is more transparent.

Being in the LED research business, the IBM group wondered if this doping effect could be extended to a crystal host with visible emission. There were two candidates, GaAsP and AlGaAs. Whereas Rupperecht tried to do GaAsP epitaxy via LPE, Woodall set up an apparatus for AlGaAs. It was difficult, to form good quality GaAsP epilayers by LPE due to the 3.6 % lattice mismatch between GaP and GaAs. AlGaAs had problems of its own: “AlGaAs is lousy material” was the pervasive opinion at that time, because, as Woodall (2000) stated, “aluminum loves oxygen”. This results in the incorporation of the “luminescence killer” oxygen in AlGaAs layers; in particular, in the vapor-phase epitaxy (VPE) process but less so in the LPE process.

Without the support of IBM management, Rupperecht and Woodall “went underground” with their research, conducting the LPE AlGaAs epigrowth experiments after regular working hours and on the weekends. Woodall designed and built a “vertical dipping”-type LPE apparatus, using graphite and alumina melt containers. As an undergraduate student Woodall had majored in metallurgy at MIT and he remembered something about phase diagrams. He made an “intelligent guess” to select the Al concentrations for the LPE melts. He added Si to the melt for the first experiment, saturated the melt and then “dipped” the GaAs substrate while cooling the melt from about 925 to 850 °C. Finally, the substrate and epilayer were withdrawn from the melt, and the apparatus was returned to 300 K. Although no Si-doped p-n junction was observed, a 100 μm thick high quality layer of AlGaAs had been grown with a bandgap in the red portion of the visible spectrum (Rupperecht *et al.*, 1967, 1968).

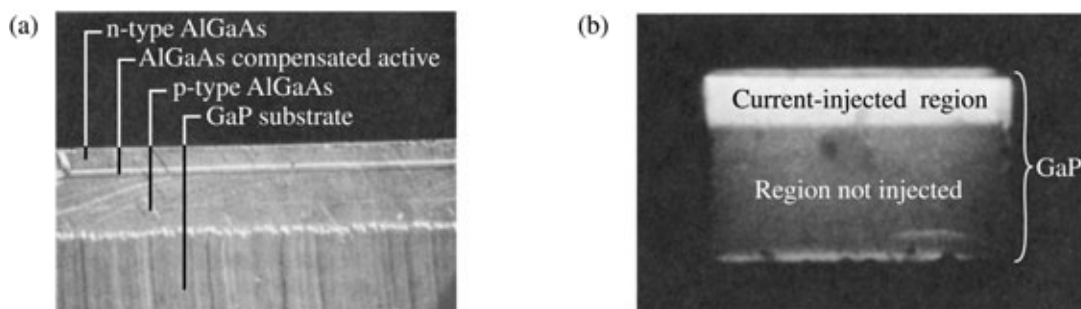


Fig. 1.3. (a) Cross section micrograph of an AlGaAs LED grown on a transparent GaP substrate. (b) Electroluminescence originating from a current-injected region located under a stripe-shaped contact viewed through the transparent GaP substrate (after Woodall *et al.*, 1972).

Visible AlGaAs were also grown on GaP, a lattice mismatched but transparent substrate. Micrographs of the structure are shown in Fig. 1.3. When AlGaAs was grown on GaP substrates, the thermodynamics of LPE made the initially grown material Al-rich due to the Al distribution coefficient in the melt. As a result, the high-Al-content AlGaAs acts as a transparent window layer for the light emitted from the low-Al-content AlGaAs active region (Woodall *et al.*, 1972).



Fig. 1.4. This classic 1964 mainframe computer IBM System 360 used high-voltage gas-discharge lamps to indicate the status of the arithmetic unit. In later models, the lamps were replaced by LEDs. The cabinet-sized 360 had a performance comparable to a current low-end laptop computer.

Pilkuhn, also an “IBM’er” who had worked with Rupprecht on GaAsP LEDs and lasers (Pilkuhn and Rupprecht, 1965), had built a small battery-powered circuit with an LED emitting visible red light, which he showed to his colleagues and management at IBM (Pilkuhn, 2000). The reactions ranged from “nice but useless” to “great and useful”. However it was soon realized that the latter was true, i.e. that LEDs were extremely useful devices. The first application of the GaAsP LEDs was as indicator lights on circuit boards, where the LEDs indicated the status and proper function of the circuit board. LEDs were also used to show the status of the data processing unit of the classic IBM System 360 mainframe computer shown in Fig. 1.4.

According to Rostky (1997), the first commercial GaAs LED was offered by the Texas Instruments Corporation in the early 1960s. The LED emitted infrared radiation near 870 nm. The manufacturing quantities of the product were low, probably caused by the high price for one LED, which reportedly was 130 US\$.

The *resonant-cavity light-emitting diode* (RCLED) was first demonstrated in the AlGaAs/GaAs materials system (Schubert *et al.*, 1992, 1994). RCLEDs represented a new class

of LEDs making use of spontaneous emission enhancement occurring in microscopic optical resonators or *microcavities*. Enhancement is greatest for wavelengths in resonance with the fundamental mode of the cavity. The emission enhancement is mediated by changes in the optical mode density within the cavity. RCLEDs have higher emission intensities along the optical axis of the cavity, which allows for higher coupling efficiencies to optical fibers.

At the present time, infrared GaAs/AlGaAs LEDs are widely used in video and audio remote controls and as sources for local-area communication networks. In addition, red AlGaAs/AlGaAs LEDs are used as high-brightness visible LEDs with efficiencies higher than the GaAsP/GaAs red LEDs but lower than the AlGaInP/GaAs red LEDs.

1.3 History of GaAsP LEDs

The beginning of visible LEDs dates back to the year 1962 when Holonyak and Bevacqua (1962) reported on the emission of coherent visible light emission from GaAsP junctions in the first volume of *Applied Physics Letters*. Although the emission of coherent light was only observed at low temperatures, the devices worked as LEDs and emitted visible light even at room temperature. This publication marks the beginning of viable p-n junction LEDs emitting in the visible wavelength range.

Nick Holonyak Jr., who in 1962 worked at General Electric in Syracuse, New York, and who later joined the University of Illinois, had used vapor-phase epitaxy (VPE) of GaAsP on GaAs substrates. This technique is suited for large-volume growth of wafers in a research as well as a manufacturing environment. Holonyak (2000) recalls that when he had made these first LEDs, at that time he had already envisioned many applications these new devices may have, including indicator lights, seven-segment numeric displays, and alphanumeric displays.

However, despite the early success of the Holonyak group, the goal of demonstrating a semiconductor laser, working at room temperature, remained elusive (Holonyak, 1963, 1964). It remained elusive for good reasons. The GaAsP material system grown on GaAs substrates has several problems which Holonyak and co-workers discovered.

Although excellent electrical junction characteristics were obtained (Holonyak *et al.*, 1963a), the optical properties degraded. When the phosphorus content in GaAsP was about 45 – 50 %, a strong decrease in the LED radiative efficiency was found. These difficulties were attributed to the direct–indirect transition of the bandgap of GaAsP (Holonyak *et al.*, 1963b, 1966; Pilkuhn and Rupprecht, 1964, 1965). It was determined that the 300 K efficiency of GaAsP alloy devices

dropped to less than 0.005 % when the phosphorus concentration exceeded 44 % (Maruska and Pankove, 1967).

The first commercial GaAsP LED was offered by the General Electric (GE) Corporation in the early 1960s. The LED emitted visible radiation in the red part of the spectrum. The manufactured quantities of the product were low, probably due to the high price which was 260 US\$ for a single LED. The product was offered in the Allied Radio catalog, a widely distributed catalog for amateur radio electronics (Rostky, 1997).

The Monsanto Corporation was the first commercial entity to start mass production of LEDs. In 1968, the company had set up a factory, produced low-cost GaAsP LEDs, and sold them to customers. The era of solid-state lamps had started. In the period 1968 – 1970, sales were skyrocketing, doubling every few months (Rostky, 1997). The Monsanto LEDs were based on GaAsP p-n junctions grown on GaAs substrates emitting in the visible red wavelength range (Herzog *et al.*, 1969; Craford *et al.*, 1972).

Monsanto developed a friendly collaboration with Hewlett-Packard (HP) Corporation, expecting that HP would make light-emitting diodes and LED displays while Monsanto would provide the raw material – the GaAsP. In the mid-1960s, Monsanto had sent one of its scientists from Saint Louis, Missouri, to Palo Alto, California, to help HP develop an LED business using Monsanto's GaAsP material. However, HP felt nervous about depending on a single source for the GaAsP material. The informal relationship ended and HP started growing its own GaAsP (Rostky, 1997).

For several years, from the late 1960s to the mid-1970s, the emerging market was in numeric displays, driven at first by calculators, then by wristwatches, following Hamilton Watch Corporation's introduction of the Pulsar digital watch in 1972. For a while, the early contenders, Monsanto and HP, took turns leaping into first place with a more advanced multiple-numeric or alphanumeric LED displays (Rostky, 1997).

A key technical innovator and manager at Monsanto was M. George Craford who has made numerous contributions to LEDs including the first demonstration of a yellow LED (Craford *et al.*, 1972). It employed an N-doped GaAsP active region grown on a GaAs substrate. When Monsanto sold off its optoelectronics business in 1979, Craford joined HP and became the key person in the company's LED business. A profile of Craford was published by Perry (1995).

It soon became clear that the large lattice mismatch between the GaAs substrate and the GaAsP epilayer resulted in a high density of dislocations (Wolfe *et al.*, 1965; Nuese *et al.*, 1966). As a result, the external efficiency of these LEDs was quite low, about 0.2 % or less (Isihamatsu

and Okuno, 1989). The importance of the growth conditions and thickness of a *buffer layer* was realized by Nuese *et al.* (1969) who pointed out that a thick graded GaAsP buffer layer yields improved brightness red LEDs. It is understood today that the thick graded buffer layer reduces the high dislocation density in the GaAsP epitaxial layer originating near the GaAsP-epilayer-to-GaAs-substrate boundary.

The direct–indirect transitions as well as the high dislocation density limit the brightness attainable with GaAsP LEDs. Today this material system is primarily used for standard (low-brightness) red LEDs for indicator lamp applications.

1.4 History of GaP and GaAsP LEDs doped with optically active impurities

Ralph Logan's and his co-workers' pioneering work on GaP LEDs was done while working at Bell Laboratories in Murray Hill, New Jersey, in the early 1960s, where they developed a manufacturing process for GaP-based red and green LEDs. At that time, semiconductors had been employed to demonstrate both bipolar and field-effect transistors for switching and amplifying electrical currents. Engineers and scientists back then also began to realize that semiconductors would be perfectly suitable for light-emitting devices.

Logan (2000) recalls that his interest was stimulated by the first reports of GaP p-n junction LEDs by Allen *et al.* (1963) and Grimmeiss and Scholz (1964). These devices emitted red light at a useful efficiency so that the light could be clearly seen with the naked eye under ambient daylight conditions. The Grimmeiss–Scholz junctions were reported to be made by alloying Sn, an n-type dopant, into p-type GaP.

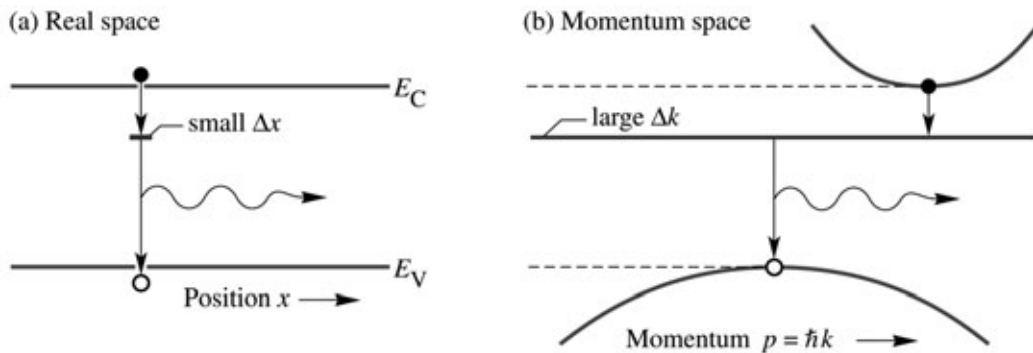


Fig. 1.5. (a) Real-space and (b) momentum-space optical transitions in GaP doped with an optically active impurity such as O or N, emitting in the red and green parts of the visible spectrum, respectively. GaP LEDs employ the *uncertainty principle* ($\Delta x \Delta p \geq \hbar$) which predicts that an electron wave function localized in real space is delocalized in momentum space, thereby making momentum-conserving (vertical) transitions possible.

GaP is an indirect-gap semiconductor that does not emit significant amounts of light due to the requirement of momentum conservation in optical transitions. Figure 1.5 shows the band diagram of GaP illustrating that the band extremum points occur at different values in momentum space. If GaP is doped with an *optically active isoelectronic impurity* such as N, strong optical transitions are obtained, as shown by Thomas *et al.* (1965), because the impurity levels are smeared out in momentum space. GaP doped with optically active impurities is a wonderful example of a practical device based on the *Heisenberg uncertainty principle*, which predicts that an impurity with a strongly localized wavefunction in position space (small Δx) will have a delocalized level in momentum space (large Δk), so that optical transitions can occur via the deep level state.

The growth of GaP was accomplished by using platelets grown from a solution containing Ga and P. The platelets had lateral dimensions of $0.5\text{ cm} \times 1\text{ cm}$ and were grown to a thickness of about 1 mm. This was the standard initial method of growing GaP and overcame the problem of dealing with the P overpressure required when growing GaP at high temperatures. No one at Bell Labs could immediately reproduce the remarkable results reported by Grimmeiss and Scholz. However, a big research effort in electroluminescence was launched at AT&T Bell Labs as a result.

In the solution growth of the GaP platelets, the dopants used were Zn and O (the latter from Ga_2O_3), but it was not generally realized that ordinary ambient air usually contains enough S to be a good n-dopant. The growth kinetics of the platelets had the intriguing result that the compensated melts generally produced an n-type layer on one platelet surface so that a p-n junction was formed under the GaP platelet surface. This was thought to explain Grimmeiss' results. Logan *et al.* (1967a) published these findings at once.

Logan's research group also reported the first demonstration of reproducible growth of efficient LEDs (Logan *et al.* 1967b). These junctions were formed by growing an n-type GaP layer onto Zn-O-doped GaP wafers that were polished out of large solution-grown wafers with typical sizes of $1 \times 1\text{ inch}^2$. Logan *et al.* found that post-growth annealing in the range $400 - 725\text{ }^\circ\text{C}$ could increase the LED efficiency by as much as an order of magnitude, yielding efficiencies exceeding 2 %. The annealing was thought to diffuse the Zn to the O atoms, thereby increasing the density of isoelectronic Zn-O centers that mediated the electroluminescence.

At the end of the 1960s ingots of GaP grown from melts at high temperature and pressure were becoming available, suitable for being cut into real substrates as we know them today. Green LEDs were formed with efficiencies as high as 0.6 % by doping the GaP with N

isoelectronic impurities (Logan *et al.*, 1968, 1971). The N was added in the form of GaN to the growth melts used to form the p-n junctions. While the external quantum efficiency of green LEDs is less than for the red LEDs, the sensitivity of the human eye to green light is more than 10 times higher than in the red, so the apparent brightness of the LEDs is comparable.

Other research laboratories such as IBM, RCA, and GE also looked into the possibility of making visible LEDs that were more efficient than those made of GaAsP. Research on GaP LEDs was pursued at IBM Corporation's Thomas J. Watson Research Center in Yorktown Heights in New York State. Manfred Pilkuhn and co-workers demonstrated an LPE-grown red GaP LED doped with Zn and O. The picture of a GaP LED with top and bottom contacts is shown in Fig. 1.6. The *IBM Research Journal* proudly reported that “brilliant red light” was emitted from the p-n junction. Note that in the 1960's, monochromatic colors were mostly generated by filtering incandescent light, so that the narrow-spectral-width LED light appeared to the observer as an impressively pure and “brilliant” color.

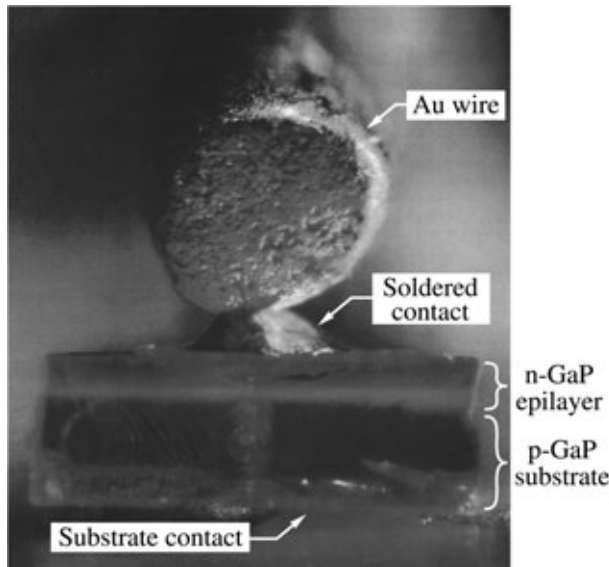


Fig. 1.6. GaP light-emitting diode grown by liquid-phase epitaxy emitting “brilliant red light” from the Zn- and O-doped p-n junction region (courtesy of Pilkuhn, 2000).

The active region of Pilkuhn's GaP LEDs were co-doped with acceptors, e.g. Zn acceptors, and donors, e.g. Te, S, or Se donors, so that light was generated predominantly by donor–acceptor pair recombination processes. The energy of the light was below the bandgap of GaP. It was also found that co-doping of GaP with Zn and O resulted in a particularly large wavelength shift so that the emission occurred in the red wavelength range (Foster and Pilkuhn, 1965). Oxygen in GaP is neither donor nor acceptor but was identified as a deep level (see, for example, Pilkuhn, 1981).

Logan and co-workers, and their management team at AT&T immediately realized that there were many possible applications for LEDs. Indicator lamps were becoming useful in the telephone business. All such lights used at this time operated using 110 V. An example is the “Princess” telephone, which was intended to be used in bedrooms – the dial lit up when the phone was picked up from its cradle. The “Princess” was a prestigious telephone and the latest fad in the 1960s but had to be installed near a 110 V outlet. A service call to the local phone company was needed if the bulb burned out. If LEDs were to replace the 110 V light bulbs, the phone line could power the LEDs and a 110 V outlet would no longer be needed. In addition, GaP LEDs had an expected lifetime exceeding 50 years when used in telephones, much longer than 110 V light bulbs, so that this reliability promised substantial cost savings for the “Bell System” or simply “Ma Bell”, as the phone company was called at that time.

More important was the multi-line “key telephone”. This is the multi-line telephone used in large offices mostly by operators and secretaries where indicator lamps tell which line is being called and busy. To switch the telephone lines and the 110 V indicator lamps, a remote switch was used with dozens of wires to each phone. Installing and servicing these phones was very costly. In present-day telephones, the LED indicator lamps are powered over the phone line. A compatible circuit inside the phone handles the switching of the indicator lamps and phone lines. The savings in telephone manufacturing, installation, and service were impressive.

With the demonstration of the reproducible growth of efficient green N-doped GaP LEDs and red Zn-O-codoped LEDs both of which were about equally bright and useful, the Bell Laboratories Development Department decided to manufacture the LEDs at its Reading, Pennsylvania, facility.



Fig. 1.7. AT&T telephone set (“Trimline” model) with the dial pad illuminated by two green N-doped GaP LEDs. The illuminated dial pad was one of the first applications of green GaP:N LEDs.

Telephone lines typically operate with a line voltage of approximately 40 V DC with currents of several milliamps. The only effect of inserting an LED into this circuit is to reduce the drive voltage by approximately 2 V, a negligible effect, while the efficient LED makes a good indicator lamp. As a result, many phone models were equipped with an illuminated dial pad. Both red and green LED illumination was available, and green was the final choice made by telephone designers. Figure 1.7 shows a 1990 version of the AT&T Trimline telephone – still using GaP:N green LEDs for the dial pad illumination. Red and green LEDs were also incorporated in the multi-line “key telephones”.

Should the reader ever be near Murray Hill, New Jersey, visiting the Bell Labs Museum, located at 600 Mountain Avenue, should be considered. Many technical artifacts including Logan *et al.*'s green GaP:N LED are displayed in the museum.

The Monsanto team applied N doping to GaAsP to attain emission in the red, orange, yellow, and green wavelength ranges (Groves *et al.*, 1971; Craford *et al.*, 1972; for a review see Duke and Holonyak, 1973). Many parameters, such as the emission and the absorption wavelength and the solubility of N in GaAsP and GaP were investigated. A useful growth method became vapor phase epitaxy (VPE), because it allowed for N-doping in the vicinity of the p-n junction only. This resulted in less absorption of light in the layers adjoining the p-n junction and higher overall LED efficiencies (Groves *et al.*, 1977, 1978a, 1978b). Today, GaP:N is the primary material for green emitters used in low-brightness applications such as indicator lights.



Fig. 1.8. Programmable pocket calculators Model SR-56 of the Texas Instruments Corporation and Model HP-67 of the Hewlett-Packard Corporation, both first manufactured in 1976. Seven-segment numeric characters composed of GaAsP LEDs were used in the display. The SR-56 came with a “huge” program memory of 100 steps. The HP-67 came with a magnetic card reader and had several freely programmable keys.



Another early application of LEDs was the numeric display in pocket calculators and digital wristwatches. Figure 1.8 shows two programmable calculators of the mid 1970s, the Texas Instruments Corporation's SR-56 and the Hewlett-Packard Corporation's HP-67. Both used red GaAsP LEDs in the seven-segment numeric display. All calculators using LED displays shared a significant problem: the display could not be read under bright outdoor conditions because the light emitted by the LEDs was simply too dim; furthermore, the power consumption of LED displays was high. Digital watches using LEDs shared the same problem. Liquid crystal displays (LCDs), introduced at the end of the 1970s, had much lower power requirements. Consequently, LCDs totally replaced LED displays in calculators and watches by the beginning of the 1980s.

1.5 History of GaN metal–semiconductor emitters

In the late 1960s, the *Radio Corporation of America* (RCA) was one of the premier manufacturers of color televisions, using cathode ray tubes (CRT) with three electron guns, to display images. At RCA's central research laboratory in Princeton, New Jersey, James Tietjen had become the director of the Materials Research Division, and he wanted to develop a flat-panel television display that could be hung on the wall like a painting. To create a full color image, the display must contain red, green, and blue pixels. Tietjen realized that red LEDs using GaAsP and green LEDs using GaP:N technology were already available. All that was needed for a flat TV based on LEDs was a bright blue LED.

In May 1968, Tietjen approached a young man in his group, Paul Maruska, and challenged him to find a method for growing single-crystal films of GaN, which Tietjen felt would yield blue LEDs. Maruska had been growing GaAsP red LEDs using the metal-halide vapor phase epitaxy (MHVPE) approach. He gained much experience with the promises and perils of III–V compounds including phosphorus, a pyrophoric substance. On a day in 1968, phosphorus caused a garbage truck to catch fire on New Jersey's Route 1 in Princeton shortly after picking up some phosphorus-containing laboratory waste at the RCA Labs. The driver of the truck decided to immediately return the burning and smoking load to RCA and dump it on the front lawn of the research labs (Maruska, 2000).

When Maruska started working on GaN, he first went to the library at Princeton University and thoroughly studied copies of all the old papers on GaN from the 1930s and 1940s (Juza and Hahn, 1938). GaN had been prepared as a powder by reacting ammonia with liquid gallium metal at elevated temperatures. He chose sapphire as the substrate because it is a robust material that is not reactive with ammonia. Unfortunately, he misinterpreted the results of Lorenz and

Binkowski (1962), who had reported the decomposition of GaN in vacuum at temperatures as low as 600 °C. All of his early GaN films were grown at temperatures below 600 °C to prevent decomposition, and hence were polycrystalline. Finally in March 1969, Maruska realized that in an ammonia environment, growth would occur rather than decomposition, and thus he raised the furnace temperature to 850 °C, the temperature typically used for GaAs growth. The sapphire appeared to be uncoated, because the GaN film was clear and had a specular surface. He rushed down to the RCA analytical center, and a Laue pattern revealed that the deposit was indeed the first single-crystal film of GaN (Maruska and Tietjen, 1969).

Maruska found that all of the GaN films were n-type without intentional doping. He sought to find a p-type dopant so that he could make a p-n junction. Zinc seemed to be an appropriate acceptor because it worked for GaAs and GaP. With heavy Zn concentrations, GaN films proved to be insulating. But they were never conducting p-type (Maruska, 2000).

During 1969, Jacques Pankove spent a sabbatical year at Berkeley University writing his classic textbook, *“Optical Processes in Semiconductors”*. When he returned to RCA Labs in January 1970, he immediately became interested and strongly involved in the new GaN films. Pankove *et al.* undertook a study of optical absorption and photoluminescence of thin-film GaN (Pankove *et al.*, 1970a, 1970b). The first example of electroluminescence from GaN was announced at RCA in the summer of 1971 (Pankove *et al.* 1971a). The sample consisted of an insulating Zn-doped layer which was contacted with two surface probes, and blue light centered at 475 nm was emitted. Pankove and co-workers then made a device consisting of an undoped n-type region, an insulating Zn-doped layer, and an indium surface contact (Pankove *et al.*, 1971b, 1972). This **metal–insulator–semiconductor (MIS) diode** was the first current-injected GaN light emitter, and it emitted green and blue light.

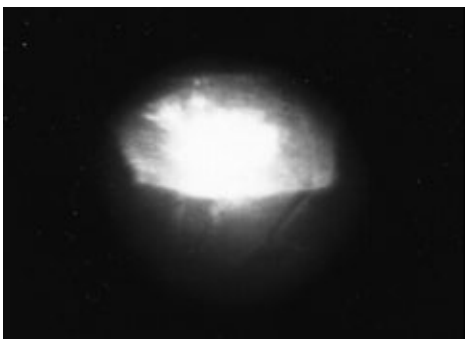


Fig. 1.9. Blue light emission found in 1972 caused by recombining electron–hole pairs created in a highly resistive GaN structure doped with Si and Mg (courtesy of Maruska, 2000)

The RCA team speculated that magnesium might be a better choice of p-type dopant than zinc. They began growing Mg-doped GaN films using the MHPVE technique, and in July 1972, obtained blue and violet emission centered at 430 nm as shown in Fig. 1.9 (Maruska *et al.*, 1972). One of these Mg-doped blue light MIS emitters continues to emit light even today. Maruska *et al.* (1973) described these efforts in a paper entitled “Violet luminescence of Mg-doped GaN”. Note that the GaN films, even though Mg doped, did not exhibit p-type conductivity, so that the luminescence in these films was probably mediated by minority carrier injection or impact ionization in the high-field insulating regions of the films. Pankove and the RCA team offered a model for the operation of these devices based on impact ionization and Fowler–Nordheim tunneling, because the characteristics were virtually independent of temperature (Pankove and Lampert, 1974; Maruska *et al.*, 1974a, 1974b). Of course, these devices were inefficient, and as a consequence, Tietjen, who had stimulated the work, now terminated it by ordering “stop this garbage” – words that Maruska (2000) still vividly remembers.

1.6 History of blue, green, and white LEDs based on GaInN p-n junctions

After the research efforts of Pankove and co-workers had ended, work on GaN virtually ceased. In 1982 only a single paper was published on GaN. However, Isamu Akasaki and co-workers in Nagoya, Japan refused to give up, and in 1989 they demonstrated the first true p-type doping and p-type conductivity in GaN. The stubborn Mg acceptors were activated by ***electron-beam irradiation*** (Amano *et al.*, 1989). It was later shown that a high-temperature ***post-growth anneal*** of Mg-doped GaN also activates Mg dopants in GaN (Nakamura *et al.*, 1994a). Superlattice doping (Schubert *et al.*, 1996) further enhances the activation efficiency of deep acceptors. These p-type doping breakthroughs opened the door to efficient p-n junction LEDs and laser diodes. Today, Mg-doping of GaN is the basis for all nitride-based LEDs and laser diodes.

Subsequent to the attainment of p-type doping, the first GaN p-n-homojunction LED was reported by Akasaki *et al.* (1992). The LED that emitted light in the ultraviolet (UV) and blue spectral range, was grown on a sapphire substrate. The result was presented at the “GaAs and related compounds” Conference held in Karuizawa, Japan in 1992. The LED had an efficiency of approximately 1 %. This was a surprisingly high value for the highly dislocated GaN material grown on the mismatched sapphire substrate. It was also the first demonstration that nitride LED efficiencies are not affected by dislocations in the same adverse manner as III–V arsenide and phosphide light emitters.



Fig. 1.10. Array of GaInN/GaN blue LEDs manufactured by the Nichia Corporation (after Nakamura and Fasol, 1997).

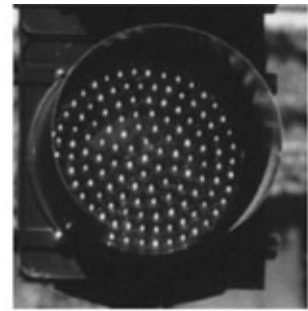
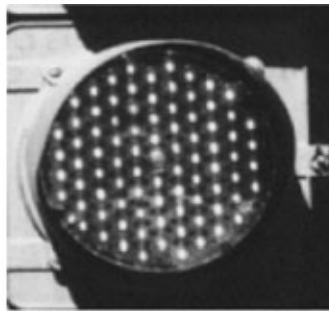
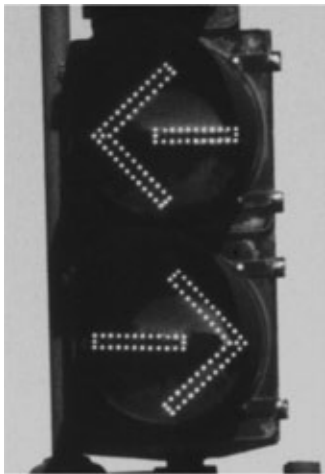


Fig. 1.11. Green traffic signals are one of the ubiquitous applications of GaInN/GaN green LEDs.

A name that is closely associated with GaN LEDs and lasers is the Nichia Chemical Industries Corporation, Japan. A team of researchers that included Shuji Nakamura has made numerous contributions to the development of GaN growth, LEDs, and lasers. Their contributions included the development of a two-flow organometallic vapor-phase epitaxy (OMVPE) growth system, the demonstration of the first viable blue and green GaInN double-heterostructure LEDs (Nakamura *et al.* 1993a, 1993b, 1994b) that achieved efficiencies of 10 % (Nakamura *et al.*, 1995), and the demonstration of the first pulsed and cw GaInN/GaN current injection **blue laser** operating at room temperature (Nakamura *et al.*, 1996). A detailed account of the team's contributions was given by Nakamura and Fasol (1997) in the book "*The blue laser diode*".

Blue LEDs made by the Nichia Corporation are shown in Fig. 1.10. A common application of high-brightness GaInN green LEDs is traffic signals as shown in Fig. 1.11. The earlier

mentioned GaP:N green LEDs are not suited for this application due to their much lower brightness.

In 1990, when Nakamura entered the field of GaN devices while working for the Nichia Corporation, he was a 36-year-old engineer without a Ph.D., not a single publication, and no conference contribution (Nakamura and Fasol, 1997). At the end of the 1990s, he had become a Professor at the University of California in Santa Barbara and a consultant for the Cree Lighting Corporation, a fierce competitor of Nichia. In 2001, he strongly criticized the Nichia Corporation and Japanese society. In the book entitled “*Breakthrough with anger*”, Nakamura (2001) stated “There is something wrong with this country. Industry and universities are terribly sick.”

The GaInN material system is also suited for **white LEDs**. There are different approaches to white LEDs, including white LEDs based on phosphor wavelength converters (see, for example, Nakamura and Fasol, 1997) and on semiconductor wavelength converters (Guo *et al.*, 1999). Much progress is expected in the area of white LEDs, since they have the potential to deliver a substantially higher luminous efficiency compared with conventional incandescent and fluorescent light sources. Whereas conventional light sources have typical (demonstrated) luminous efficiencies of 15 – 100 lm/W, white LEDs have the potential for luminous efficiencies exceeding 300 lm/W.

1.7 History of AlGaInP visible-spectrum LEDs

The AlGaInP material system is suited for high-brightness emission in the red (625 nm), orange (610 nm) and yellow (590 nm) spectral range and today is the dominant material system for high-brightness emitters in that wavelength range. Figure 1.12 shows some of the common signage applications of red and yellow AlGaInP LEDs.

The AlGaInP material system was first developed in Japan for visible lasers (Kobayashi *et al.*, 1985; Ohba *et al.*, 1986; Ikeda *et al.*, 1986; Itaya *et al.*, 1990). Efforts started with AlGaInP/GaInP double-heterostructure lasers using $\text{Ga}_{0.5}\text{In}_{0.5}\text{P}$ as the active material, which is lattice matched to GaAs substrates. The bandgap energy of lattice-matched GaInP is approximately 1.9 eV (650 nm), making the material suitable for visible lasers emitting in the red. These lasers are used, for example, in laser pointers and in digital video disc (DVD) players.

The addition of Al to the GaInP active region allows one to attain shorter emission wavelengths including the orange and yellow spectral region. However, $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ becomes an indirect semiconductor at Al compositions of $x \approx 0.53$, so that the radiative efficiency strongly decreases at wavelengths near and, in particular, below 600 nm.

Consequently, AlGaInP is not suited for high-efficiency emission at wavelengths below 570 nm.

Subsequent to the AlGaInP laser development that occurred in the early 1980s, AlGaInP LED development started at the end of the 1980s (Kuo *et al.* 1990; Fletcher *et al.*, 1991; Sugawara *et al.* 1991). In contrast to the AlGaInP laser structures, the LED structures typically employ current-spreading layers so that the entire p-n junction plane of the LED chip lights up and not just the region below the top ohmic contact. Further improvements were attained by using multiple quantum well (MQW) active regions (Huang and Chen, 1997), coherently strained MQW active regions (Chang and Chang, 1998a, 1998b), distributed Bragg reflectors (Huang and Chen, 1997; Chang *et al.*, 1997), and transparent GaP substrate technology (Kish and Fletcher, 1997). Comprehensive reviews of the AlGaInP material system and AlGaInP LEDs can be found in the literature (Stringfellow and Craford, 1997; Mueller, 2000; Krames *et al.*, 2002).



Fig. 1.12. Examples of red and amber AlGaInP/GaAs LEDs used in signage applications.

References

- Akasaki I., Amano H., Itoh K., Koide N., and Manabe K. "GaN based UV/blue light-emitting devices" GaAs and Related Compounds conference, *Inst. Phys. Conf. Ser.* **129**, 851 (1992)
- Allen J. W., Moncaster M. E., and Starkiewicz J. "Electroluminescent devices using carrier injection in gallium phosphide" *Solid State Electronics* **6**, 95 (1963)
- Amano H., Kito M., Hiramatsu K., Akasaki I. "P-type conduction in Mg-doped GaN treated with low-energy electron beam irradiation (LEEBI)" *Jpn. J. Appl. Phys.* **28**, L2112 (Dec. 1989)

- Chang S. J. and Chang C. S. “AlGaInP–GaInP compressively strained multiquantum well light-emitting diodes for polymer fiber applications” *IEEE Photonics Technol. Lett.* **10**, 772 (1998a)
- Chang S. J. and Chang C. S. “650 nm AlGaInP/GaInP compressively strained multi-quantum well light-emitting diodes” *Jpn. J. Appl. Phys* **37**, L653 (1998b)
- Chang S. J., Chang C. S., Su Y. K., Chang P. T., Wu Y. R., Huang K. H. and Chen T. P. “AlGaInP multiquantum well light-emitting diodes” *IEE Proceedings – Optoelectronics* **144**, 1 (1997)
- Craford M. G., Shaw R. W., Herzog A. H., and Groves W. O. “Radiative recombination mechanisms in GaAsP diodes with and without nitrogen doping” *J. Appl. Phys.* **43**, 4075 (1972)
- Destriau G. “Scintillations of zinc sulfides with alpha-rays” *J. Chimie Physique* **33**, 587 (1936)
- Duke C. B. and Holonyak Jr. N. “Advances in light-emitting diodes” *Physics Today*, December issue, p. 23 (1973)
- Edmond J. A., Kong H. S., and Carter Jr. C. H. “Blue LEDs, UV photodiodes and high-temperature rectifiers in 6 H-SiC” *Physica B* **185**, 453 (1993)
- Fletcher R. M., Kuo C., Osentowski T. D., and Robbins V. M. “Light-emitting diode with an electrically conductive window” U.S. Patent 5,008,718 (1991)
- Foster L. M. and Pilkuhn M. “Electroluminescence near bandgap in GaP containing shallow donor and acceptor levels” *Appl. Phys. Lett.* **7**, 65 (1965)
- Grimmeiss H. G. and Scholz H. J. “Efficiency of recombination radiation in GaP” *Phys. Lett.* **8**, 233 (1964)
- Groves W. O., Herzog A. H., and Craford M. G. “The effect of nitrogen doping on GaAs_{1-x}P_x electroluminescent diodes” *Appl. Phys. Lett.* **19**, 184 (1971)
- Groves W. O. and Epstein A. S. “Epitaxial deposition of III–V compounds containing isoelectronic impurities” US Patent 4,001,056 issued Jan. 4 (1977)
- Groves W. O., Herzog A. H., and Craford M. G. “Process for the preparation of electroluminescent III–V materials containing isoelectronic impurities” US Patent Re. 29,648 issued May 30 (1978a)
- Groves W. O., Herzog A. H., and Craford M. G. “GaAsP electroluminescent device doped with isoelectronic impurities” US Patent Re. 29,845 issued Nov. 21 (1978b)
- Guo X., Graff J. W., and Schubert E. F. “Photon recycling semiconductor light-emitting diode” *IEDM Technical Digest*, **IEDM-99**, 600 (Dec. 1999)

- Hall R. N., Fenner G. E., Kingsley J. D., Soltys T. J., and Carlson R. O. “Coherent light emission from GaAs junctions” *Phys. Rev. Lett.* **9**, 366 (1962)
- Herzog A. H., Groves W. O., and Craford M. G. “Electroluminescence of diffused GaAs_{1-x}P_x diodes with low donor concentrations” *J. Appl. Phys.* **40**, 1830 (1969)
- Holonyak Jr. N. “Active region in visible-light diode laser” *Electronics* **36**, 35 (1963)
- Holonyak Jr. N. “Laser action in Ga(AsP) and GaAs” *Proc. IEEE* **52**, 104 (1964)
- Holonyak Jr. N., personal communication (2000)
- Holonyak Jr. N. and Bevacqua S. F. “Coherent (visible) light emission from Ga(As_{1-x}P_x) junctions” *Appl. Phys. Lett.* **1**, 82 (1962)
- Holonyak Jr. N., Bevacqua S. F., Bielan C. V., Carranti F. A., Hess B. G., and Lubowski S. J. “Electrical properties of Ga(AsP) p-n junctions” *Proc. IEEE* **51**, 364 (1963a)
- Holonyak Jr. N., Bevacqua S. F., Bielan C. V., and Lubowski S. J. “The ‘direct–indirect’ transition in Ga(As_{1-x}P_x) p-n junctions” *Appl. Phys. Lett.* **3**, 47 (1963b)
- Holonyak Jr. N., Nuese C. J., Sirkis M. D., and Stillman G. E. “Effect of donor impurities on the direct–indirect transition in Ga(AsP)” *Appl. Phys. Lett.* **8**, 83 (1966)
- Huang K.-H. and Chen T.-P. “Light-emitting diode structure” U.S. Patent 5,661,742 (1997)
- Ikeda M., Nakano K., Mori Y., Kaneko K. and Watanabe N. “MOCVD growth of AlGaInP at atmospheric pressure using triethylmetals and phosphine” *J. Crystal Growth* **77**, 380 (1986)
- Isihamatsu S. and Okuno Y. “High efficiency GaAlAs LED” *Optoelectronics – Devices and Technologies* **4**, 21 (1989)
- Itaya K., Ishikawa M. and Uematsu Y. “636 nm room temperature cw operation by heterobarrier blocking structure InGaAlP laser diodes” *Electron. Lett.* **26**, 839 (1990)
- Juza R. and Hahn H. “On the crystal structure of Cu₃N, GaN and InN (translated from German)” *Zeitschrift fuer anorganische und allgemeine Chemie* **239**, 282 (1938)
- Kish F. A. and Fletcher R. M. “AlGaInP light-emitting diodes” in “High brightness light-emitting diodes” edited by G. B. Stringfellow and M. G. Craford *Semiconductor and Semimetals* Vol. **48**, p. 149 (Academic Press, San Diego, 1997)
- Kobayashi K., Kawata S., Gomyo A., Hino I. and Suzuki T. “Room-temperature cw operation of AlGaInP double-heterostructure visible lasers” *Electron. Lett.* **21**, 931 (1985)
- Krames M. R., Amano H., Brown J. J., and Heremans P. L. “High-efficiency light-emitting diodes” Special Issue of *IEEE J. Select. Top. Quant. Electronics* **8**, 185 (2002)

- Kuo C. P., Fletcher R. M., Osentowski T. D., Lardizabel M. C., Craford M. G., and Robbins V. M. “High performance AlGaInP visible light-emitting diodes” *Appl. Phys. Lett.* **57**, 2937 (1990)
- Loebner E. E. “Subhistories of the light-emitting diode” *IEEE Trans. Electron Devices* **ED-23**, 675 (1976)
- Logan R. A., personal communication (2000)
- Logan R. A., White H. G., and Trumbore F. A. “P-n junctions in compensated solution grown GaP” *J. Appl. Phys.* **38**, 2500 (1967a)
- Logan R. A., White H. G., and Trumbore F. A. “P-n junctions in GaP with external electroluminescence efficiencies $\sim 2\%$ at 25 °C” *Appl. Phys. Lett.* **10**, 206 (1967b)
- Logan R. A., White H. G., Wiegmann W. “Efficient green electroluminescence in nitrogen-doped GaP p-n junctions” *Appl. Phys. Lett.* **13**, 139 (1968)
- Logan R. A., White H. G., and Wiegmann W. “Efficient green electroluminescent junctions in GaP” *Solid State Electronics* **14**, 55 (1971)
- Lorenz M. R. and Binkowski B. B. “Preparation, stability, and luminescence of gallium nitride” *J. Electrochem. Soc.* **109**, 24 (1962)
- Lossev O. V. “Luminous carborundum detector and detection effect and oscillations with crystals” *Philosophical Magazine* **6**, 1024 (1928)
- Maruska H. P., personal communication. The photograph of a GaN MIS LED is gratefully acknowledged (2000)
- Maruska H. P. and Pankove J. I. “Efficiency of $\text{GaAs}_{1-x}\text{P}_x$ electroluminescent diodes” *Solid State Electronics* **10**, 917 (1967)
- Maruska H. P. and Tietjen J. J. “The preparation and properties of vapour-deposited single-crystal-line GaN” *Appl. Phys. Lett.* **15**, 327 (1969)
- Maruska H. P., Rhines W. C., Stevenson D. A. “Preparation of Mg-doped GaN diodes exhibiting violet electroluminescence” *Mat. Res. Bull.* **7**, 777 (1972)
- Maruska H. P., Stevenson D. A., Pankove J. I. “Violet luminescence of Mg-doped GaN (light-emitting diode properties)” *Appl. Phys. Lett.* **22**, 303 (1973)
- Maruska H. P., Anderson L. J., Stevenson D. A. “Microstructural observations on gallium nitride light-emitting diodes” *J. Electrochem. Soc.* **121**, 1202 (1974a)
- Maruska H. P. and Stevenson D. A. “Mechanism of light production in metal-insulator-semiconductor diodes; GaN:Mg violet light-emitting diodes” *Solid State Electronics* **17**, 1171 (1974b)

- Mueller G. (Editor) "Electroluminescence I" *Semiconductors and Semimetals* Vol. **64** (Academic Press, San Diego, 2000)
- Nakamura S. "Breakthrough with anger" (Shueisha, Tokyo, 2001). See also *Compound Semiconductors* Vol. **7**, No. 7, 25 (Aug. 2001) and Vol. **7**, No. 9, 15 (Oct. 2001)
- Nakamura S. and Fasol G. "The blue laser diode" (Springer, Berlin, 1997)
- Nakamura S., Senoh M., and Mukai T. "P-GaN/n-InGaN/n-GaN double-heterostructure blue-light-emitting diodes" *Jpn. J. Appl. Phys.* **32**, L8 (1993a)
- Nakamura S., Senoh M., and Mukai T. "High-power InGaN/GaN double-heterostructure violet light-emitting diodes" *Appl. Phys. Lett.* **62**, 2390 (1993b)
- Nakamura S., Iwasa N., and Senoh M. "Method of manufacturing p-type compound semiconductor" US patent 5,306,662 (1994a)
- Nakamura S., Mukai T. and Senoh M. "Candela-class high-brightness InGaN/AlGaIn double-heterostructure blue-light-emitting diodes" *Appl. Phys. Lett.* **64**, 1687 (1994b)
- Nakamura S., Senoh M., Iwasa N., Nagahama S. "High-brightness InGaIn blue, green, and yellow light-emitting diodes with quantum well structures" *Jpn. J. Appl. Phys.* **34**, L797 (1995)
- Nakamura S., Senoh M., Nagahama S., Iwasa N., Yamada T., Matsushita T., Sugimoto Y., and Kiyoku H. "Room-temperature continuous-wave operation of InGaIn multi-quantum-well structure laser diodes" *Appl. Phys. Lett.* **69**, 4056 (1996)
- Nathan M. I., Dumke W. P., Burns G., Dill Jr. F. H., and Lasher G. J. "Stimulated emission of radiation from GaAs p-n junctions" *Appl. Phys. Lett.* **1**, 62 (1962)
- Nuese C. J., Stillman G. E., Sirkis M. D., and Holonyak Jr. N., "Gallium arsenide-phosphide: crystal, diffusion, and laser properties" *Solid State Electronics* **9**, 735 (1966)
- Nuese C. J., Tietjen J. J., Gannon J. J., and Gossenberger H. F. "Optimization of electroluminescent efficiencies for vapor-grown GaAsP diodes" *J. Electrochem. Soc.: Solid State Sci.* **116**, 248 (1969)
- Ohba Y., Ishikawa M., Sugawara H., Yamamoto T., and Nakanishi T. "Growth of high-quality InGaAlP epilayers by MOCVD using methyl metalorganics and their application to visible semiconductor lasers" *J. Cryst. Growth* **77**, 374 (1986)
- Pankove J. I. and Berkeyheiser J. E. "A light source modulated at microwave frequencies" *Proc. IRE*, **50**, 1976 (1962)
- Pankove J. I. and Massoulié M. J. "Injection luminescence from GaAs" *Bull. Am. Phys. Soc.* **7**, 88 (1962)

- Pankove J. I., Berkeyheiser J. E., Maruska H. P., and Wittke J. "Luminescent properties of GaN" *Solid State Commun.* **8** 1051 (1970a)
- Pankove J. I., Maruska H. P., and Berkeyheiser J. E. "Optical absorption of GaN" *Appl. Phys. Lett.* **17**, 197 (1970b)
- Pankove J. I., Miller E. A., Richman D., and Berkeyheiser J. E. "Electroluminescence in GaN" *J. Luminescence* **4**, 63 (1971a)
- Pankove J. I., Miller E. A., and Berkeyheiser J. E. "GaN electroluminescent diodes" *RCA Review* **32**, 383 (1971b)
- Pankove J. I., Miller E. A., and Berkeyheiser J. E. "GaN blue light-emitting diodes" *J. Luminescence* **5**, 84 (1972)
- Pankove J. I. and Lampert M. A. "Model for electroluminescence in GaN" *Phys. Rev. Lett.* **33**, 361 (1974)
- Perry T. S. "M. George Craford" *IEEE Spectrum*, February issue, p. 52 (1995)
- Pilkuhn M. H., personal communication. The photograph of GaP:Zn-O LED is gratefully acknowledged (2000)
- Pilkuhn M. H. "Light-emitting diodes" in "Handbook of semiconductors" edited by T. S. Moss, Vol. **4**, edited by C. Hilsum, p. 539 (1981)
- Pilkuhn M. H. and Rupprecht H. "Light emission from GaAs_xP_{1-x} diodes" *Trans. Metallurgical Soc. AIME* **230**, 282 (1964)
- Pilkuhn M. H. and Rupprecht H. "Electroluminescence and lasing action in GaAs_xP_{1-x}" *J. Appl. Phys.* **36**, 684 (1965)
- Potter R. M., Blank J. M., Addamiano A. "Silicon carbide light-emitting diodes" *J. Appl. Phys.* **40**, 2253 (1969)
- Quist T. M., Rediker R. H., Keyes R. J., Krag W. E., Lax B., McWhorter A. L. and Zeigler H. J. "Semiconductor maser of GaAs" *Appl. Phys. Lett.* **1**, 91 (1962)
- Rostky G. "LEDs cast Monsanto in unfamiliar role" *Electronic Engineering Times (EETimes)* on the internet, see < <http://eetimes.com/anniversary/designclassics/monsanto.html> > Issue 944, Monday, March 10 (1997)
- Round H. J. "A note on carborundum" *Electrical World*, **19**, 309 (1907)
- Rupprecht H., personal communication (2000)
- Rupprecht H., Pilkuhn M., and Woodall J. M. "Continuous stimulated emission from GaAs diodes at 77 K" (1st report of 77 K cw laser) *Proc. IEEE* **51**, 1243 (1963)

- Rupprecht H., Woodall J. M., Konnerth K., and Pettit D. G. "Efficient electro-luminescence from GaAs diodes at 300 K" *Appl. Phys. Lett.* **9**, 221 (1966)
- Rupprecht H., Woodall J. M., and Pettit G. D. "Efficient visible electroluminescence at 300 K from AlGaAs pn junctions grown by liquid phase epitaxy" *Appl. Phys. Lett.* **11**, 81 (1967).
- Rupprecht H., Woodall J. M., Pettit G. D., Crowe J. W., and Quinn H. F. "Stimulated emission from AlGaAs diodes at 77 K" *Quant. Elect.* **4**, 35 (1968)
- Schubert E. F., Wang Y.-H., Cho A. Y., Tu L.-W., and Zydzik G. J. "Resonant cavity light-emitting diode" *Appl. Phys. Lett.* **60**, 921 (1992)
- Schubert E. F., Hunt N. E. J., Micovic M., Malik R. J., Sivco D. L., Cho A. Y., and Zydzik G. J. "Highly efficient light-emitting diodes with microcavities" *Science* **265**, 943 (1994)
- Schubert E. F., Grieshaber W., and Goepfert I. D. "Enhancement of deep acceptor activation in semiconductors by superlattice doping" *Appl. Phys. Lett.* **69**, 3737 (1996)
- Stringfellow G. B. and Craford M. G. (Editors) "High brightness light-emitting diodes" *Semiconductors and Semimetals* Vol. **48** (Academic Press, San Diego, 1997)
- Sugawara H., Ishikawa M., Kokubun Y., Nishikawa Y., and Naritsuka S. "Semiconductor light-emitting device" U.S. Patent 5,048,035 (1991)
- Thomas D. G., Hopfield J. J., and Frosch C. J. "Isoelectronic traps due to nitrogen in gallium phosphide" *Phys. Rev. Lett.* **15**, 857 (1965)
- Violin E. E., Kalnin A. A., Pasynkov V. V., Tairov Y. M., Yaskov D. A. "Silicon Carbide – 1968" *2nd International Conference on Silicon Carbide*, published as a special issue of the *Materials Research Bulletin* page 231 (1969)
- Welker H. "On new semiconducting compounds (translated from German)" *Zeitschrift fuer Naturforschung* **7a**, 744 (1952)
- Welker H. "On new semiconducting compounds II (translated from German)" *Zeitschrift fuer Naturforschung* **8a**, 248 (1953)
- Wolfe C. M., Nuese C. J., and Holonyak Jr. N. "Growth and dislocation structure of single-crystal Ga(AsP)" *J. Appl. Phys.* **36**, 3790 (1965)
- Woodall J. M., personal communication (2000)
- Woodall J. M., Potemski R. M., Blum S. E., and Lynch R. "AlGaAs LED structures grown on GaP substrates," *Appl. Phys. Lett.* **20**, 375 (1972)

2

Radiative and non-radiative recombination

Electrons and holes in semiconductors recombine either radiatively, i.e. accompanied by the emission of a photon, or non-radiatively. In light-emitting devices, the former is clearly the preferred process. However, non-radiative recombination can, under practical conditions, never be reduced to zero. Thus, there is competition between radiative and non-radiative recombination. Maximization of the radiative process and minimization of the non-radiative process can be attained in a number of ways which will be discussed below.

2.1 Radiative electron–hole recombination

Any undoped or doped semiconductor has two types of free carriers, electrons and holes. Under equilibrium conditions, i.e. without external stimuli such as light or current, the law of mass action states that the product of the electron and hole concentrations is, at a given temperature, a constant, i.e.

$$\boxed{n_0 p_0 = n_i^2} \quad (2.1)$$

where n_0 and p_0 are the equilibrium electron and hole concentrations and n_i is the intrinsic carrier concentration. The validity of the law of mass action is limited to non-degenerately doped semiconductors (see, for example, Schubert, 1993).

Excess carriers in semiconductors can be generated by either absorption of light or by an injection current. The total carrier concentration is then given by the sum of equilibrium and excess carrier concentrations, i.e.

$$n = n_0 + \Delta n \quad \text{and} \quad p = p_0 + \Delta p \quad (2.2)$$

where Δn and Δp are the excess electron and hole concentration, respectively.

Next, we consider recombination of carriers. The band diagram of a semiconductor with electrons and holes is shown in **Fig. 2.1**. We are interested in the rate at which the carrier concentration decreases and denote the recombination rate as R . Consider a free electron in the conduction band. The probability that the electron recombines with a hole is proportional to the hole concentration, that is, $R \propto p$. The number of recombination events will also be proportional to the concentration of electrons, as indicated in **Fig. 2.1**. Thus the recombination rate is proportional to the product of electron and hole concentrations, that is, $R \propto n p$. Using a proportionality constant, the recombination rate per unit time per unit volume can be written as

$$R = -\frac{dn}{dt} = -\frac{dp}{dt} = B n p \quad (2.3)$$

This equation is the **bimolecular rate equation** and the proportionality constant B is called the **bimolecular recombination coefficient**. It has typical values of $10^{-11} - 10^{-9} \text{ cm}^3/\text{s}$ for III–V semiconductors. The bimolecular recombination coefficient will be calculated in a subsequent section using the van Roosbroeck–Shockley model.

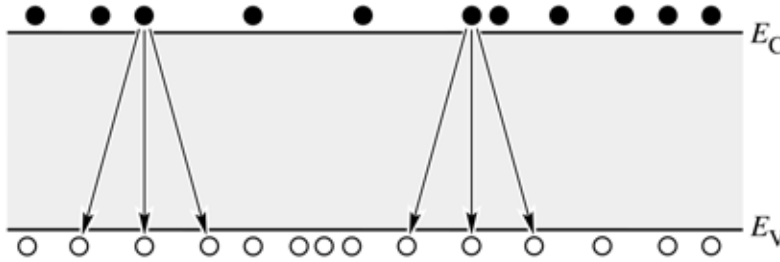


Fig. 2.1. Illustration of electron-hole recombination. The number of recombination events per unit time per unit volume is proportional to the product of electron and hole concentrations, *i. e.* $R \propto n p$.

2.2 Radiative recombination for low-level excitation

Next, we discuss the recombination dynamics as a function of time. Consider a semiconductor subjected to photoexcitation. The equilibrium and excess electron and hole concentrations are n_0 , p_0 , Δn , and Δp , respectively. Since electrons and holes are generated and annihilated (by recombination) in pairs, the steady-state electron and hole excess concentrations are equal,

$$\Delta n(t) = \Delta p(t) . \quad (2.4)$$

Using the bimolecular rate equation, the recombination rate is given by

$$R = B [n_0 + \Delta n(t)] [p_0 + \Delta p(t)] . \quad (2.5)$$

For the case of **low-level excitation**, the photogenerated carrier concentration is much smaller than the majority carrier concentration, i.e. $\Delta n \ll (n_0 + p_0)$. Using $\Delta n(t) = \Delta p(t)$ one obtains from Eq. (2.5)

$$\begin{aligned} R &= B n_i^2 + B (n_0 + p_0) \Delta n(t) \\ &= R_0 + R_{\text{excess}} . \end{aligned} \quad (2.6)$$

The first summand on the right-hand side of the equation can be identified as the *equilibrium recombination rate* and the second term as the *excess recombination rate*.

The time-dependent carrier concentration can be calculated from the rate equation

$$\frac{dn(t)}{dt} = G - R = (G_0 + G_{\text{excess}}) - (R_0 + R_{\text{excess}}) \quad (2.7)$$

where G_0 and R_0 are the equilibrium generation and recombination rates, respectively.

Next, we assume that the semiconductor has been illuminated with light and excess carriers are generated. At the time $t = 0$, the illumination is switched off (i.e. $G_{\text{excess}} = 0$) as indicated in Fig. 2.2. The recombination rate can then be calculated by insertion of Eq. (2.6) into Eq. (2.7) and using $G_0 = R_0$. This yields the differential equation

$$\frac{d}{dt} \Delta n(t) = -B (n_0 + p_0) \Delta n(t) . \quad (2.8)$$

The solution of the differential equation can be obtained by separation of variables. One obtains

$$\boxed{\Delta n(t) = \Delta n_0 e^{-B(n_0 + p_0)t}} \quad (2.9)$$

where $\Delta n_0 = \Delta n(t = 0)$. Rewriting the result as

$$\Delta n(t) = \Delta n_0 e^{-t/\tau} \quad (2.10)$$

allows one to identify the *carrier lifetime* τ as

$$\tau = \left[B (n_0 + p_0) \right]^{-1} \quad (2.11)$$

For semiconductors with a specific doping type, Eq. (2.11) reduces to

$$\tau_n = \frac{1}{B p_0} = \frac{1}{B N_A} \quad \text{for p-type semiconductors} \quad (2.12)$$

and

$$\tau_p = \frac{1}{B n_0} = \frac{1}{B N_D} \quad \text{for n-type semiconductors} \quad (2.13)$$

where τ_n and τ_p is the electron and hole lifetime, respectively. Using this result, the rate equation, Eq. (2.8), can be simplified for semiconductors of a specific conductivity type. One obtains the *monomolecular rate equations*:

$$\frac{d}{dt} \Delta n(t) = - \frac{\Delta n(t)}{\tau_n} \quad \text{for p-type semiconductors} \quad (2.14)$$

and

$$\frac{d}{dt} \Delta p(t) = - \frac{\Delta p(t)}{\tau_p} \quad \text{for n-type semiconductors.} \quad (2.15)$$

Figure 2.2 shows the majority and minority carrier concentration in a p-type semiconductor as a function of time (similar considerations apply if an n-type semiconductor were chosen). Note that Fig. 2.2 shows the case of *low-level excitation* in which the photogenerated carrier concentration is much smaller than the majority carrier concentration. However, the photogenerated carrier concentration is much larger than the minority carrier concentration.

Once photoexcitation is terminated, the minority carrier concentration decays exponentially with a characteristic time constant denoted as the *minority carrier lifetime* τ . It is the mean time between generation and recombination of a minority carrier.

Note that the majority carrier concentration also decays with the same time constant τ . However, only a very *small fraction* of the majority carriers disappear by recombination, as illustrated in Fig. 2.2. For low-level excitation, the average time it takes for a majority carrier to

recombine is much longer than the minority carrier lifetime. For many practical purposes, the majority carrier lifetime can be assumed to be infinitely long.

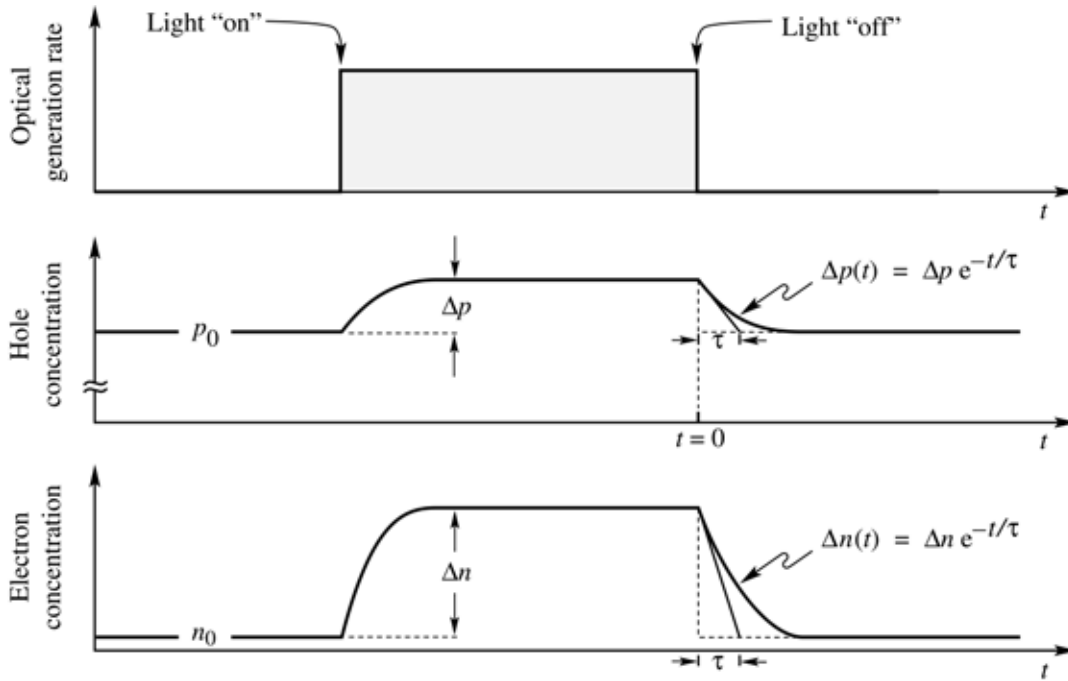


Fig. 2.2. Carrier concentration as a function of time before, during, and after an optical excitation pulse. The semiconductor is assumed to be p-type and thus it is $p_0 \gg n_0$. Electrons and holes are generated in pairs, thus $\Delta p = \Delta n$. Under low-level excitation as shown here, it is $\Delta n \ll p_0$. In most practical cases the equilibrium minority carrier concentration is extremely small so that $n_0 \ll \Delta n$.

Theoretical and experimental values for the minority carrier lifetime in GaAs as a function of the doping concentration are shown in Fig. 2.3. The theoretical result is calculated from Eq. (2.10) using $B = 10^{-10} \text{ cm}^3/\text{s}$. In nominally undoped material, minority carrier lifetimes as long as $15 \mu\text{s}$ have been measured in GaAs at room temperature (Nelson and Sobers, 1978a, 1978b).

Exercise. Minority carrier lifetimes. Calculate the minority carrier lifetime in p-type GaAs at a doping concentration of 10^{15} and 10^{18} cm^{-3} using a bimolecular recombination coefficient of $B = 10^{-10} \text{ cm}^3/\text{s}$. Assume that one could fabricate GaAs without any impurities. What would be the carrier lifetime in intrinsic GaAs with a carrier concentration of $2 \times 10^6 \text{ cm}^{-3}$?

Solution: $\tau_n = 10 \mu\text{s}$ for $N_A = 10^{15} \text{ cm}^{-3}$

$$\begin{aligned}\tau_n &= 10 \text{ ns} && \text{for } N_A = 10^{18} \text{ cm}^{-3} \\ \tau &= 2500 \text{ s} && \text{for undoped GaAs.}\end{aligned}$$

Discuss how the modulation speed of communication LEDs is affected by the radiative lifetime and the doping concentration.

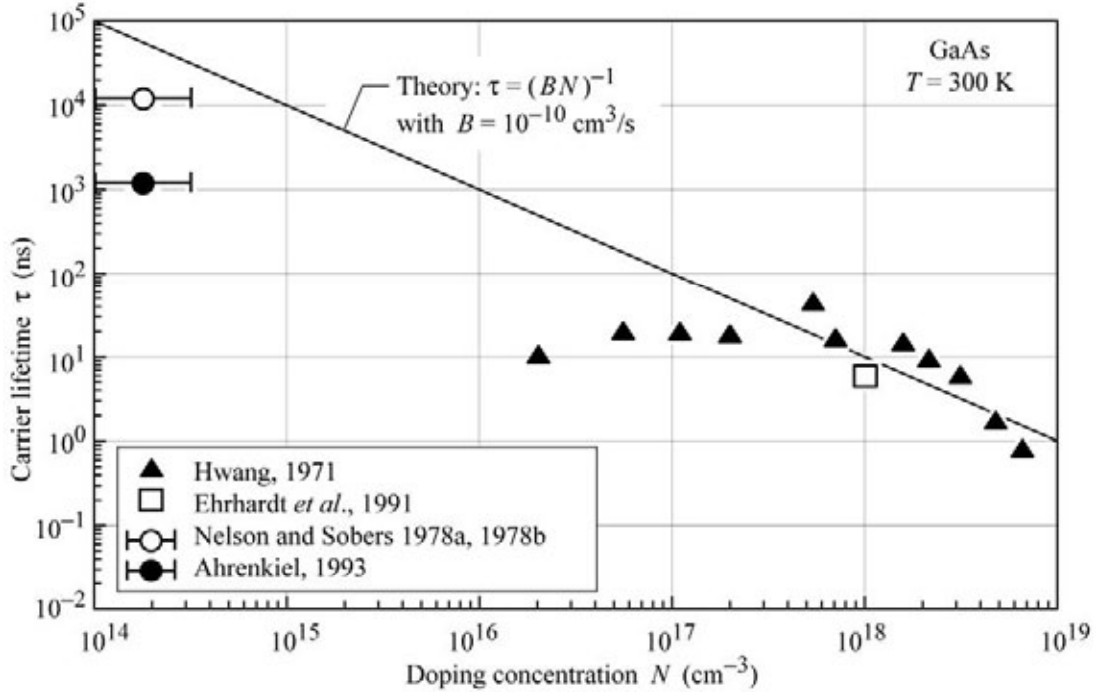


Fig. 2.3. Minority carrier lifetime as a function of doping concentration in GaAs at 300 K. The lifetime was inferred from luminescence decay measurements. The data points of Nelson and Sobers (1978a, 1978b) and Ahrenkiel (1993) were measured on nominally undoped material with a doping concentration $\ll 10^{15}$ cm⁻³.

2.3 Radiative recombination for high-level excitation

For the case of **high-level excitation**, the photogenerated carrier concentration is larger than the equilibrium carrier concentration, i.e. $\Delta n \gg (n_0 + p_0)$. The bimolecular rate equation (Eq. 2.3) is then given by

$$\frac{d\Delta n(t)}{dt} = -B \Delta n^2. \quad (2.16)$$

Solving this differential equation by the separation of variables method and using the initial condition $\Delta n(0) = \Delta n_0$ yields the solution

$$\Delta n(t) = \frac{1}{Bt + \Delta n_0^{-1}} \quad (2.17)$$

This solution represents, in contrast to the low-density approximation, a *non-exponential* carrier decay.

In an exponential decay, it takes the time constant τ for the carrier concentration to decrease from Δn_0 to $\Delta n_0 e^{-1}$. Using the same definition for the non-exponential decay given by Eq. (2.17), the “time constant” can be calculated from the slope of the decay by using the equation

$$\tau(t) = - \frac{\Delta n(t)}{\frac{d\Delta n(t)}{dt}} . \quad (2.18)$$

Using this definition for the non-exponential decay of Eq. (2.17), one obtains the “time constant”

$$\tau(t) = t + \frac{1}{B \Delta n_0} . \quad (2.19)$$

Thus, for non-exponential decays the “time constant” depends on time. Equation (2.19) shows that the minority carrier lifetime increases with time. For sufficiently long times, low-level excitation conditions will be reached and τ will approach the low-level value.

2.4 Bimolecular rate equations for quantum well structures

Quantum wells provide a means of confining the free carriers to a narrow quantum well region by means of the two barrier regions cladding the quantum well. Assume that the well region has a thickness of L_{QW} . Assume further that the conduction band and valence band wells have carrier densities of n^{2D} and p^{2D} , respectively. The effective 3D carrier concentration for electrons and holes is then given by n^{2D} / L_{QW} and p^{2D} / L_{QW} , respectively. Using these values as the 3D carrier concentration, the recombination rate can be inferred from Eq. (2.5), and it is given by

$$R = B \frac{n^{2D}}{L_{QW}} \frac{p^{2D}}{L_{QW}} . \quad (2.20)$$

This equation illustrates one of the essential advantages of quantum well and double heterostructures. A decrease of the quantum well thickness allows one to attain high three-dimensional (3D) carrier concentrations (carriers per cm^{-3}). As a result, the carrier lifetime for radiative recombination is reduced, as inferred from Eq. (2.11), and the radiative efficiency is increased.

For sufficiently small quantum well thicknesses, the wave function no longer scales with the physical well width. L_{QW} must be replaced by the carrier distribution width, which for sufficiently small well thicknesses is larger than L_{QW} , since the wave function will extend into the barriers. This effect should be considered for well thicknesses $< 100 \text{ \AA}$ in the AlGaAs/GaAs material system.

2.5 Luminescence decay

The carrier decay in semiconductors can be measured by the decay of the luminescence after a short optical excitation pulse. The luminescence intensity is proportional to the recombination rate. Calculating the recombination rate for the low and high excitation cases (Eqs. 2.9 and 2.17), one obtains

$$R = -\frac{dn(t)}{dt} = \frac{\Delta n_0}{\tau} e^{-t/\tau} \quad \text{for low excitation} \quad (2.21)$$

and

$$R = -\frac{dn(t)}{dt} = \frac{-B}{\left(Bt + \Delta n_0^{-1}\right)^2} \quad \text{for high excitation} . \quad (2.22)$$

Figure 2.4 illustrates schematically the decay of the luminescence after optical excitation by a short pulse. For the case of low excitation density, the decay is exponential with a time constant τ . For the case of high excitation, the decay is non-exponential. All non-exponential decay functions can be expressed by an exponential function *with a time-dependent time constant*, i.e. $\exp[-t/\tau(t)]$. The time constant increases with time. This type of decay function is frequently called a ***stretched exponential function***.

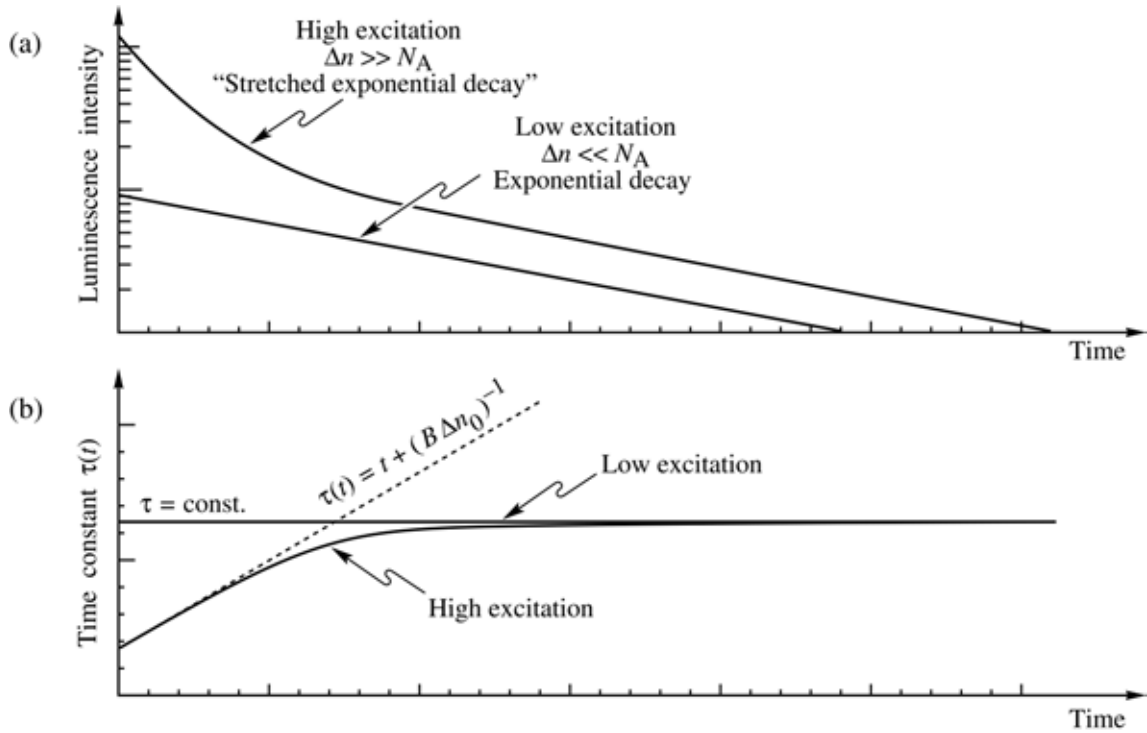


Fig. 2.4. (a) Luminescence decay for low and high excitation densities. (b) Time constants for low and high excitation densities.

The recombination dynamics of carriers in LEDs is one of the factors that limits the time it takes to switch an LED on and off. The modulation speed of LEDs used for communication applications can be limited by the minority carrier lifetime. The carrier lifetime can be reduced by either a high doping of the active region or a high concentration of injected carriers in the active region. Heterostructures that confine free carriers to the small well region are frequently employed to obtain high carrier concentrations and thus short carrier lifetimes.

2.6 Non-radiative recombination in the bulk

There are two basic recombination mechanisms in semiconductors, namely *radiative* recombination and *non-radiative* recombination. In a radiative recombination event, one photon with energy equal to the bandgap energy of the semiconductor is emitted, as illustrated in Fig. 2.5. During non-radiative recombination, the electron energy is converted to vibrational energy of lattice atoms, i.e. phonons. Thus, the electron energy is converted to heat. For obvious reasons, non-radiative recombination events are unwanted in light-emitting devices.

There are several physical mechanisms by which non-radiative recombination can occur. The most common cause for non-radiative recombination events are defects in the crystal structure. These defects include unwanted foreign atoms, native defects, dislocations, and any complexes

of defects, foreign atoms, or dislocations. In compound semiconductors, native defects include interstitials, vacancies, and antisite defects (Longini and Greene, 1956; Baraff and Schluter, 1985). All such defects have energy level structures that are different from substitutional semiconductor atoms. It is quite common for such defects to form one or several energy levels within the forbidden gap of the semiconductor.

Energy levels within the gap of the semiconductor are efficient recombination centers; in particular, if the energy level is close to the middle of the gap. The recombination of carriers via a trap level is shown schematically in Fig. 2.6. Owing to the promotion of non-radiative processes, such deep levels or traps are called *luminescence killers*.

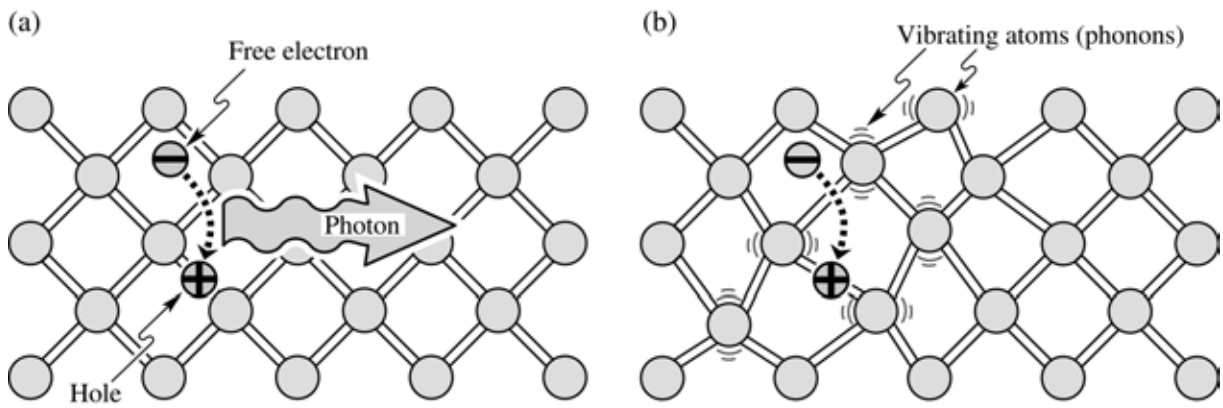


Fig. 2.5. (a) Radiative recombination of an electron–hole pair accompanied by the emission of a photon with energy $h\nu \approx E_g$. (b) In non-radiative recombination events, the energy released during the electron–hole recombination is converted to phonons (adopted from Shockley, 1950).

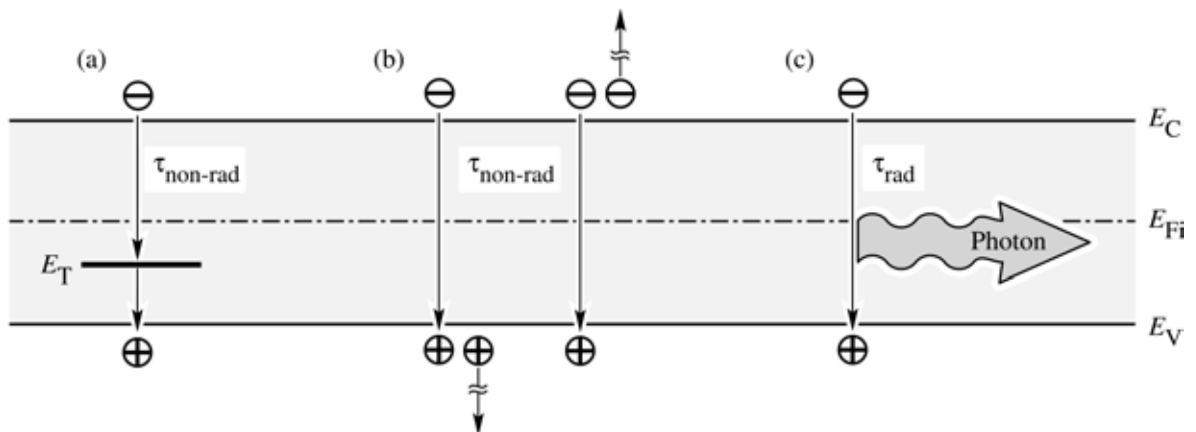


Fig. 2.6. Band diagram illustrating non-radiative recombination: (a) via a deep level, (b) via an Auger process and (c) radiative recombination.

The recombination of free carriers via deep levels was first analyzed by Shockley, Read, and Hall (Hall, 1952; Shockley and Read, 1952). The non-radiative recombination rate through a deep level with trap energy E_T and concentration N_T is given by (Shockley and Read, 1952)

$$R_{SR} = \frac{p_0 \Delta n + n_0 \Delta p + \Delta n \Delta p}{(N_T v_p \sigma_p)^{-1} (n_0 + n_1 + \Delta n) + (N_T v_n \sigma_n)^{-1} (p_0 + p_1 + \Delta p)} \quad (2.23)$$

where $\Delta n = \Delta p$; v_n and v_p are the electron and hole thermal velocities, and σ_n and σ_p are the capture cross sections of the traps. The quantities n_1 and p_1 are the electron and hole concentrations if the Fermi energy is located at the trap level. These quantities are given by

$$n_1 = n_i \exp\left(\frac{E_T - E_{Fi}}{kT}\right) \quad \text{and} \quad p_1 = n_i \exp\left(\frac{E_{Fi} - E_T}{kT}\right). \quad (2.24)$$

where E_{Fi} is the Fermi level in the intrinsic semiconductor.

The non-radiative lifetime of excess electrons can be deduced from the equation $R_{SR} = \Delta n / \tau$. Consequently, the lifetime is given by

$$\frac{1}{\tau} = \frac{p_0 + n_0 + \Delta n}{(N_T v_p \sigma_p)^{-1} (n_0 + n_1 + \Delta n) + (N_T v_n \sigma_n)^{-1} (p_0 + p_1 + \Delta p)}. \quad (2.25)$$

We now differentiate between majority carriers and minority carriers and assume that the semiconductor is p-type. Then holes are in the majority, *i.e.* $p_0 \gg n_0$ and $p_0 \gg p_1$. If we further assume a small deviation from equilibrium *i.e.* $\Delta n \ll p_0$, then the minority carrier lifetime is given by

$$\frac{1}{\tau} = \frac{1}{\tau_{n_0}} = N_T v_n \sigma_n. \quad (2.26)$$

If electrons were the majority carriers, the lifetime would be obtained in an analogous way, *i.e.*

$$\frac{1}{\tau} = \frac{1}{\tau_{p_0}} = N_T v_p \sigma_p. \quad (2.27)$$

The results show that the Shockley–Read recombination rate is limited by the rate of capture of minority carriers. This result suggests itself since the capture of *majority* carriers is a much more *likely* event than the capture of *minority* carriers. Equation (2.25) can then be written as

$$\frac{1}{\tau} = \frac{p_0 + n_0 + \Delta n}{\tau_{p_0}(n_0 + n_1 + \Delta n) + \tau_{n_0}(p_0 + p_1 + \Delta p)}. \quad (2.28)$$

For small deviations from equilibrium, i.e. $\Delta n \ll p_0$, the equation simplifies to

$$\tau = \tau_{n_0} \frac{p_0 + p_1}{p_0 + n_0} + \tau_{p_0} \frac{n_0 + n_1 + \Delta n}{p_0 + n_0} \approx \tau_{n_0} \frac{p_0 + p_1}{p_0 + n_0}. \quad (2.29)$$

Inspection of the equation reveals that the lifetime does not change for small deviations from equilibrium in an extrinsic semiconductor.

For further insight, we assume that the trap captures electrons and holes at the same rate, i.e. $\nu_n \sigma_n = \nu_p \sigma_p$ and $\tau_{n_0} = \tau_{p_0}$. One obtains from Eq. (2.29)

$$\tau = \tau_{n_0} \left(1 + \frac{p_0 + p_1}{p_0 + n_0} \right). \quad (2.30)$$

For the special case of intrinsic material, i.e. $n_0 = p_0 = n_i$, the equation simplifies to

$$\tau_i = \tau_{n_0} \left(1 + \frac{p_1 + n_1}{2n_i} \right) = \tau_{n_0} \left[1 + \cosh \left(\frac{E_T - E_{Fi}}{kT} \right) \right] \quad (2.31)$$

where E_{Fi} is the intrinsic Fermi level, which is typically close to the middle of the gap. The cosh function has a minimum when the argument of the function is zero. Thus the non-radiative lifetime is minimized, if $E_T - E_{Fi}$ is zero; i.e. when the trap level is at or close to the midgap energy. For such midgap levels, the lifetime is given by $\tau = 2 \tau_{n_0}$. This result demonstrates that deep levels are effective recombination centers if they are near the middle of the gap.

Inspection of Eq. (2.31) also reveals the temperature dependence of Shockley–Read recombination. As T increases the non-radiative recombination lifetime *decreases*. As a result, the radiative band-to-band recombination efficiency *decreases* at high temperatures. The highest

band-to-band radiative efficiencies of direct-gap semiconductors can be obtained at cryogenic temperatures.

However, some devices are based on radiative recombination through a deep state. A well-known example of radiative recombination mediated by a deep level is N-doped GaP. It follows from the Shockley–Read model that the deep-level recombination rate increases with increasing temperature.

In *indirect-gap* semiconductors such as GaP, radiative transitions are mediated by phonons. That is, radiative recombination must be accompanied by absorption or emission of a phonon. Since phonons are more abundant at high temperatures, radiative recombination (mediated by the absorption of a phonon) can increase with temperature.

In the vicinity of a deep level, the luminescence intensity decreases. A single *point defect* will be difficult to observe due to its relatively small effect. Frequently, however, defects group into clusters of defects or *extended defects*. Such extended defects are, for example, threading dislocations and misfit dislocations occurring when epitaxial semiconductors are grown on mismatched substrates, i.e. substrates with a different lattice constant from that of the epitaxial layer. There are also many other types of extended defects. The luminescence-killing nature of extended defects is illustrated in Fig. 2.7, which shows a cathodoluminescence micrograph of a GaAs layer measured at room temperature. The figure reveals several dark spots. Luminescence in the vicinity of the defects is reduced due to the non-radiative recombination channels so that the defects manifest themselves as dark spots. The size of the dark spots depends on the size of the defect and the minority carrier diffusion length.

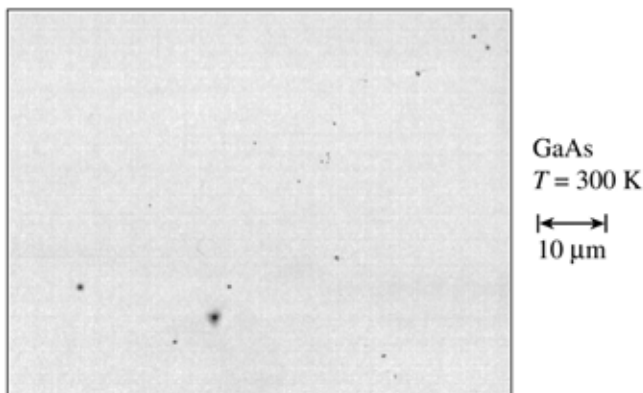


Fig. 2.7. Cathodoluminescence micrograph of a GaAs epitaxial layer. The dark spots are due to large clusters of non-radiative recombination centers (after Schubert, 1995).

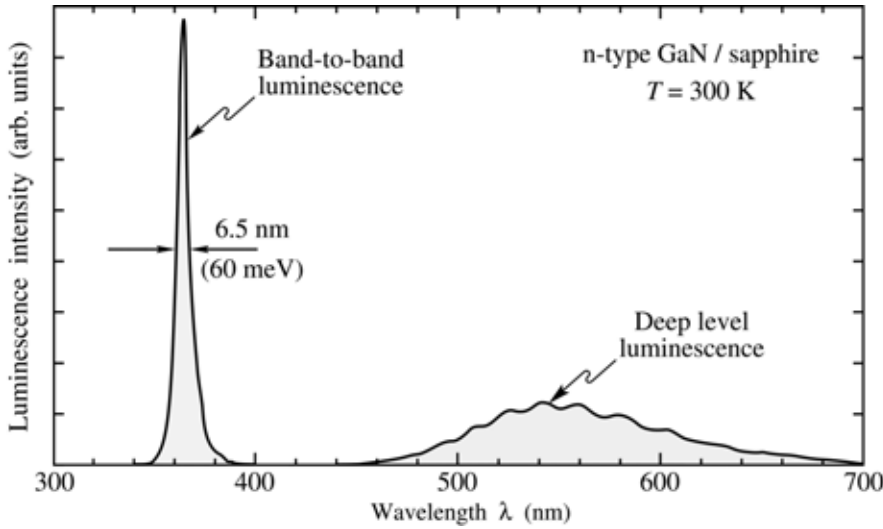


Fig. 2.8. Room-temperature photoluminescence spectrum of GaN with a band-to-band optical transition at 365 nm and a second transition at 550 nm, which was identified as an optically active deep-level transition (after Grieshaber *et al.*, 1996).

While most deep-level transitions are non-radiative, some deep-level transitions are radiative. An example of a radiative deep-level transition in GaN is shown in Fig. 2.8 (Grieshaber *et al.*, 1996). The luminescence spectrum shows the band-to-band transition at 365 nm and a broad deep-level transition around 550 nm. The deep-level transition occurs near the yellow range of the visible spectrum. The yellow luminescence line has been shown to be due to Ga vacancies (Neugebauer and Van de Walle, 1996; Schubert *et al.* 1997; Saarinen *et al.* 1997), a common point defect in n-type GaN.

Deep levels can be caused by **native defects** such as Ga vacancies, unwanted foreign impurities, dislocations, impurity–defect complexes, and combinations of different types of defects. Frequently it takes many years to unambiguously identify the atomic nature of a defect. For a review of defects in semiconductors see, for example, Pantelides (1992).

Another important non-radiative recombination mechanism is **Auger recombination**. In this process the energy becoming available through electron–hole recombination (approximately E_g), is dissipated by the excitation of a free electron high into the conduction band, or by a hole deeply excited into the valence band. The processes are shown schematically in Fig. 2.6 (b). The highly excited carriers will subsequently lose energy by multiple phonon emission until they are close to the band edge.

The recombination rates due to the two Auger processes shown in Fig. 2.6 (b) are given by

$$R_{\text{Auger}} = C_p n p^2 \quad (2.32)$$

and

$$R_{\text{Auger}} = C_n n^2 p. \quad (2.33)$$

Auger recombination is proportional to the square of the carrier concentration (either p^2 or n^2) since two carriers of the same type (either two holes or two electrons) are required for the recombination process. The first process (see Eq. 2.32) is more likely to happen in p-type semiconductors due to the abundance of holes. The second process (see Eq. 2.33) is more likely in n-type semiconductors due to the abundance of electrons.

During Auger recombination, energy and momentum must be conserved. Owing to the differences in conduction and valence band structure in semiconductors, the two Auger coefficients C_p and C_n are generally different.

In the high-excitation limit in which the non-equilibrium carriers have a higher concentration than equilibrium carriers, the Auger rate equations reduce to

$$R_{\text{Auger}} = (C_p + C_n) n^3 = C n^3 \quad (2.34)$$

where C is the **Auger coefficient**. Typical values for the Auger coefficient are $10^{-28} - 10^{-29} \text{ cm}^6/\text{s}$ for III–V semiconductors (see, for example, Olshansky *et al.*, 1984; Agrawal and Dutta, 1986).

Auger recombination reduces the luminescence efficiency in semiconductors only at very high excitation intensity or at very high carrier injection currents. This is due to the cubic carrier concentration dependence. At lower carrier concentrations, the Auger recombination rate is very small and can be neglected for practical purposes. The numerical values of Auger recombination can be determined by a quantum mechanical calculation that takes into account the band structure of the semiconductor (see, for example, Agrawal and Dutta, 1986).

2.7 Non-radiative recombination at surfaces

Substantial non-radiative recombination can occur at semiconductor surfaces. Surfaces are a strong perturbation of the periodicity of a crystal lattice. Recall that the band diagram model is based on the strict periodicity of a lattice. Since this periodicity ends at a surface, the band diagram will need to be modified at a semiconductor surface. This modification includes the addition of electronic states within the forbidden gap of the semiconductor.

Next, we consider a semiconductor surface from a chemical point of view. Atoms at the surface cannot have the same bonding structure as bulk atoms due to the lack of neighboring atoms. Thus some of the valence orbitals do not form a chemical bond. These partially filled electron orbitals, or **dangling bonds**, are electronic states that can be located in the forbidden gap

of the semiconductor where they act as recombination centers. Depending on the charge state of these valence orbitals, the states can be acceptor-like or donor-like states.

The dangling bonds may also rearrange themselves and form bonds between neighboring atoms in the same surface plane. This **surface reconstruction** can lead to a *locally* new atomic structure with state energies different from bulk atomic states. The surface bonding structure depends on the specific nature of the semiconductor surface. The energetic location of surface states is very difficult to predict, even with powerful theoretical models. Thus phenomenological models of surface recombination are commonly used.

It has been shown that electronic states within the forbidden gap appear at semiconductor surfaces. Bardeen and Shockley (Shockley, 1950) have pioneered the understanding of surface states and their role as recombination centers.

Next, we calculate the effect of surface recombination on the carrier distribution in a p-type semiconductor subjected to illumination. Assume that the illumination causes a uniform steady-state generation rate G . The one-dimensional continuity equation must be fulfilled at any point in the semiconductor. The continuity equation for electrons is given by

$$\frac{\partial \Delta n(x,t)}{\partial t} = G - R + \frac{1}{e} \frac{\partial}{\partial x} J_n \quad (2.35)$$

where J_n is the current density caused by electrons flowing to the surface. In the bulk of the uniform semiconductor, there is no dependence on space and thus the continuity equation reduces to $G = R$ under steady-state conditions. Using the recombination rate in the bulk as given by Eq. (2.14), the excess carrier concentration in the bulk is given by $\Delta n_\infty = G \tau_n$ as indicated in Fig. 2.9. Assuming that the electron current is a diffusion current of the form

$$J_n = e D_n \frac{\partial \Delta n(x,t)}{\partial x} \quad (2.36)$$

and inserting the diffusion current into Eq. (2.35), yields the continuity equation for diffusive currents, i.e.

$$\frac{\partial \Delta n(x,t)}{\partial t} = G - \frac{\Delta n(x,t)}{\tau_n} + D_n \frac{\partial^2 \Delta n(x,t)}{\partial x^2} . \quad (2.37)$$

At the semiconductor surface carriers will recombine rapidly due to surface states. The **boundary condition** at the surface is given by

$$eD_n \left. \frac{\partial \Delta n(x,t)}{\partial x} \right|_{x=0} = e S \Delta n(x,t) \big|_{x=0} \quad (2.38)$$

where S is the surface recombination velocity. The boundary condition states that minority carriers diffusing to the surface will recombine at the surface. We assume that the generation rate is constant with time, and thus the minority carrier concentration has no time dependence. The steady-state solution to the differential equation with the above boundary condition is given by

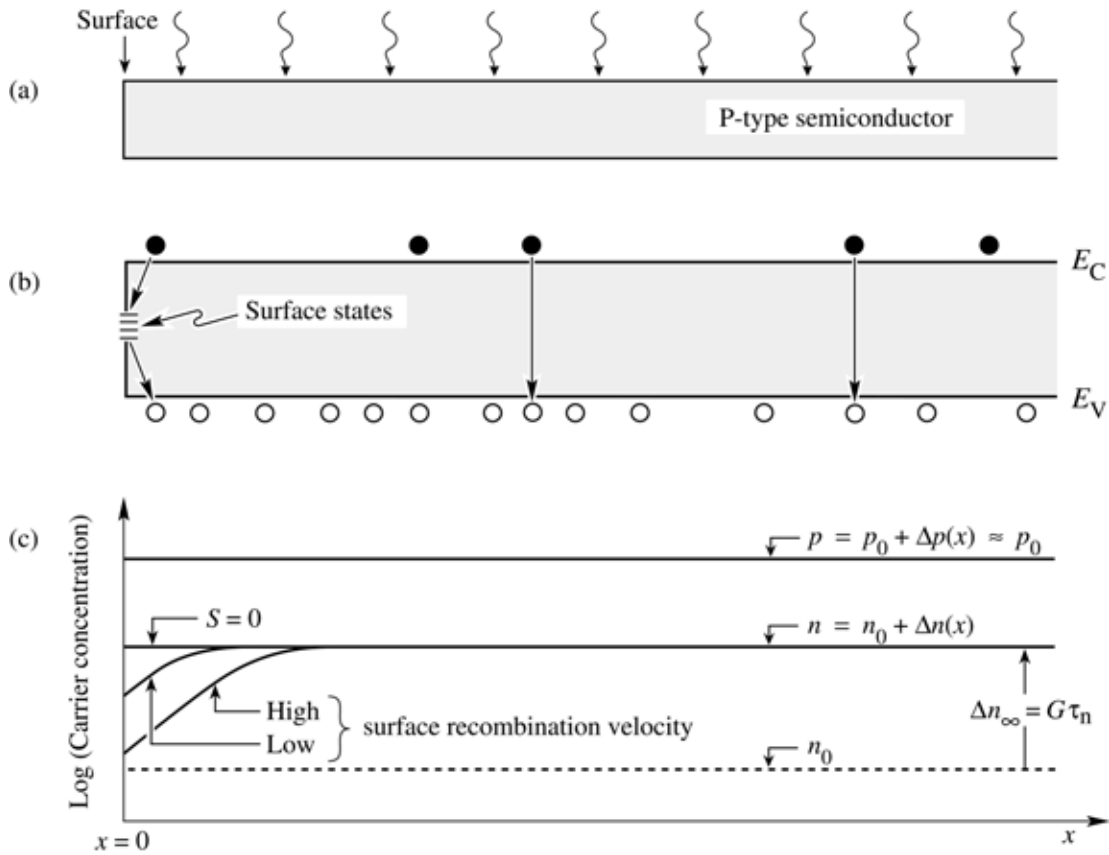


Fig. 2.9. (a) Illuminated p-type semiconductor, (b) band diagram, and (c) minority and majority carrier concentration near the surface assuming uniform carrier generation due to illumination. The excess carrier concentrations are Δn and Δp .

$$n(x) = n_0 + \Delta n(x) = n_0 + \Delta n_\infty \left[1 - \frac{\tau_n S \exp(-x/L_n)}{L_n + \tau_n S} \right]. \quad (2.39)$$

The carrier concentration near a semiconductor surface is shown in Fig. 2.9 for different surface recombination velocities. For $S \rightarrow 0$, the minority carrier concentration at the surface is identical to the bulk value, i.e. $n(0) \rightarrow n_0 + \Delta n_\infty$. For $S \rightarrow \infty$, the minority carrier concentration at the surface approaches the equilibrium value, i.e. $n(0) \rightarrow n_0$.

Surface recombination leads to a reduced luminescence efficiency and also to heating of the surface due to non-radiative recombination at the surface. Both effects are unwanted in electroluminescent devices. The surface recombination velocities for several semiconductors are summarized in Table 2.1. The data shown in the table shows that GaAs has a particularly high surface recombination velocity.

Semiconductor	Surface recombination velocity
GaAs	$S = 10^6$ cm/s
InP	$S = 10^3$ cm/s
Si	$S = 10^1$ cm/s

Table 2.1. Surface recombination velocities of several semiconductors.

Experimental evidence of surface recombination is illustrated in Fig. 2.10, which shows the luminescence emanating from a stripe-like current injection contact on a GaAs chip. The luminescence is viewed from the substrate side so that the stripe-like metal contact is “behind” the light-emitting region. Figure 2.10 clearly reveals that the luminescence decreases in the near-surface region.

Surface recombination can occur only when both types of carriers are present. It is important in the design of LEDs that the carrier-injected active region, in which naturally both types of carriers are present, *be far removed from any surface*. This can be achieved, for example, by carrier injection under a contact that is much smaller than the semiconductor die. Furthermore, the contact must be sufficiently far away from the side surfaces of the die. If the current flow is confined to the region below the contact, carriers will not “see” any semiconductor surfaces. Note that unipolar regions of a semiconductor device, e.g. the confinement regions, are not affected by surface recombination due to the lack of minority carriers.

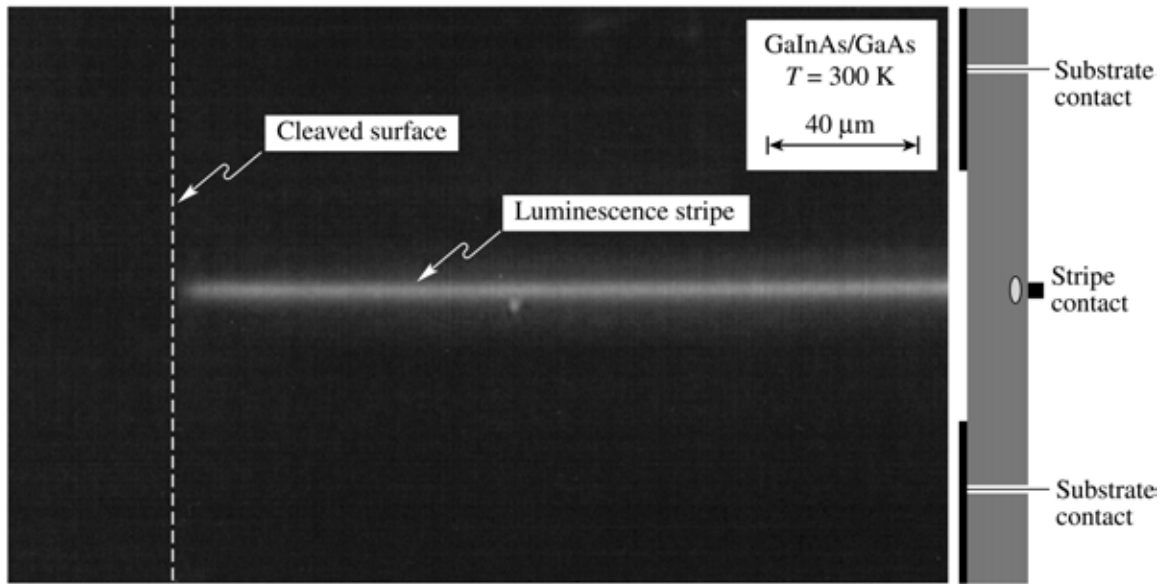


Fig. 2.10. Micrograph of a GaInAs/GaAs structure with a stripe-shaped contact on the top side and a contact widow on the substrate side of the device under current injection conditions. The luminescence emanating from the active region located below the stripe contact clearly decreases in the vicinity of the surface due to surface recombination.

Several passivation techniques have been developed to reduce the surface recombination in semiconductors, including treatments with sulfur, and other chemicals (Lipsanen *et al.*, 1999).

2.8 Competition between radiative and non-radiative recombination

So far we have seen that several mechanisms for non-radiative recombination exist, including Shockley–Read, Auger, and surface recombination. Even though non-radiative recombination can be reduced, it can never be totally eliminated. For example, surface recombination can be drastically reduced by device designs that spatially separate the active region from any surfaces. However even if the separation is large, a few carriers will still diffuse to the surface and recombine there.

Just as for surface recombination, non-radiative bulk recombination and Auger recombination can never be totally avoided. Any semiconductor crystal will have some native defects. Even though the concentration of these native defects can be low, it is never zero. Thermodynamic considerations predict that if an energy E_a is needed to create a specific point defect in a crystal lattice, the probability that such a defect does indeed form at a specific lattice site, is given by the Boltzmann factor, i.e. $\exp(-E_a / kT)$. The product of the concentration of lattice sites and the Boltzmann factor gives the concentration of defects. A native point defect or

extended defect may form a deep state in the gap and thus be a non-radiative recombination center.

Exercise. Concentration of point defects. Assume that the energy required to move a substitutional lattice atom into an interstitial position is $E_a = 1.1$ eV. What is the equilibrium concentration of interstitial defects of a simple cubic lattice with lattice constant $a_0 = 2.5$ Å?

Solution: The concentration of lattice atoms of a simple cubic lattice is given by $N = a_0^{-3} = 6.4 \times 10^{22} \text{ cm}^{-3}$. The concentration of interstitial defects under equilibrium conditions at room temperature is then given by

$$N_{\text{defect}} = N \exp(-E_a/kT) = 2.7 \times 10^4 \text{ cm}^{-3}.$$

Note that the calculated concentration of defects is small when compared to the typical concentrations of electrons and holes. If the defect discussed here forms a level in the gap, non-radiative recombination through the defect level can occur.

Another issue is the chemical purity of semiconductors. It is difficult to fabricate materials with impurity levels lower than the *parts per billion* (ppb) range. Thus, even the purest semiconductors contain impurities in the 10^{12} cm^{-3} range. Some elements may form deep levels and thus reduce the luminescent efficiency.

In the 1960's, when the first III–V semiconductors had been demonstrated, the internal luminescence efficiencies at room temperature were very low, typically a fraction of 1 %. At the present time, high-quality bulk semiconductors and quantum well structures can have internal efficiencies exceeding 90 %, and in some cases even 99 %. This remarkable progress is due to improved crystal quality, and reduced defect and impurity concentrations.

Next, we calculate the internal quantum efficiency in a semiconductor with non-radiative recombination centers. If the radiative lifetime is denoted as τ_r and the non-radiative lifetime is denoted as τ_{nr} , then the total probability of recombination is given by the sum of the radiative and non-radiative probabilities:

$$\tau^{-1} = \tau_r^{-1} + \tau_{\text{nr}}^{-1} \quad (2.40)$$

The relative probability of radiative recombination is given by the radiative probability over the total probability of recombination. Thus the probability of radiative recombination or *internal quantum efficiency* is given by

$$\eta_{\text{int}} = \frac{\tau_r^{-1}}{\tau_r^{-1} + \tau_{\text{nr}}^{-1}} . \quad (2.41)$$

The internal quantum efficiency gives the ratio of the number of light quanta emitted inside the semiconductor to the number of charge quanta undergoing recombination. Note that not all photons emitted internally may escape from the semiconductor due to the light-escape problem, reabsorption in the substrate, or other reabsorption mechanisms.

References

- Agrawal G. P. and Dutta N. K. “Long wavelength semiconductor lasers” (John Wiley and Sons, New York, 1986)
- Ahrenkiel R. K. “Minority-carrier lifetime in III–V semiconductors” in “Minority carriers in III–V semiconductors: physics and applications” edited by R. K. Ahrenkiel and M. S. Lundstrom *Semiconductors and Semimetals* Vol. **39**, 40 (Academic Press, San Diego, 1993)
- Baraff G. A. and Schluter M. “Electronic structure, total energies, and abundances of the elementary point defects in GaAs” *Phys. Rev. Lett.* **55**, 1327 (1985)
- Ehrhardt A., Wettling W., and Bett A. “Transient photoluminescence decay study of minority carrier lifetime in GaAs heteroface solar cell structures” *Appl. Phys. A (Solids and Surfaces)* **A53**, 123 (1991)
- Grieshaber W., Schubert E. F., Goepfert I. D., Karlicek R. F. Jr., Schurman M. J. and Tran C. “Competition between band gap and yellow luminescence in GaN and its relevance for optoelectronic devices” *J. Appl. Phys.* **80**, 4615 (1996)
- Hall R. N. “Electron–hole recombination in germanium” *Phys. Rev.* **87**, 387 (1952)
- Hwang C. J. “Doping dependence of hole lifetime in n-type GaAs” *J. Appl. Phys.* **42**, 4408 (1971)
- Lipsanen H., Sopanen M., Ahopelto J., Sandman J., and Feldmann J. “Effect of InP passivation on carrier recombination in $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ surface quantum wells” *Jpn. J. Appl. Phys.* **38**, 1133 (1999)

- Longini R. L. and Greene “Ionization interaction between impurities in semiconductors and insulators” *Phys. Rev.* **102**, 992 (1956)
- Nelson R. J. and Sobers R. G. “Interfacial recombination velocity in GaAlAs/GaAs heterostructures” *Appl. Phys. Lett.* **32**, 761 (1978a)
- Nelson R. J. and Sobers R. G. “Minority-carrier lifetime and internal quantum efficiency of surface-free GaAs” *Appl. Phys. Lett.* **49**, 6103 (1978b)
- Neugebauer J. and Van de Walle C. “Gallium vacancies and the yellow luminescence in GaN” *Appl. Phys. Lett.* **69**, 503 (1996)
- Olshansky R., Su C. B., Manning J., and Powazinik W. “Measurement of radiative and non-radiative recombination rates in InGaAsP and AlGaAs light sources” *IEEE J. Quantum Electronics* **QE-20**, 838 (1984)
- Pantelides S. T. (Editor) “Deep centers in semiconductors” (Gordon and Breach, Yverdon, Switzerland, 1992)
- Saarinén K. *et al.* “Observation of native Ga vacancies in GaN by positron annihilation” *Phys. Rev. Lett.* **79**, 3030 (1997)
- Schubert E. F., unpublished (1995)
- Schubert E. F. “Doping in III–V semiconductors” (Cambridge University Press, Cambridge UK, 1993)
- Schubert E. F., Goepfert I. and Redwing J. M. “Evidence of compensating centers as origin of yellow luminescence in GaN” *Appl. Phys. Lett.* **71**, 3224 (1997)
- Shockley W. “Electrons and holes in semiconductors” (D. Van Nostrand Company, New York, 1950)
- Shockley W. and Read W. T. “Statistics of the recombinations of holes and electrons” *Phys. Rev.* **87**, 835 (1952)

3

Theory of radiative recombination

In this chapter, the theory of radiative recombination is discussed in terms of a semiclassical model based on equilibrium generation and recombination. This model was developed by van Roosbroeck and Shockley (1954). Subsequently, the Einstein model of spontaneous and stimulated transitions in a two-level atom will be discussed. For a quantum-mechanical discussion of spontaneous recombination, the reader is referred to the literature (Bebb and Williams, 1972; Dutta, 1993).

3.1 The van Roosbroeck–Shockley model

The van Roosbroeck–Shockley model allows one to calculate the spontaneous radiative recombination rate under equilibrium and non-equilibrium conditions. To calculate the recombination rate, the model requires the knowledge of only a few basic parameters, namely the bandgap energy, the absorption coefficient, and the refractive index. All of these parameters can be determined by simple, well-known experimental methods.

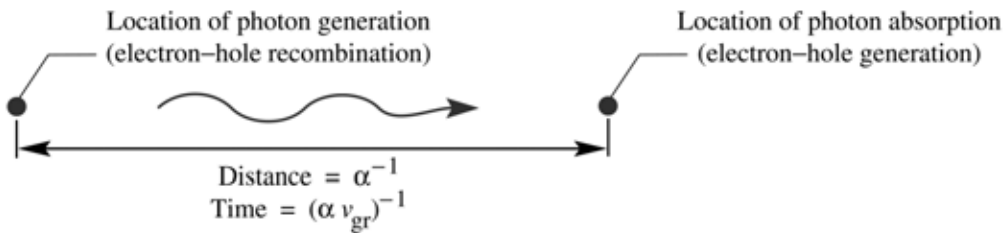


Fig. 3.1. Illustration of distance and elapsed time between a photon generation and absorption event.

Consider a semiconductor with an absorption coefficient $\alpha(\nu)$ given in units of cm^{-1} . A photon is generated in the semiconductor by electron–hole recombination and is subsequently absorbed, as illustrated in Fig. 3.1. The mean distance a photon of frequency ν travels before

being absorbed is simply $\alpha(\nu)^{-1}$ as illustrated in Fig. 3.1. The time it takes for a photon to be absorbed is given by

$$\tau(\nu) = \frac{1}{\alpha(\nu) v_{\text{gr}}} \quad (3.1)$$

where v_{gr} is the group velocity of photons propagating in the semiconductor. The group velocity of photons is given by

$$v_{\text{gr}} = \frac{d\omega}{dk} = \frac{d\nu}{d(1/\lambda)} = c \frac{d\nu}{d(\bar{n}\nu)} \quad (3.2)$$

where \bar{n} is the refractive index. Inserting the group velocity into Eq. (3.1) yields

$$\boxed{\frac{1}{\tau(\nu)} = \alpha(\nu) v_{\text{gr}} = \alpha(\nu) c \frac{d\nu}{d(\bar{n}\nu)}} \quad (3.3)$$

This equation gives the *inverse photon lifetime* or *photon absorption probability* per unit time. The product of the absorption probability and the photon density yields the photon absorption rate per unit time per unit volume.

Under equilibrium conditions, the density of photons per unit volume in a medium with refractive index \bar{n} is given by Planck's black body radiation formula

$$N(\lambda) d\lambda = \frac{8\pi}{\lambda^4} \frac{1}{e^{h\nu/kT} - 1} d\lambda \quad (3.4)$$

from which we can readily obtain $N(\nu) d\nu$, the number of photons having frequencies in the interval ν and $\nu + d\nu$. We have $\lambda = c / (\bar{n}\nu)$, so that

$$d\lambda = - \frac{c}{(\bar{n}\nu)^2} \frac{d(\bar{n}\nu)}{d\nu} d\nu. \quad (3.5)$$

Inserting this value in Eq. (3.4), yields Planck's black body photon distribution as a function of frequency

$$N(\nu) d\nu = \frac{8\pi \nu^2 \bar{n}^2}{c^3} \frac{d(\bar{n}\nu)}{d\nu} \frac{1}{e^{h\nu/kT} - 1} d\nu \quad (3.6)$$

The absorption rate per unit volume in the frequency interval ν and $\nu + d\nu$ is given by the photon density divided by the mean lifetime of photons

$$R_0(\nu) = \frac{N(\nu)}{\tau(\nu)} = \frac{8\pi \nu^2 \bar{n}^2}{c^3} \frac{d(\bar{n}\nu)}{d\nu} \frac{1}{e^{h\nu/kT} - 1} \alpha(\nu) c \frac{d\nu}{d(\bar{n}\nu)}. \quad (3.7)$$

Integration over all frequencies yields the absorption rate per unit volume

$$R_0 = \int_0^\infty R_0(\nu) d\nu = \int_0^\infty \frac{8\pi \nu^2 \bar{n}^2}{c^2} \frac{\alpha(\nu)}{e^{h\nu/kT} - 1} d\nu \quad (3.8)$$

which is the celebrated **van Roosbroeck–Shockley equation**. The van Roosbroeck–Shockley equation can be simplified by writing the absorption coefficient as

$$\alpha = \alpha_0 \sqrt{(E - E_g)/E_g}. \quad (3.9)$$

The square-root dependence of the absorption coefficient is motivated by the proportionality of the absorption coefficient and the density of states, which in turn follows a square-root dependence on energy. Note that α_0 is the absorption coefficient at $h\nu = 2 E_g$. Approximate values of α_0 for several semiconductors are given in Table 3.1.

The van Roosbroeck–Shockley equation can be simplified further by neglecting the frequency dependence of the refractive index and using the refractive index value at the band edge. One obtains

$$R_0 = 8\pi c \bar{n}^2 \alpha_0 \sqrt{\frac{kT}{E_g}} \left(\frac{kT}{ch}\right)^3 \int_{x_g}^\infty \frac{x^2 \sqrt{x - x_g}}{e^x - 1} dx \quad (3.10)$$

where $x = h\nu / (kT) = E / (kT)$ and $x_g = E_g / (kT)$. Owing to the strong increase of the exponential function with x , only a small range of energies close to the bandgap contributes to the integral. The integral has no simple analytical solution and it needs to be evaluated by a numerical method.

Under equilibrium conditions, the carrier generation rate (photon absorption rate) is equal to the carrier recombination rate (photon emission rate). Thus, the van Roosbroeck–Shockley model provides the *equilibrium* recombination rate. As discussed earlier, the bimolecular rate equation, which applies to both equilibrium and non-equilibrium conditions, gives the number of recombination events occurring per unit volume per unit time:

$$R = B n p \quad (3.11)$$

Next, we use the van Roosbroeck–Shockley model to calculate the bimolecular recombination coefficient B . Under equilibrium conditions, it is $R = R_0 = B n_i^2$. Thus the bimolecular recombination coefficient is related to the equilibrium recombination rate by

$$B = \frac{R_0}{n_i^2} \quad (3.12)$$

Material	E_g (eV)	α_0 (cm ⁻¹)	\bar{n} (-)	R_0 (cm ⁻³ s ⁻¹)	n_i (cm ⁻³)	B (cm ³ s ⁻¹)	τ_{spont} (s)
GaAs	1.42	2×10^4	3.3	7.9×10^2	2×10^6	2.0×10^{-10}	5.1×10^{-9}
InP	1.35	2×10^4	3.4	1.2×10^4	1×10^7	1.2×10^{-10}	8.5×10^{-9}
GaN	3.4	2×10^5	2.5	8.9×10^{-30}	2×10^{-10}	2.2×10^{-10}	4.5×10^{-9}
GaP	2.26	2×10^3	3.0	1.0×10^{-12}	1.6×10^0	3.9×10^{-13}	2.6×10^{-6}
Si	1.12	1×10^3	3.4	3.3×10^6	1×10^{10}	3.2×10^{-14}	3.0×10^{-5}
Ge	0.66	1×10^3	4.0	1.1×10^{14}	2×10^{13}	2.8×10^{-13}	3.5×10^{-6}

Table 3.1. Bimolecular recombination coefficient at 300 K for different semiconductors as calculated from the energy gap, absorption coefficient, and refractive index at the bandgap energy. The spontaneous lifetime is given by $B^{-1} N_{D,A}^{-1}$ and it is calculated for a majority carrier concentration of 10^{18} cm^{-3} .

Table 3.1 shows the bimolecular recombination coefficient for different semiconductors as calculated from Eqs. (3.10) and (3.12). All material parameters used in the calculation are given in the table. The calculated results reveal that $B = 10^{-9} - 10^{-11} \text{ cm}^3/\text{s}$ for direct-gap III–V

semiconductors. This calculated result agrees well with experimental results. GaP, Si, and Ge, all indirect-gap semiconductors, have a much smaller bimolecular recombination coefficient compared with direct-gap III–V semiconductors.

There are several other ways to calculate the bimolecular recombination coefficient. An early calculation by Hall (1960) using a two-band model produced a bimolecular recombination coefficient of

$$B = 5.8 \times 10^{-13} \frac{\text{cm}^3}{\text{s}} \left(\frac{m_h^*}{m_e} + \frac{m_e^*}{m_e} \right)^{-3/2} \left(1 + \frac{m_e}{m_h^*} + \frac{m_e}{m_e^*} \right) \left(\frac{300 \text{ K}}{T} \right)^{3/2} \left(\frac{E_g}{1 \text{ eV}} \right)^2 \bar{n}, \quad (3.13)$$

where m_e^* , m_h^* , and m_e are the effective electron mass, effective hole mass, and free electron mass, respectively. Garbuzov (1982) described a simple quantum-mechanical calculation for direct-gap semiconductors and obtained the following expression for the bimolecular recombination coefficient:

$$B = 3.0 \times 10^{-10} \frac{\text{cm}^3}{\text{s}} \left(\frac{300 \text{ K}}{T} \right)^{3/2} \left(\frac{E_g}{1.5 \text{ eV}} \right)^2. \quad (3.14)$$

All of the methods described here to calculate B give reasonably similar results. The B coefficient is in the $10^{-10} \text{ cm}^3/\text{s}$ range when calculated for GaAs at 300 K using Eqs. (3.10) and (3.12) – (3.14).

The bimolecular rate equation, $R = B n p$, applies to *dilute* carrier concentrations, i.e. to non-degenerately doped semiconductors. Thus the bimolecular recombination coefficient applies to semiconductors with non-degenerate carrier concentrations. In this case, the bimolecular recombination coefficient is independent of the carrier concentration. For very high carrier concentrations, however, the bimolecular recombination coefficient decreases due to an increasing momentum mismatch between electrons and holes. The bimolecular recombination coefficient in the degenerate regime can be expressed as

$$B|_{\text{high concentrations}} = B - \frac{n}{N_c} B^*. \quad (3.15)$$

That is, the recombination coefficient is reduced at high concentrations. A detailed discussion of the bimolecular recombination coefficient at high concentrations including numerical values for B^* can be found in the literature (Agrawal and Dutta, 1986; Olshansky *et al.*, 1984).

Exercise. Radiative efficiency. Analyze the temperature dependence of the radiative van Roosbroeck–Shockley model and of the non-radiative Shockley–Read model and predict the temperature dependence of the radiative efficiency in semiconductors.

3.2 The Einstein model

The first theory of optical transitions was developed by Albert Einstein. The Einstein model includes *spontaneous* and *stimulated* or *induced transitions*. *Spontaneous transitions* occur spontaneously, that is, without an external stimulus. In contrast, *stimulated transitions* are induced by an external stimulus, namely a photon. Thus the induced transition rates are proportional to the photon density or radiation density.

The coefficients \mathcal{A} and \mathcal{B} describe spontaneous and stimulated transitions in an atom with two quantized levels. These transitions are illustrated schematically in Fig. 3.2. Denoting the two levels as “1” and “2”, as shown in Fig. 3.2, Einstein postulated the probability per unit time for the downward transition ($2 \rightarrow 1$) and upward transition ($1 \rightarrow 2$) as

$$\boxed{W_{2 \rightarrow 1} = \mathcal{B}_{2 \rightarrow 1} \rho(\nu) + \mathcal{A}} \quad (3.16)$$

and

$$\boxed{W_{1 \rightarrow 2} = \mathcal{B}_{1 \rightarrow 2} \rho(\nu)} \quad (3.17)$$

The downward transition probability (per atom) has *two* terms, namely the induced term and the spontaneous term. The induced term, $\mathcal{B}_{2 \rightarrow 1} \rho(\nu)$, is proportional to the radiation density $\rho(\nu)$. The spontaneous downward transition probability is a constant \mathcal{A} . Note that the spontaneous lifetime is given by $\tau_{\text{spont}} = \mathcal{A}^{-1}$. The upward probability is just $\mathcal{B}_{1 \rightarrow 2} \rho(\nu)$.

The Einstein \mathcal{A} coefficient in an atom corresponds to the bimolecular recombination coefficient in a semiconductor. In an atom, the concentration terms (i.e. n and p in the bimolecular rate equation $R = B n p$) do not come into play, since for a downward transition to occur, the upper level must be occupied (one electron, “ $n = 1$ ”) and the lower level must be unoccupied (one hole, “ $p = 1$ ”).

Einstein showed that $\mathcal{B} = \mathcal{B}_{2 \rightarrow 1} = \mathcal{B}_{1 \rightarrow 2}$. Thus stimulated absorption and stimulated emission are complementary processes. He also showed that the ratio of the coefficients at a frequency ν in an isotropic medium with refractive index \bar{n} is a constant given by $\mathcal{A}/\mathcal{B} = 8 \pi \bar{n}^3 h \nu^3 / c^3$.

The equivalence of $\mathcal{B}_{2 \rightarrow 1}$ and $\mathcal{B}_{1 \rightarrow 2}$ can also be shown by quantum mechanical considerations, namely by using Fermi's Golden Rule. However, a detailed discussion of the Einstein model goes beyond the scope of this discussion.

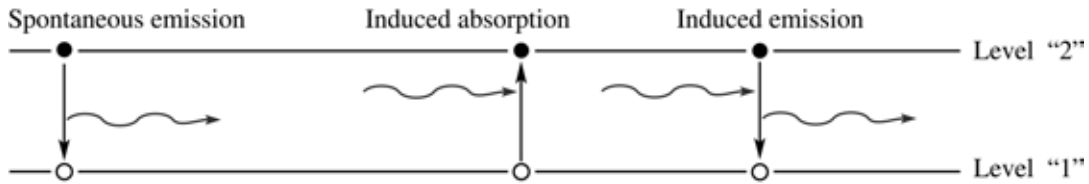


Fig. 3.2. Spontaneous emission, induced absorption, and induced emission events in the two-level atom model.

References

- Agrawal G. P. and Dutta N. K. "Long-wavelength semiconductor lasers" Chap. 3 (Van Nostrand Reinhold, New York, 1986)
- Bebb H. B. and Williams E. W. "Photoluminescence I: Theory" in "Transport and optical phenomena" edited by R. K. Willardson and A. C. Beer *Semiconductors and Semimetals* Vol. **8**, p. 181 (Academic Press, New York, 1972)
- Dutta N. K. "Radiative transitions in GaAs and other III–V compounds" in "Minority carriers in III–V semiconductors: physics and applications" edited by R. K. Ahrenkiel and M. S. Lundstrom *Semiconductors and Semimetals* Vol. **39**, p. 1 (Academic Press, San Diego, 1993)
- Garbuzov D. Z. "Radiation effects, lifetimes and probabilities of band-to-band transitions in direct A_3B_5 compounds of GaAs type" *J. Luminescence* **27**, 109 (1982)
- Hall R. N. "Recombination processes in semiconductors" *Proc. Inst. Elect. Eng.* **106 B**, Suppl. 17, 983 (1960)
- Olshansky R., Su C. B., Manning J., and Powazinik W. "Measurement of radiative and non-radiative recombination rates in InGaAsP and AlGaAs light sources" *IEEE J. Quantum Electron.* **QE-20**, 838 (1984)
- Van Roosbroeck W. and Shockley W. "Photon-radiative recombination of electrons and holes in germanium" *Phys. Rev.* **94**, 1558 (1954)

LED basics: electrical properties

4.1 Diode current–voltage characteristic

The electrical characteristics of p-n junctions will be summarized, however, a detailed derivation of the results will not be provided in this chapter. We consider an *abrupt p-n junction* with a donor concentration of N_D and an acceptor concentration of N_A . All dopants are assumed to be fully ionized so that the free electron concentration is given by $n = N_D$ and the free hole concentration is given by $p = N_A$. It is further assumed that no *compensation* of the dopants occurs by unintentional impurities and defects.

In the vicinity of an unbiased p-n junction, electrons originating from donors on the n-type side diffuse over to the p-type side where they encounter many holes with which they recombine. A corresponding process occurs with holes that diffuse to the n-type side. As a result, a region near the p-n junction is depleted of free carriers. This region is known as the **depletion region**.

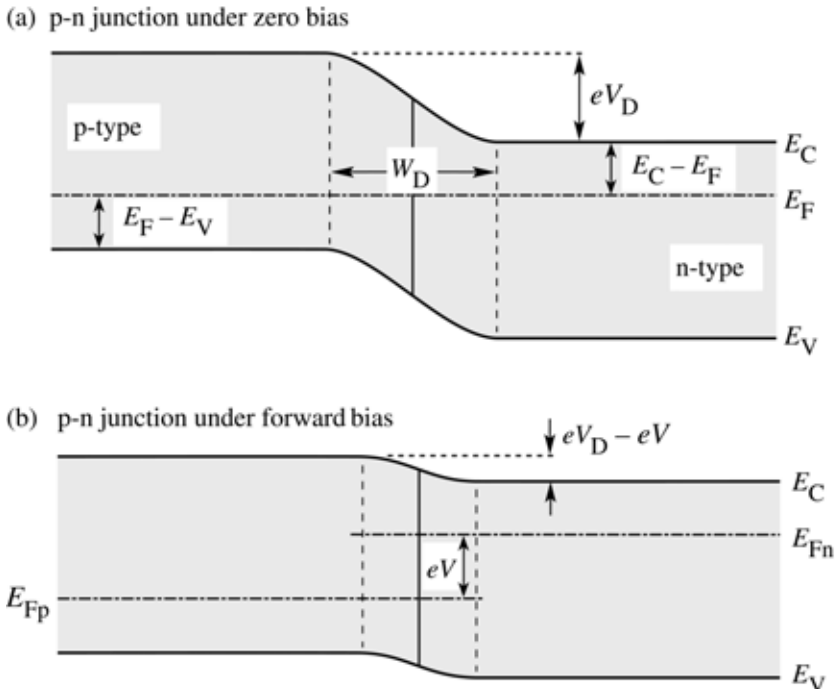


Fig. 4.1. P-n junction under (a) zero bias and (b) forward bias. Under forward bias conditions, minority carriers diffuse into the neutral regions where they recombine.

In the absence of free carriers in the depletion region, the only charge in the depletion region is from ionized donors and acceptors. These dopants form a space charge region, i.e. donors on the n-type side and acceptors on the p-type side. The space charge region produces a potential that is called the **diffusion voltage**, V_D . The diffusion voltage is given by

$$V_D = \frac{kT}{e} \ln \frac{N_A N_D}{n_i^2} \quad (4.1)$$

where N_A and N_D is the acceptor and donor concentration, respectively, and n_i is the intrinsic carrier concentration of the semiconductor. The diffusion voltage is shown in the band diagram of Fig. 4.1. The diffusion voltage represents the barrier that free carriers must overcome in order to reach the neutral region of opposite conductivity type.

The width of the depletion region, the charge in the depletion region, and the diffusion voltage are related by the Poisson equation. It is therefore possible to determine the depletion layer width from the diffusion voltage. The depletion layer width is given by

$$W_D = \sqrt{\frac{2\varepsilon}{e} (V - V_D) \left(\frac{1}{N_A} + \frac{1}{N_D} \right)} \quad (4.2)$$

where $\varepsilon = \varepsilon_r \varepsilon_0$ is the dielectric permittivity of the semiconductor and V is the diode bias voltage.

Upon application of the bias voltage to the p-n junction, the voltage is going to drop across the depletion region. This region is highly resistive due to the fact that it is depleted of free carriers. An external bias therefore decreases or increases the p-n junction barrier for forward or reverse bias, respectively. Under *forward* bias conditions, electrons and holes are injected into the region with opposite conductivity type and current flow *increases*. The carriers diffuse into the regions of opposite conductivity type where they will eventually recombine, thereby emitting a photon.

The current–voltage (I – V) characteristic of a p-n junction was first developed by Shockley and the equation describing the I – V curve of a p-n junction diode is therefore referred to as the **Shockley equation**. The Shockley equation for a diode with cross-sectional area A is given by

$$I = eA \left(\sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} + \sqrt{\frac{D_n}{\tau_n}} \frac{n_i^2}{N_A} \right) (e^{eV/kT} - 1) \quad (4.3)$$

where $D_{n,p}$ and $\tau_{n,p}$ are the electron and hole diffusion constants and the electron and hole minority carrier lifetimes, respectively.

Under reverse bias conditions, the diode current saturates and the saturation current is given by the factor preceding the exponential function in the Shockley equation. The diode I - V characteristic can be written as

$$I = I_s \left(e^{eV/kT} - 1 \right) \quad \text{with} \quad I_s = eA \left(\sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} + \sqrt{\frac{D_n}{\tau_n}} \frac{n_i^2}{N_A} \right). \quad (4.4)$$

Under typical forward bias conditions, the diode voltage is $V \gg kT/e$, and thus $[\exp(eV/kT) - 1] \approx \exp(eV/kT)$. Using Eq. (4.1), the Shockley equation can be rewritten, for forward bias conditions, as

$$I = eA \left(\sqrt{\frac{D_p}{\tau_p}} N_A + \sqrt{\frac{D_n}{\tau_n}} N_D \right) e^{e(V-V_D)/kT}. \quad (4.5)$$

The exponent of the exponential function in Eq. (4.5) illustrates that the current strongly increases as the diode voltage approaches the diffusion voltage, i.e. $V \approx V_D$. The voltage at which the current strongly increases is called the **threshold voltage** and this voltage is given by $V_{th} \approx V_D$.

The band diagram of a p-n junction, shown in Fig. 4.1, also illustrates the separation of the Fermi level from the conduction and valence band edge. The difference in energy between the Fermi level and the band edges can be inferred from Boltzmann statistics and is given by

$$E_C - E_F = -kT \ln \frac{n}{N_C} \quad \text{for the n-type side} \quad (4.6)$$

and

$$E_F - E_V = -kT \ln \frac{p}{N_V} \quad \text{for the p-type side.} \quad (4.7)$$

The band diagram shown in Fig. 4.1 illustrates that the following sum of energies is zero:

$$eV_D - E_g + (E_F - E_V) + (E_C - E_F) = 0 . \quad (4.8)$$

In highly doped semiconductors, the separation between the band edges and the Fermi level is small compared with the bandgap energy, i.e. $E_C - E_F \ll E_g$ on the n-type side and $E_F - E_V \ll E_g$ on the p-type side. Furthermore, these quantities depend only weakly (logarithmic dependence) on the doping concentration as inferred from Eqs. (4.6) and (4.7). Thus the third and fourth summand of Eq. (4.8) can be neglected and the diffusion voltage can be approximated by the bandgap energy divided by the elementary charge

$$V_{th} \approx V_D \approx E_g / e . \quad (4.9)$$

Several diode I - V characteristics of semiconductors made from different materials are shown in Fig. 4.2 along with the bandgap energy of these materials. The experimental threshold voltages shown in the figure and the comparison with the bandgap energy of these materials indicates that the energy gap and the threshold voltage indeed agree reasonably well.

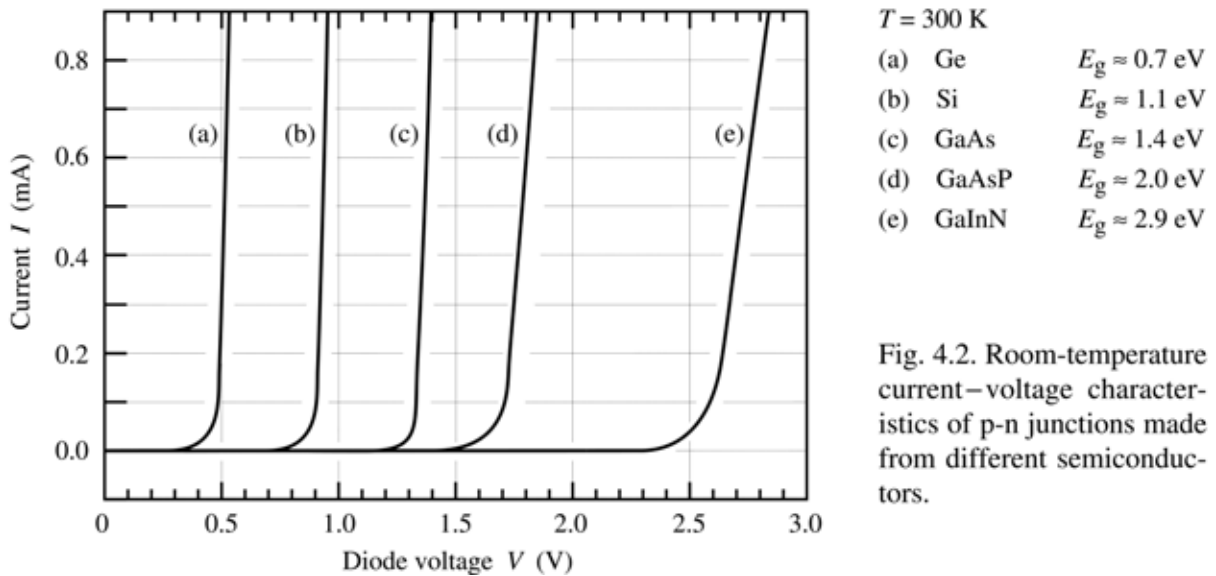


Fig. 4.2. Room-temperature current-voltage characteristics of p-n junctions made from different semiconductors.

The forward diode voltage at a diode current of 20 mA versus bandgap energy for LEDs emitting in the ultraviolet, visible, and infrared wavelength ranges is shown in Fig. 4.3 (Krames *et al.*, 2000). The solid line illustrates the expected forward diode voltage. The line equals the bandgap energy divided by the elementary charge. Inspection of the figure reveals that most semiconductor LEDs follow the solid line, except for LEDs based on III-V nitrides. This

peculiarity is due to several reasons. *First*, large bandgap discontinuities occur in the nitride material system which cause an additional voltage drop. *Secondly*, the contact technology is less mature in the nitride material system, which causes an additional voltage drop at the ohmic contacts. *Thirdly*, the p-type conductivity in bulk GaN is generally low. *Lastly*, a parasitic voltage drop can occur in the n-type buffer layer.

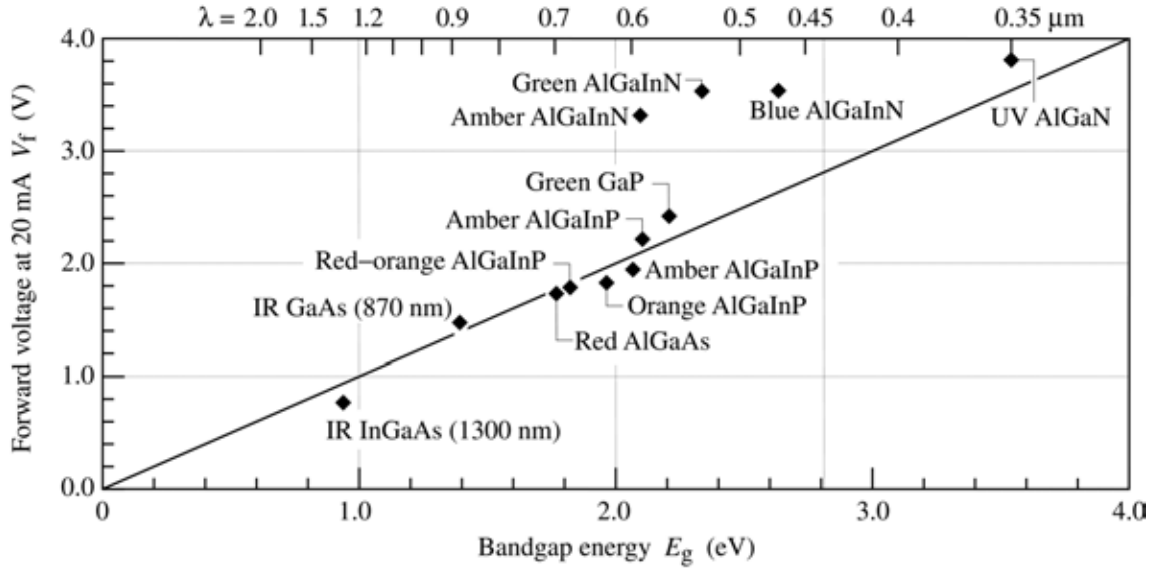


Fig. 4.3. Typical diode forward voltage versus bandgap energy for LEDs made from different materials (after Krames *et al.*, 2000; updated with UV LED data of Emerson *et al.*, 2002).

4.2 Deviations from the ideal I - V characteristic

The Shockley equation gives the expected theoretical I - V characteristic of a p-n junction. To describe experimentally measured characteristics, the following equation is used:

$$I = I_s e^{eV/(n_{\text{ideal}} kT)} \quad (4.10)$$

where n_{ideal} is the **ideality factor** of the diode. For a perfect diode, the ideality factor has a value of unity ($n_{\text{ideal}} = 1.0$). For real diodes, the ideality factor assumes values of typically $n_{\text{ideal}} = 1.1 - 1.5$. However, values as high as $n_{\text{ideal}} = 2.0$ have been found for III-V arsenide and phosphide diodes. Values as high as $n_{\text{ideal}} = 6.0$ have been found for GaN/GaInN diodes. For a detailed analysis of the diode ideality factor (see, for example, Rhoderick and Williams, 1988).

Frequently a diode has unwanted or **parasitic resistances**. The effect of a series resistance and a parallel resistance is shown in Fig. 4.4. A series resistance can be caused by excessive

contact resistance or by the resistance of the neutral regions. A parallel resistance can be caused by any channel that bypasses the p-n junction. This bypass can be caused by damaged regions of the p-n junction or by surface imperfections.

The diode I - V characteristic, as given by the Shockley equation, needs to be modified in order to take into account parasitic resistances. Assuming a shunt with resistance R_p (parallel to the ideal diode) and a series resistance R_s (in series with the ideal diode *and* the shunt), the I - V characteristic of a forward-biased p-n junction diode is given by

$$I - \frac{(V - IR_s)}{R_p} = I_s e^{e(V - IR_s)/(n_{\text{ideal}} kT)} \quad (4.11)$$

For $R_p \rightarrow \infty$ and $R_s \rightarrow 0$, this equation reduces to the Shockley equation.

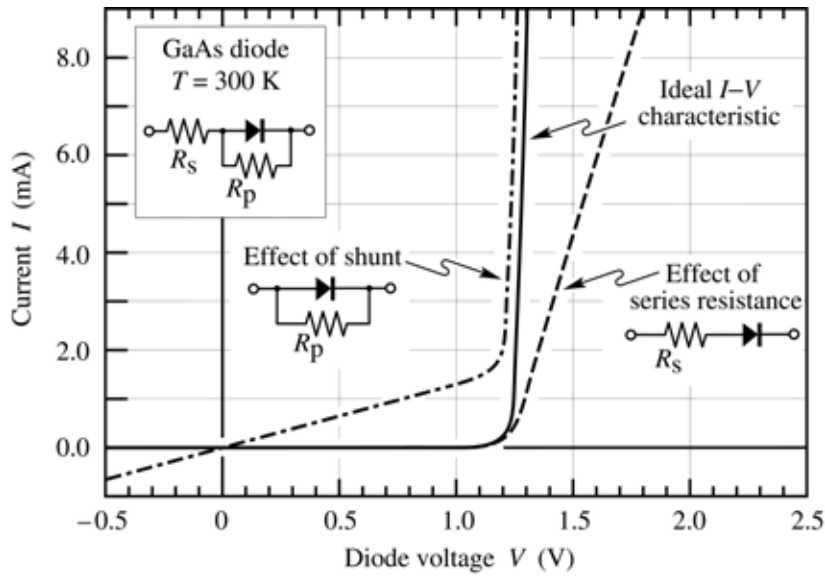


Fig. 4.4. Effect of a series resistance and a parallel resistance (shunt) on the I - V characteristic of a p-n junction diode.

4.3 Evaluation of diode parasitic resistances

The diode parallel resistance can be evaluated near the origin of the I - V diagram where $V \ll E_g / e$. For this voltage range, the p-n junction current can be neglected and the parallel resistance is given by

$$R_p = \left. \frac{dV}{dI} \right|_{\text{near origin}} \quad (4.12)$$

Note that in any reasonable diode, the parallel resistance is much larger than the series resistance so that the series resistance need not be taken into account when evaluating the parallel resistance.

The series resistance can be evaluated at a high voltage where $V > E_g / e$. For sufficiently large voltages, the diode I-V characteristic becomes linear and the series resistance is given by

$$R_s = \left. dV / dI \right|_{\text{at voltages exceeding turn-on}} \quad (4.13)$$

However, it may not be practical to evaluate the diode resistance at high voltages due to device heating effects. For this case, the following procedure will be suitable.

For devices with a high parallel resistance ($R_p \rightarrow \infty$), the diode I - V characteristic, given in Eq. (4.11), can be written as

$$I = I_s e^{e(V - IR_s) / (n_{\text{ideal}} kT)} \quad (4.14)$$

Solving the equation for V and then differentiating V with respect to I yields

$$\frac{dV}{dI} = R_s + \frac{n_{\text{ideal}} kT}{e} \frac{1}{I}, \quad (4.15)$$

where the second summand on the right-hand side of the equation represents the differential p-n junction resistance. Multiplication of the equation by I allows one to identify the series resistance of the diode as the slope of a $(I dV / dI)$ -versus- I plot, as shown in Fig. 4.5.

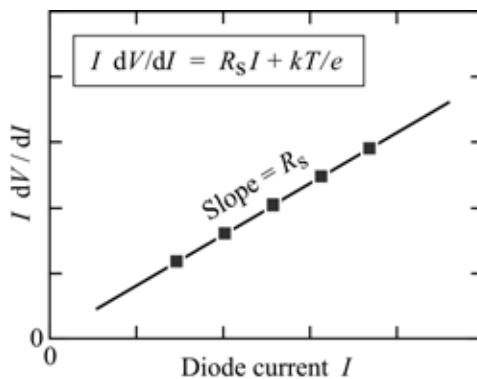


Fig. 4.5. Method for evaluating the diode series resistance. The equation shown as an inset is valid for forward bias ($V \gg kT/e$).

4.4 Emission energy

The energy of photons emitted from a semiconductor with energy gap E_g is given by the bandgap energy, i.e.

$$h \nu \approx E_g . \quad (4.16)$$

In an ideal diode, every electron injected into the active region will generate a photon. Conservation of energy thus requires that the energy with which an electron is injected is equal to the photon energy. Thus energy conservation requires

$$e V = h \nu . \quad (4.17)$$

That is, the voltage applied to the LED multiplied by the elementary charge is equal to the photon energy. There are several effects that can change the diode voltage from the ideal value given by Eq. (4.17). These effects will be discussed below.

4.5 Carrier distribution in p-n homojunctions

The carrier distribution in p-n homojunctions, i.e. p-n junctions consisting of a single material, depends on the diffusion constant of the carriers. The *diffusion constant* of carriers is not easily measured. Much more common is the measurement of the *carrier mobility*; for example, by the Hall effect. The diffusion constant can be inferred from the carrier mobility by the **Einstein relation** which, for non-degenerate semiconductors, is given by

$$D_n = \frac{kT}{e} \mu_n \quad \text{and} \quad D_p = \frac{kT}{e} \mu_p . \quad (4.18)$$

Carriers injected into a neutral semiconductor, with no external electric field applied, propagate by diffusion. If carriers are injected into a region with opposite conductivity type, the minority carriers will eventually recombine. The mean distance a minority carrier diffuses before recombination is the **diffusion length**. Electrons injected into a p-type region will, on average, diffuse over the diffusion length L_n , before recombining with holes. The diffusion length is given by

$$L_n = \sqrt{D_n \tau_n} \quad \text{and} \quad L_p = \sqrt{D_p \tau_p} \quad (4.19)$$

where τ_n and τ_p are the electron and hole minority carrier lifetimes, respectively. In typical semiconductors, the diffusion length is of the order of a several micrometers. For example, the diffusion length of electrons in p-type GaAs is given by $L_n = (220 \text{ cm}^2/\text{s} \times 10^{-8} \text{ s})^{1/2} \approx 15 \text{ } \mu\text{m}$. Thus, minority carriers are distributed over a region several microns thick.

The distribution of carriers in a p-n junction under zero and under forward bias is shown in Figs. 4.6 (a) and (b), respectively. Note that minority carriers are distributed over quite a large distance. Furthermore, the minority carrier concentration decreases as these carriers diffuse further into the adjacent region. Thus recombination occurs over a large region, with a strongly changing minority carrier concentration. As will be shown below, the *large recombination region in homojunctions* is not beneficial for efficient recombination.

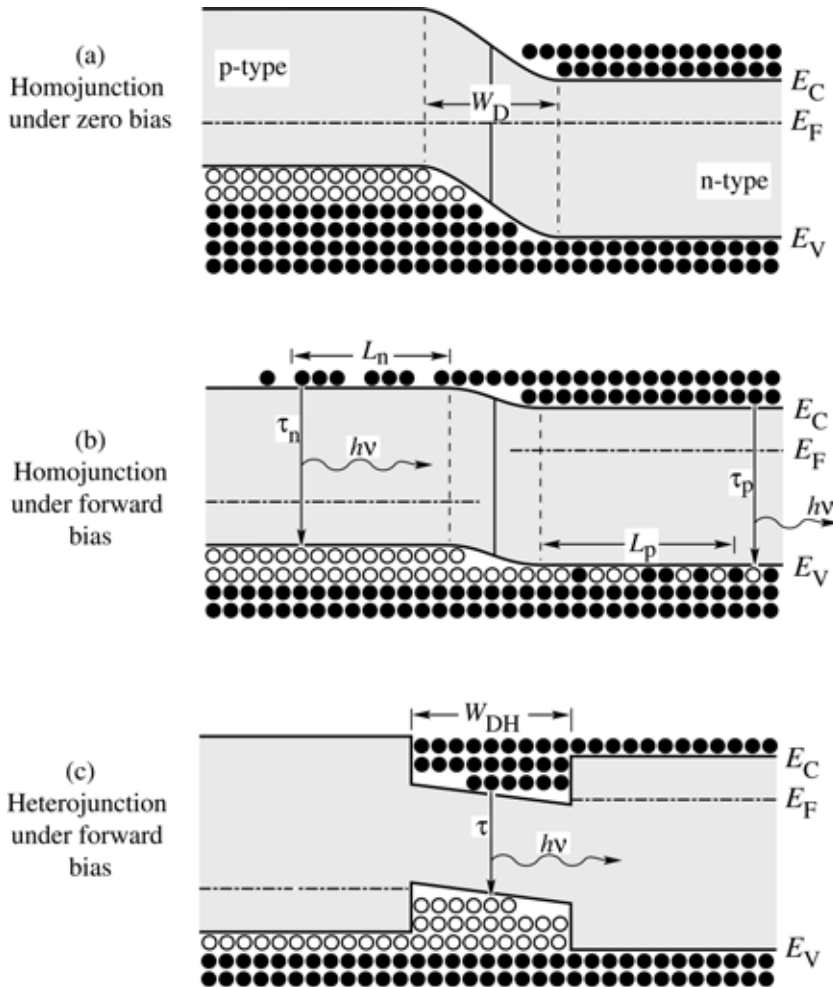


Fig. 4.6. P-n homojunction under (a) zero and (b) forward bias. P-n heterojunction (c) under forward bias. In homojunctions, carriers diffuse, on average, over the diffusion lengths L_n and L_p before recombining. In heterojunctions, carriers are confined by the heterojunction barriers.

4.6 Carrier distribution in p-n heterojunctions

All high-intensity light-emitting diodes do *not* use the homojunction design but rather employ heterojunctions, which have clear advantages over homojunction devices. Heterojunction devices

employ two types of semiconductors, namely a small-bandgap active region and a large-bandgap barrier region. If a structure consists of *two* barriers, i.e. two large-bandgap semiconductors, then the structure is called a ***double heterostructure*** (frequently abbreviated as ***DH***).

The effect of heterojunctions on the carrier distribution is shown in Fig. 4.6 (c). Carriers injected into the active region of the double heterostructure are confined to the active region by means of the barriers. As a result, the thickness of the region in which carriers recombine *is given by the thickness of the active region rather than the diffusion length*.

The consequences of this change are significant. We assume that the thickness of the active region is much smaller than the typical diffusion length. Diffusion lengths may range from 1 to 20 μm . The active region of double heterojunctions may range from 0.01 to 1.0 μm . Thus, carriers in the active region of a double heterostructure have a much higher concentration than carriers in homojunctions, which are distributed over several diffusion lengths. Recalling that the radiative recombination rate is given by the bimolecular recombination equation, i.e.

$$R = B n p \quad (4.20)$$

it is clear that a high concentration of carriers in the active region *increases* the radiative recombination rate and decreases the recombination lifetime. For this reason, all high-efficiency LED designs employ double heterostructure or quantum well designs.

4.7 The effect of heterojunctions on device resistance

The employment of heterostructures allows one to improve the efficiency of LEDs by confining carriers to the active region, thereby avoiding diffusion of minority carriers over long distances. Heterostructures can also be used to confine light to waveguide regions; in particular, in edge-emitting LEDs. Generally, modern semiconductor LEDs and lasers have many heterojunctions, e.g. for contact layers, active regions, and waveguiding regions. Although heterostructures allow for improved LED designs, there are also problems associated with heterojunctions.

One of the problems introduced by heterostructures is the resistance caused by the heterointerface. The origin of the resistance is illustrated in Fig. 4.7 (a), which shows the band diagram of a heterostructure. The heterostructure consists of two semiconductors with different bandgap energy and it is assumed that both sides of the heterostructure are of n-type conductivity. Carriers in the large-bandgap material will diffuse over to the small-bandgap material where they occupy conduction band states of lower energy. As a result of the electron

transfer, an electrostatic dipole forms, consisting of a positively charged depletion layer with ionized donors in the large-bandgap material, and a negatively charged electron accumulation layer in the small-bandgap material. The charge transfer leads to the band bending illustrated in Fig. 4.7 (a). Carriers transferring from one semiconductor to the other must overcome this barrier by either tunneling or by thermal emission over the barrier. The resistance caused by heterojunctions can have a strong deleterious effect on device performance, especially in high-power devices. The thermal power produced by heterostructure resistances leads to heating of the active region, thereby decreasing the radiative efficiency.

It has been shown that heterostructure band discontinuities can be completely eliminated by **grading** of the chemical composition of the semiconductor in the vicinity of the heterostructure (Schubert *et al.*, 1992). The band diagram of a graded heterostructure is shown in Fig. 4.7 (b). Inspection of the figure reveals that there is no longer a spike in the conduction band which hinders the electron flow. It has been shown that the resistance of parabolically graded heterostructures is comparable to bulk material resistance. Thus, the additional resistance introduced by abrupt heterostructures can be completely eliminated by parabolic grading.

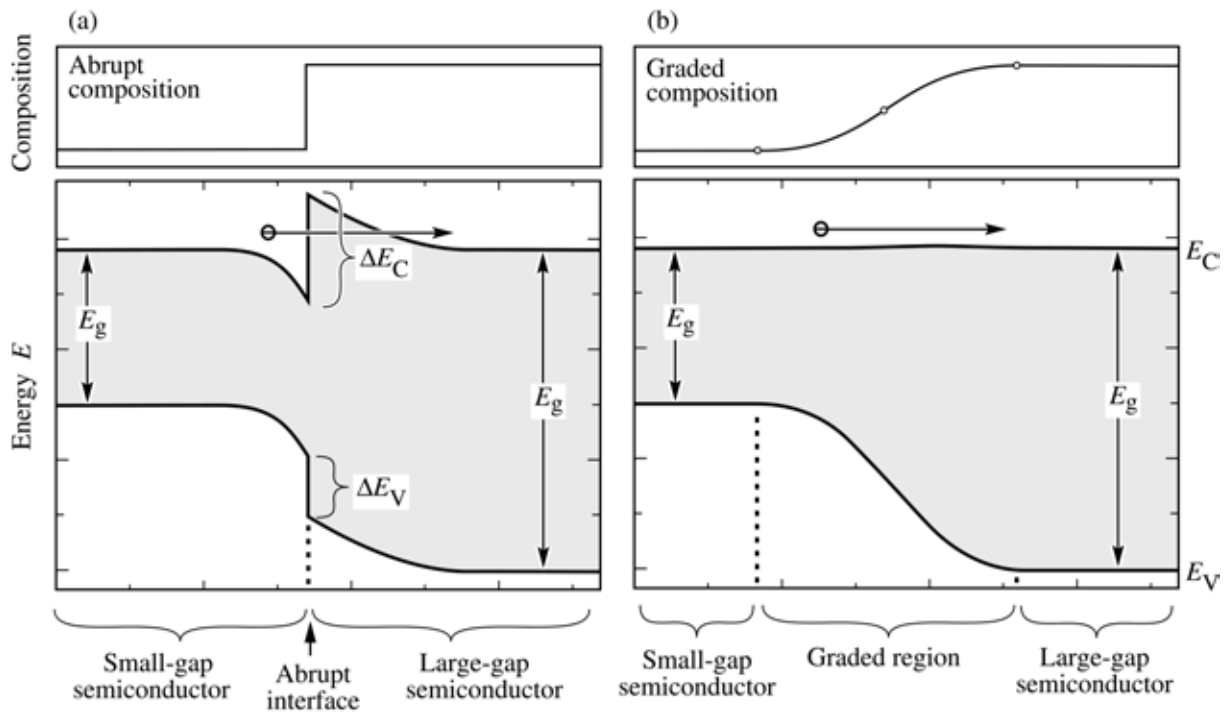


Fig. 4.7. Band diagram of (a) an abrupt n-type-n-type heterojunction and (b) a graded heterojunction of two semiconductors with different bandgap energy. The abrupt junction is more resistive than the graded junction due to the electron barrier forming at the abrupt junctions (after Schubert *et al.*, 1992).

The shape of the graded region should be parabolic for the following reason. The large-bandgap material will be depleted of free carriers due to electron transfer to the small-bandgap material. Thus the charge concentration in the large-bandgap material will be the donor concentration. Assuming that the donor concentration N_D is a constant throughout the heterostructure, the solution of the Poisson equation yields the electrostatic potential

$$\Phi = \frac{e N_D}{2 \varepsilon} x^2 . \quad (4.21)$$

The equation reveals that the potential depends *quadratically* on the spatial coordinate x , i.e. the potential has a *parabolic* shape. In order to compensate for the parabolic shape of the depletion potential, the composition of the semiconductor is *varied parabolically as well*, so that an overall *flat potential* results. It is assumed here that the parabolic variation of the chemical composition results in a parabolic change of the bandgap energy, i.e. that the bandgap energy depends linearly on the chemical composition and that bandgap *bowing* can be neglected.

Next, an approximate design rule for the grading of a heterostructure is given. Assume that the conduction band discontinuity of an abrupt heterojunction is given by ΔE_C and that the structure is uniformly doped with doping concentration N_D . Let us assume that carriers have transferred to the small-bandgap semiconductor, thus causing a depletion region of thickness W_D in the large-bandgap semiconductor. If the potential created in the depletion region is equal to $\Delta E_C / e$, then electrons will no longer transfer to the small-bandgap material. The thickness of the depletion region can be inferred from Eq. (4.21) to be

$$W_D = \sqrt{\frac{2 \varepsilon \Delta E_C}{e^2 N_D}} . \quad (4.22)$$

A heterostructure interface should be graded over the distance W_D in order to minimize the resistance introduced by an abrupt heterostructure. Although the result of Eq. (4.22) is an approximation it provides excellent guidance for device design. Steps can be taken to refine the calculation. For example, the potential change due to the electron accumulation layer in the small-bandgap material can be taken into account. Several software packages are available that allow for the numerical calculation of semiconductor heterostructures, for example the software package *Atlas* by the Silvaco Corporation.

Exercise: Grading of heterostructures. Assume that the conduction band discontinuity of an AlGaAs/GaAs heterostructure is given by $\Delta E_C = 300$ meV and that the structure is uniformly doped with donors of concentration $N_D = 5 \times 10^{17} \text{ cm}^{-3}$. Over what distance should the interface be graded in order to minimize the resistance occurring in abrupt heterostructures?

Solution: Calculating the depletion layer thickness from Eq. (4.22) yields $W_D = 30$ nm. Thus the heterostructure should be graded over 30 nm to minimize the heterostructure resistance. The graded region should have *two* parabolic regions as shown in Fig. 4.7 (b).

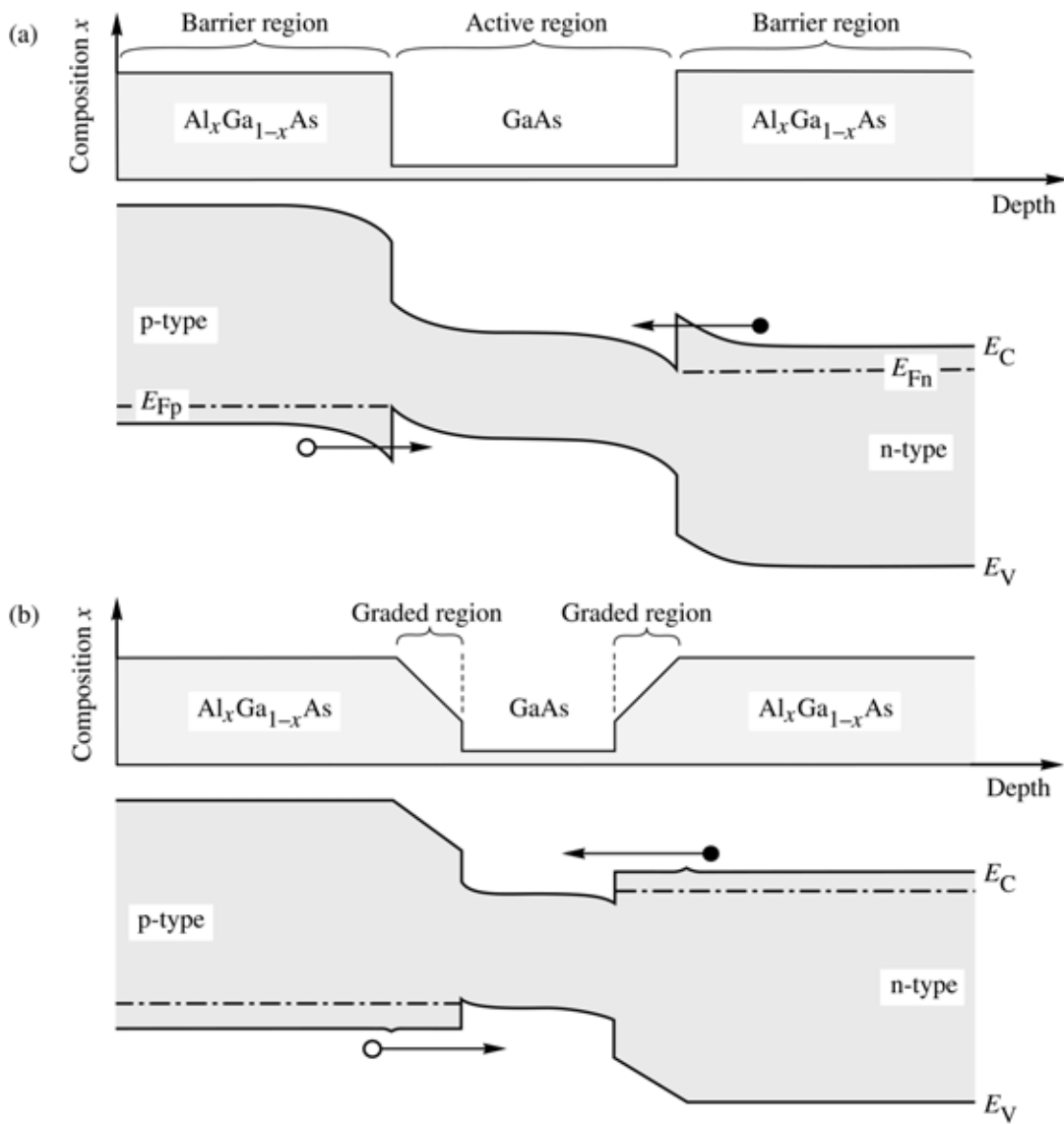


Fig. 4.8. Band diagram of (a) an abrupt double heterostructure and (b) a graded double heterostructure. The barrier–well interface of the abrupt junction is more resistive than the graded junction due to barriers forming at the interfaces.

Grading is useful for all heterostructures, including the heterostructures adjoining the active region. The effect of grading in a double heterostructure is shown in Fig. 4.8. The composition and the band diagram of an ungraded structure are shown in Fig. 4.8 (a). At both heterointerfaces, barriers develop as a result of free charge transferring to the active region. These barriers increase the device resistance under forward bias conditions.

The case of graded heterointerfaces is shown in Fig. 4.8 (b). The figure shows two linearly graded regions cladding the active region. The band diagram illustrates that barriers at the heterointerfaces can be effectively reduced or completely eliminated by grading. Note that the linear grading shown in Fig. 4.8 (b) results in small “spikes” at the interfaces between the linearly graded and the non-graded regions. These “spikes” are a result of the linear grading assumed, and would not result for parabolic grading.

Generally, the transport of carriers in heterostructures should be as *adiabatic* as possible, i.e. the carrier transport within the semiconductor device should not generate unnecessary heat. This is particularly true for high-power devices where additional heat generated inside the device leads to a performance loss due to increased operating temperature.

Finally, note that *lattice matching* is desirable in all heterostructure devices. It is also desirable in graded structures in order to minimize the number of misfit dislocations that act as non-radiative recombination centers.

4.8 Carrier loss in double heterostructures

In an ideal LED, the injected carriers are confined to the active region by the barrier layers adjoining the active regions. By means of confinement of carriers to the active region, a high carrier concentration is attained resulting in a high radiative efficiency of the recombination process.

The energy barriers confining the carriers to the active region are typically of the order of several hundred meV, i.e. much larger than kT . Nevertheless, some carriers will succeed in escaping from the active region into the barrier layers. The concentration of the escaping carriers in the barrier layers will be rather low, resulting in a low radiative efficiency of carriers in the barrier layers.

Free carriers in the active region are distributed according to the Fermi–Dirac distribution and, as a result, some carriers have a higher energy than the height of the confining barrier. Thus some of the carriers escape from the active region into the barrier regions as illustrated in Fig. 4.9.

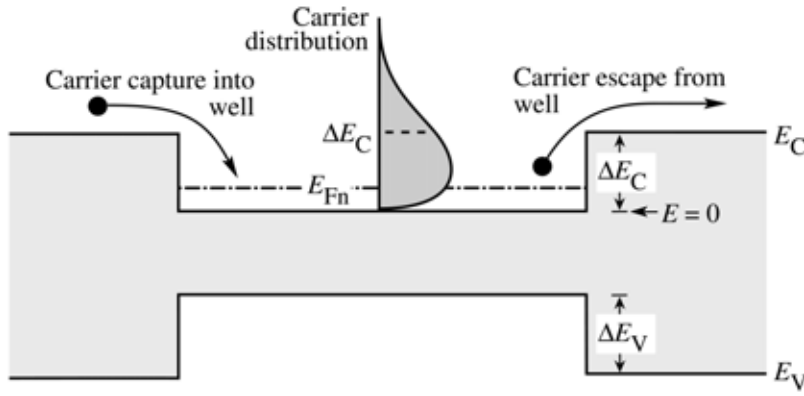


Fig. 4.9. Carrier capture and escape in a double heterostructure. Also shown is the carrier distribution in the active layer.

Consider electrons in the active region of a double heterostructure and assume that the active region is clad by a barrier with height ΔE_C , as shown in Fig. 4.9. The energy distribution of carriers is given by the Fermi–Dirac distribution. Thus, a certain fraction of the carriers residing in the active region have an energy higher than the energy of the barrier. The concentration of electrons with energy higher than the barrier is given by

$$n_B = \int_{E_B}^{\infty} \rho_{\text{DOS}} f_{\text{FD}}(E) dE \quad (4.23)$$

where f_{FD} is the Fermi–Dirac distribution function and E_B is the height of the barrier. For a bulk-type density of states, the concentration of carriers with energy higher than E_B is given by

$$n_B = \frac{1}{2\pi^2} \left(\frac{2m^*}{(h/2\pi)^2} \right)^{3/2} \int_{E_B}^{\infty} \frac{\sqrt{E - E_C}}{1 + e^{(E - E_{\text{Fn}})/kT}} dE. \quad (4.24)$$

Taking into account that we are interested in carriers with energies much higher than the Fermi energy, the Fermi–Dirac distribution can be approximated by the Boltzmann distribution. One obtains

$$n_B = N_c e^{(E_{\text{Fn}} - E_B)/kT} \quad (4.25)$$

where N_c is the effective density of states in the active region. Equation (4.25) gives the concentration of free carriers at the active-region–cladding-region interface. Minority carriers at the edge of the cladding layer will diffuse into the cladding layer. The diffusion process is

governed by the initial concentration n_B and the electron diffusion length L_n . Taking the location of the origin ($x = 0$) at the edge of the barrier, the carrier distribution can be written as

$$n_B(x) = n_B(0) e^{-x/L_n} = N_c e^{-(E_B - E_{Fn})/kT} e^{-x/L_n} \quad (4.26)$$

where $L_n = (D_n \tau_n)^{1/2}$ is the diffusion length, τ_n is the minority carrier lifetime, and D_n is the diffusion constant. The diffusion constant can be inferred from the mobility using the Einstein relation $D = \mu kT / e$.

The diffusion current density of electrons leaking over the barrier can be obtained from the carrier concentration gradient at $x = 0$, i.e.

$$J_n|_{x=0} = -eD_n \left. \frac{dn_B(x)}{dx} \right|_{x=0} = -eD_n \frac{n_B(0)}{L_n}. \quad (4.27)$$

The leakage current depends on the carrier concentration at the edge of the barrier. Thus a high barrier height is required to minimize the leakage current. Clearly, barriers must be much larger than kT for efficient confinement of carriers. Some material systems such as AlGaIn/GaN or AlGaAs/GaAs have relatively high barriers and therefore lower leakage currents over the barrier. Other material systems such as AlGaInP/AlGaInP emitting at 600 – 650 nm have lower barriers and therefore have a stronger carrier leakage over the barriers.

Note that the leakage increases exponentially with temperature. Thus, a decrease of the radiative efficiency of LEDs results as the temperature increases. To reduce the temperature dependence of the emission, high barriers are required. In addition to carrier leakage, other effects, such as Shockley–Read recombination also contribute to the lower radiative efficiency at high temperatures.

Exercise: Carrier leakage over a barrier. Electrons in the active region of a GaAs structure have a concentration of $2 \times 10^{18} \text{ cm}^{-3}$. Calculate the current density of the carrier loss over the barrier for barrier heights of 200 and 300 meV, assuming an electron mobility of $2000 \text{ cm}^2/\text{V s}$ and a minority carrier lifetime of 5 ns. Compare the calculated leakage current to typical LED injection currents of $0.1 - 10 \text{ kA/cm}^2$.

Solution: The Fermi level in GaAs at a carrier density of $2 \times 10^{18} \text{ cm}^{-3}$ is 77 meV above the conduction band edge. Assuming that the effective density of states in the barrier is the same as

in the GaAs active region, the carrier concentrations at the edge of the barrier are $3.9 \times 10^{15} \text{ cm}^{-3}$ for a 200 meV barrier and $8.3 \times 10^{13} \text{ cm}^{-3}$ for a 300 meV barrier. The diffusion constant, as inferred from the Einstein relation, is $D_n = 51.7 \text{ cm}^2/\text{s}$. The diffusion length is then given by $L_n = (D_n \tau_n)^{1/2} = 5.1 \text{ } \mu\text{m}$. The leakage current is calculated using Eq. (4.27), and one obtains 63 A/cm^2 for the 200 meV barrier and 1.3 A/cm^2 for the 300 meV barrier. The comparison with typical diode current densities ($0.1 - 10 \text{ kA/cm}^2$) yields the fact that leakage currents can be a significant loss mechanism for small barrier heights.

In the consideration above, we have assumed that electrons diffuse in the p-type region and have neglected any drift. However, if the p-type region has a substantial resistance, electron drift cannot be neglected. This drift will enhance the electron current. Furthermore, electrical contacts have been neglected. The minority carrier concentration at the contact–semiconductor interface can be assumed to be zero due to the high surface recombination velocity of such interfaces. Taking these effects into account, the leakage current was calculated by Ebeling (1993). If the distance of the contact from the active–barrier interface is denoted by x_p , the leakage current is given by

$$J_n = -eD_n n_B(0) \left(\sqrt{\frac{1}{L_n^2} + \frac{1}{L_{nf}^2}} \coth \sqrt{\frac{1}{L_n^2} + \frac{1}{L_{nf}^2}} x_p + \frac{1}{L_{nf}} \right) \quad (4.28)$$

where

$$L_{nf} = \frac{kT}{e} \frac{\sigma_p}{J_{\text{tot}}}, \quad (4.29)$$

σ_p is the conductivity of the p-type cladding region and J_{tot} is the total diode current density.

4.9 Carrier overflow in double heterostructures

The **overflow of carriers** from the active region into the confinement regions is another loss mechanism. Carrier overflow occurs at high injection current densities. As the injection current increases, the carrier concentration in the active region increases and the Fermi energy rises. For sufficiently high current densities, the Fermi energy will rise to the top of the barrier. The active region is flooded with carriers and a further increase in injection current density will *not* increase

the carrier concentration in the active region. As a result, the optical intensity saturates. At high injection current densities, carrier overflow occurs, even if the barriers are sufficiently high, so that carrier leakage over the barriers at low injection current densities can be neglected.

Consider a double heterostructure LED with an active region of thickness W_{DH} , as shown in Fig. 4.10. The rate equation of carrier supply to (by injection) and removal from the active region (by recombination) is given by

$$\frac{dn}{dt} = \frac{J}{e W_{\text{DH}}} - Bnp, \quad (4.30)$$

where B is the bimolecular recombination coefficient. For high injection densities, it is $n = p$. Solving Eq. (4.30) for n under steady-state conditions ($dn/dt = 0$), yields

$$n = \sqrt{\frac{J}{e B W_{\text{DH}}}}. \quad (4.31)$$

The carrier density increases with the current injected into the device. As a result, the Fermi energy rises. In the high-density approximation, the Fermi energy is given by

$$\frac{E_F - E_C}{kT} = \left(\frac{3\sqrt{\pi}}{4} \frac{n}{N_c} \right)^{2/3}. \quad (4.32)$$

At high injection levels, the Fermi energy rises and will eventually reach the top of the barrier. At that point, it is $E_F - E_C = \Delta E_C$. Using this value, the current density at which the active region overflows can be calculated from Eqs. (4.31) and (4.32) and one obtains

$$J_{\text{overflow}} = \left(\frac{4N_c}{3\sqrt{\pi}} \right)^2 \left(\frac{\Delta E_C}{kT} \right)^3 e B W_{\text{DH}} \quad (4.33)$$

Either the conduction band or the valence band well may overflow first depending on the effective density of states (N_c, N_v) and the band discontinuities ($\Delta E_C, \Delta E_V$).

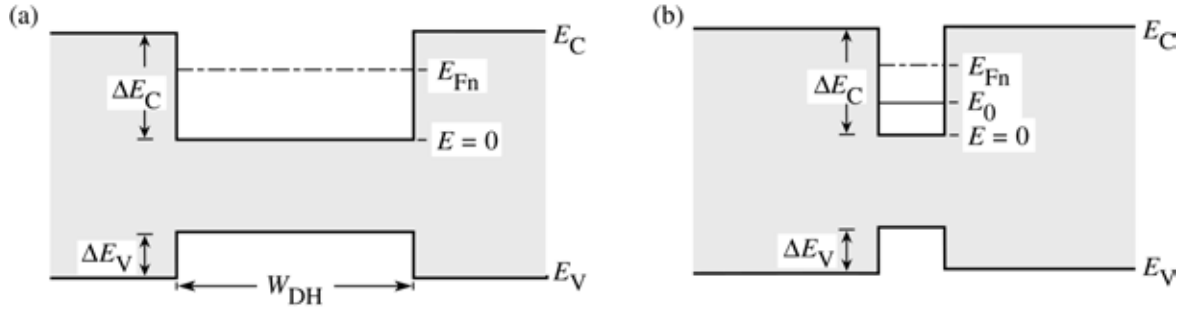


Fig. 4.10. Fermi level (E_{Fn}) and subband level (E_0) in (a) a double heterostructure and (b) a quantum well structure.

Exercise: Carrier overflow in a double heterostructure. Consider electrons in a GaAs double heterostructure with a barrier height of $\Delta E_C = 200$ meV and an active region thickness of $W_{DH} = 500$ Å. Calculate the current level at which the electron well overflows.

Solution: Using $N_c = 4.4 \times 10^{17} \text{ cm}^{-3}$ and $B = 10^{-10} \text{ cm}^3/\text{s}$, one obtains from Eq. (4.33) a current level of $J_{\max} = 3990 \text{ A/cm}^2$.

Generally, the problem of carrier overflow is more severe in structures with a small active-region volume. In particular, single-quantum-well structures and quantum-dot active regions have an inherently small volume. At a certain current density, the active region is filled with carriers, and the injection of additional carriers will not lead to an increase in the emitted light intensity.

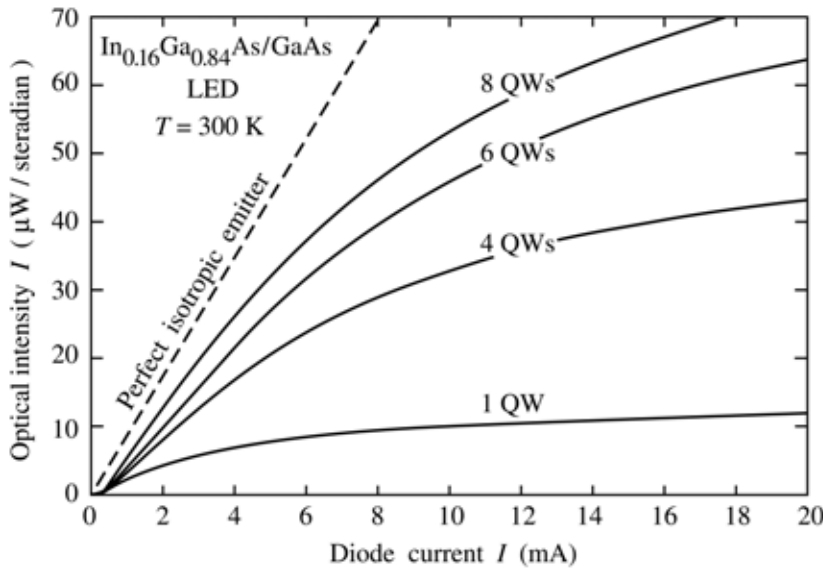


Fig. 4.11. Optical intensity emitted by $\text{In}_{0.16}\text{Ga}_{0.84}\text{As}/\text{GaAs}$ LEDs with active regions consisting of one, four, six, and eight quantum wells and theoretical intensity of a perfect isotropic emitter (dashed line) (after Hunt *et al.*, 1992).

Experimental results of an LED structure with one, four, six, and eight quantum wells (QWs) is shown in Fig. 4.11 (Hunt *et al.*, 1992). The light intensity for the single QW structure saturates at a low current level. As the number of quantum wells is increased, the current level at which saturation occurs increases, and the optical saturation intensity increases as well. The saturation of the light intensity displayed in Fig. 4.11 is caused by the overflow of carriers.

The calculation of the overflow current level is different for quantum well structures and bulk active regions. For quantum well structures, we must employ the two-dimensional (2D) density of states, rather than the 3D density of states that was used in the above calculation. The Fermi level in a QW with one quantized state with energy E_0 is given by

$$\frac{E_F - E_0}{kT} = \ln \left[\exp \left(\frac{n^{2D}}{N_c^{2D}} \right) - 1 \right] \quad (4.34)$$

where n^{2D} is the 2D carrier density per cm^2 and N_c^{2D} is the effective 2D density of states given by

$$N_c^{2D} = \frac{m^*}{\pi (h/2\pi)^2} kT . \quad (4.35)$$

Because we are dealing with high carrier densities, the high-degeneracy approximation can be employed and one obtains

$$E_F - E_0 = \frac{\pi (h/2\pi)^2}{m^*} n^{2D} . \quad (4.36)$$

Next, we write the rate equation for the quantum well. The rate equation of carrier supply to (by injection) and removal from the active region (by recombination) is given by

$$\frac{dn^{2D}}{dt} = \frac{J}{e} - B^{2D} n^{2D} p^{2D} , \quad (4.37)$$

where $B^{2D} \approx B / W_{\text{QW}}$ is the bimolecular recombination coefficient for a 2D structure. For high injection densities, it is $n^{2D} = p^{2D}$. Solving Eq. (4.37) for n^{2D} under steady-state conditions ($dn^{2D}/dt = 0$), yields

$$n^{2D} = \sqrt{\frac{J}{e B^{2D}}} = \sqrt{\frac{J W_{QW}}{e B}}. \quad (4.38)$$

At high injection levels, the Fermi energy will reach the top of the barrier. At that point, $E_F - E_0 = \Delta E_C - E_0$. The use of this value in Eq. (4.36) and subsequent elimination of n^{2D} from Eqs. (4.36) and (4.38) yields the current density at which the active region overflows

$$J|_{\text{overflow}} = \left[\frac{m^*}{\pi (h/2\pi)^2} (\Delta E_C - E_0) \right]^2 \frac{e B}{W_{QW}} \quad (4.39)$$

Thus overflow of the active region is a potential problem in double heterostructures as well as quantum well structures. In order to avoid this problem, high-current LEDs must employ *thick* double-heterostructure active regions, or *many* QWs of multiple QW (MQW) active regions, or a large injection (contact) area. By choosing these parameters, the volume of the active region can be designed in such a way that at the intended current density of operation, carrier overflow does not occur.

4.10 Electron blocking layers

Carriers tend to escape from the active layer of an LED into the confinement layers. The carrier escape can be substantial in double heterostructures with a low barrier height at the active–confinement interface. In addition, high temperatures promote carrier loss out of the active region due to the increase in carrier thermal energy.

The electron leakage current is larger than the hole leakage current due to the usually larger diffusion constant of electrons compared with holes in III–V semiconductors. To reduce carrier leakage out of the active region, carrier-blocking layers are used. In particular, ***electron-blocking layers*** or ***electron blockers*** are used in many LED structures to reduce electron escape out of the active region. Such electron blockers are regions with a high bandgap energy located at the confinement–active interface.

The band diagram of a GaInN LED with an electron-blocking layer is shown in Fig. 4.12. The LED has AlGaIn confinement layers and a GaInN/GaN multiple quantum well active region. An AlGaIn electron-blocking layer is included in the p-type confinement layer at the confinement–active interface. Figure 4.12 (a), showing the undoped structure, illustrates that the

AlGaN electron-blocking layer creates a barrier to current flow in both the conduction band as well as the valence band.

However, Fig. 4.12 (b), showing the doped structure, illustrates that the barrier in the valence band is screened by free carriers so that there is *no barrier* to the flow of *holes* in the p-type confinement layer. That is, the entire band discontinuity is located in the conduction band, i.e.

$$\text{barrier height for electrons} = E_{C,\text{confinement}} - E_{C,\text{active}} + \Delta E_g \quad (4.40)$$

where ΔE_g is the difference in bandgap energy between the confinement and the electron-blocking layer.

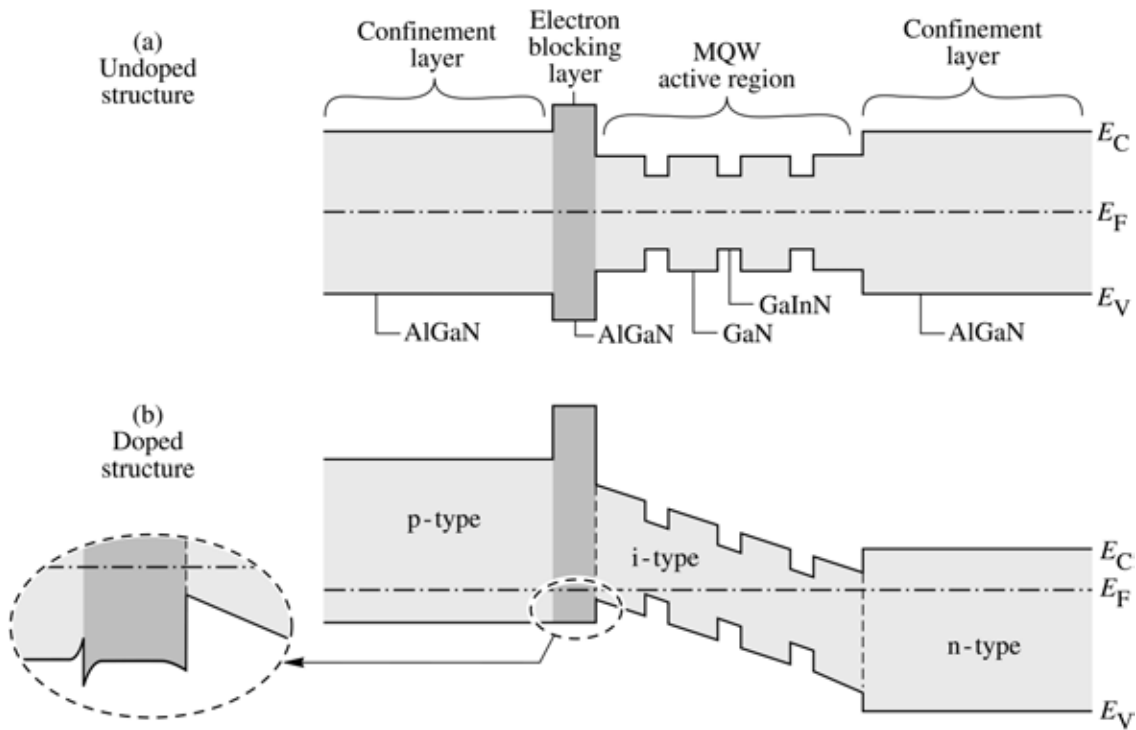


Fig. 4.12. Illustration of an AlGaN current-blocking layer in an AlGaN/GaN/GaInN multi-quantum well LED structure. (a) Band diagram without doping. (b) Band diagram with doping. The Al content in the electron-blocking layer is higher than in the p-type confinement layer.

The inset of Fig. 4.12 shows the valence band edge of the electron-blocking layer in greater detail. A potential spike (hole depletion layer in the electron blocker) and notch (hole accumulation layer in the p-type confinement layer) occur at the confinement–blocking layer interface. Holes must tunnel through the potential spike when propagating towards the active

region. Note that the valence band edge can be completely smoothed out by compositional grading at the confinement–blocking layer interface so that the electron blocking layer does not impede the hole flow at all.

4.11 Diode voltage

The energy of an injected electron is converted into optical energy upon electron–hole recombination. Thus, conservation of energy requires that the **drive voltage** or **forward voltage** of a light-emitting device is equal to (or larger than) the bandgap energy. The diode voltage is thus given by

$$V = h\nu / e \approx E_g / e. \quad (4.41)$$

There are several mechanisms causing the drive voltage to be slightly different from this value and these mechanisms will be discussed below.

First, if the diode has a significant series resistance, an additional voltage drop occurs. The additional resistance can be caused by (i) *contact resistance*, (ii) *resistances caused by abrupt heterostructures*, and (iii) *bulk resistance* occurring particularly in materials with low carrier concentrations or low carrier mobilities. A voltage drop of magnitude IR_s occurs at the series resistance thereby increasing the drive voltage.

Secondly, carrier energy may be lost upon injection into a quantum well structure or double heterostructure. An example of non-adiabatic injection is shown in Fig. 4.13, which shows a thin quantum well under forward bias conditions. The figure illustrates that upon injection into the quantum well, the electron loses energy $\Delta E_C - E_0$, where ΔE_C is the band discontinuity and E_0 is the energy of the lowest quantized state in the conduction-band quantum well. Similarly, the energy lost by holes is given by $\Delta E_V - E_0$, where ΔE_V is the band discontinuity and E_0 is the energy of the lowest state in the valence-band quantum well. Upon injection of carriers into the well, the carrier energy is dissipated by **phonon emission**, i.e. by conversion of the carrier energy to heat. The energy loss due to **non-adiabatic injection** of carriers is relevant in semiconductors with large band discontinuities, for example GaN and other group-III nitride materials.

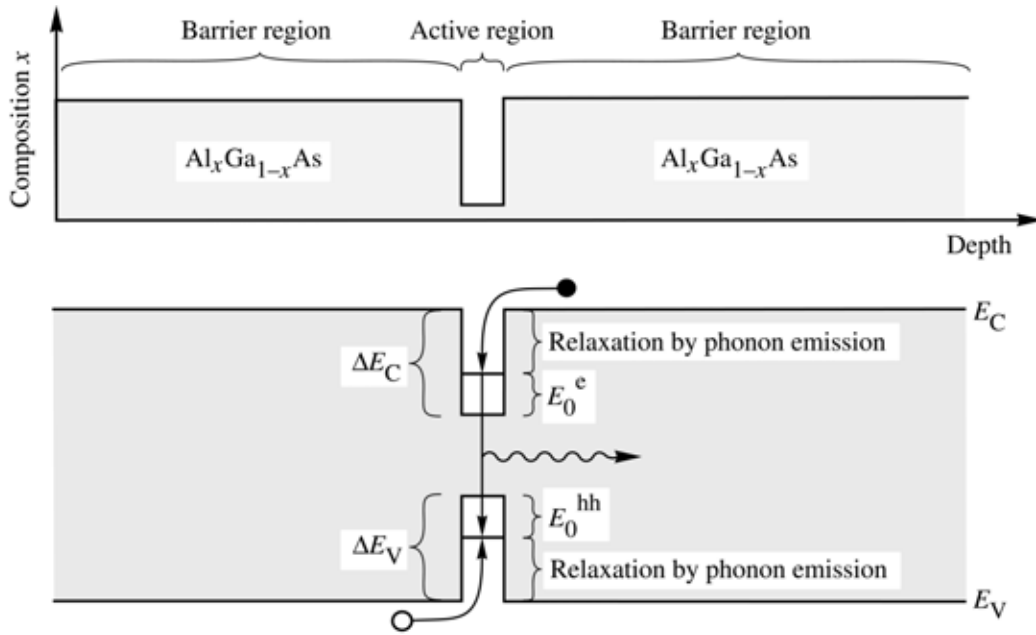


Fig. 4.13. (a) Chemical composition and (b) band diagram of a quantum well structure, illustrating the energy loss of carriers as they are captured into the quantum well.

Thus, the total voltage drop across a forward-biased LED is given by

$$V = \frac{E_g}{e} + I R_s + \frac{\Delta E_C - E_0}{e} + \frac{\Delta E_V - E_0}{e} \quad (4.42)$$

where the first summand on the right-hand side of the equation is the theoretical voltage minimum, the second summand is due to the series resistance in the device, and the third and fourth summands are due to non-adiabatic injection of carriers into the active region.

One finds experimentally that the diode voltage can be *slightly lower than the minimum* value predicted by Eq. (4.42), i.e. can be slightly lower than $E_g / e \approx h\nu / e$. Both electrons and holes carry, on average, the thermal energy kT . In a forward-biased p-n junction, high-energy carriers are more likely than low-energy carriers to diffuse over to the side of opposite conductivity type where they recombine. At room temperature, $4kT$ amounts to a voltage of about 100 mV. In low-resistance devices, the diode voltage can be 100 – 200 mV *lower* than $h\nu / e$. For example, in forward-biased GaAs LEDs ($E_g = 1.42$ eV), some photon emission with $h\nu = 1.42$ eV is observed at diode voltages of about 1.32 V, i.e. lower than the photon energy.

Exercise. Drive voltages of LEDs. Calculate the approximate forward diode voltage of LEDs

emitting in the blue, green, and red parts of the visible spectrum. Also calculate the forward diode voltage of LEDs emitting at 870 nm and 1.55 μm .

Solution:

<i>Emission color</i>	<i>Wavelength</i>	<i>Photon energy</i>	<i>Drive voltage</i>
Blue	470 nm	2.6 eV	2.6 V
Green	550 nm	2.2 eV	2.2 V
Red	650 nm	1.9 eV	1.9 V
IR	870 nm	1.4 eV	1.4 V
IR	1550 nm	0.8 eV	0.8 V

4.12 Temperature dependence of diode voltage

The I – V characteristic of an ideal p–n junction diode is given by the Shockley equation

$$I = I_s \left(e^{eV/kT} - 1 \right) \quad (4.43)$$

where I_s is the saturation current. For non-degenerate semiconductors and under forward bias conditions, the Shockley equation can be written as

$$I = I_s^* \left(e^{(eV - E_g)/kT} - 1 \right) \quad (4.44)$$

where

$$I_s^* = eA \left(\frac{D_p N_c N_v}{L_p N_D} + \frac{D_n N_c N_v}{L_n N_A} \right). \quad (4.45)$$

Let us assume that the diode is driven with a constant current I . As the temperature changes, the voltage drop across the diode will change. The temperature dependence of the diode voltage can be determined by solving Eq. (4.44) for V . One obtains

$$V(T) = \frac{kT}{e} \ln \frac{I}{I_s^*} + \frac{E_g(T)}{e}. \quad (4.46)$$

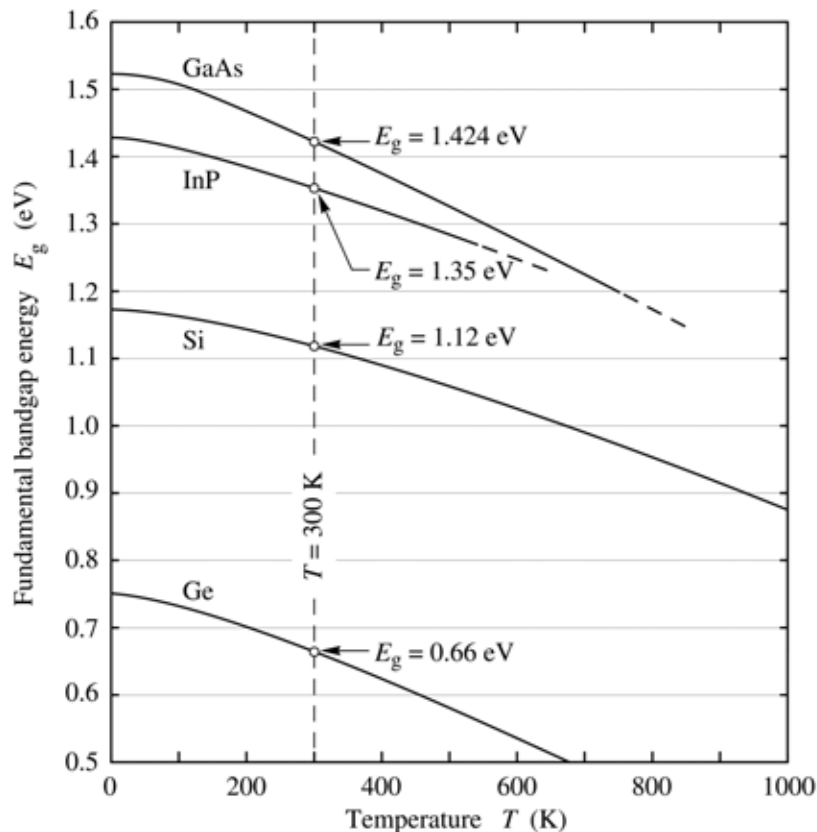
The first summand on the right-hand side of the equation is due to the changes of the Fermi

level with temperature. As the temperature increases, the Fermi level moves towards the middle of the gap so that the Fermi level separation between the p- and n-sides becomes smaller. As a result, a smaller voltage is required for a given current. The contribution of the first term is generally smaller than the contribution of the second term.

The second summand on the right-hand side of the equation is due to the changes in bandgap energy. As the temperature increases, the energy gap of semiconductors decreases. The temperature dependence of the energy gap of a semiconductor can be expressed by the formula

$$E_g = E_g|_{T=0\text{ K}} - \frac{\alpha T^2}{T + \beta} \quad (4.47)$$

where α and β are fitting parameters. The bandgap energy versus temperature for several semiconductors is shown in Fig. 4.14 along with the values for α and β . Assuming that the change in bandgap energy is the dominant factor in determining the temperature dependence of the diode voltage, the temperature dependence of the forward voltage follows directly from the temperature dependence of the bandgap energy.



$$E_g = E_g(0\text{ K}) - \frac{\alpha T^2}{T + \beta}$$

	$E_g(0\text{ K})$	$\alpha (10^{-4} \frac{\text{eV}}{\text{K}})$	$\beta (\text{K})$
GaAs	1.519	5.41	204
InP	1.425	4.50	327
Si	1.170	4.73	636
Ge	0.744	4.77	235

Fig. 4.14. Fundamental bandgap energy of GaAs, InP, Si, and Ge as a function of temperature. The bandgap energy is approximated by a parabolic equation with the fitting parameters α and β .

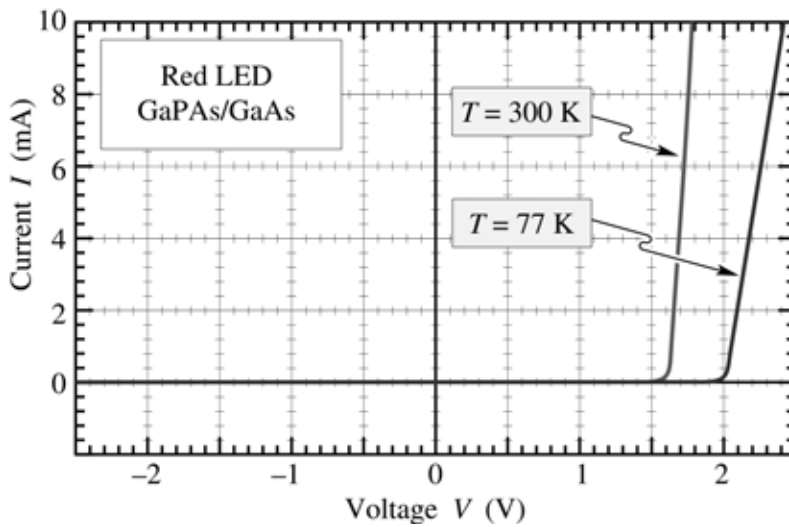


Fig. 4.15. Current–voltage characteristic of a GaAsP/GaAs LED emitting in the red part of the visible spectrum, measured at 77 and 300 K. The threshold voltages are 2.0 and 1.6 V, at 77 and 300 K, respectively.

The temperature dependence of an GaPAs/GaAs LED is illustrated in Fig. 4.15, which shows the I – V characteristic at 77 K and at room temperature. Inspection of the figure reveals that the threshold voltage as well as the series resistance of the diode increases as the diode is cooled. If the device were driven at a constant voltage, e.g. 1.9 V, a large current change would result from a change in temperature.

Exercise. Temperature dependence of the diode forward voltage. Calculate the linear temperature coefficient of the forward voltage of a GaAs LED at room temperature. Assume that the change in forward voltage is mainly due to the temperature dependence of the bandgap energy. What is the decrease in forward voltage if a GaAs diode temperature is increased from 20 to 40 °C?

Solution: The temperature dependence of the energy gap is given by

$$\frac{dE_g}{dT} = - \frac{\alpha T (T + 2\beta)}{(T + \beta)^2}. \quad (4.48)$$

For GaAs with $\alpha = 5.41 \times 10^{-4}$ eV/K and $\beta = 204$ K, one obtains at room temperature $dE_g / dT = -0.45$ meV/K. The decrease in diode voltage for the 20 °C temperature increase is $\Delta V = 9$ mV.

4.13 Constant-current and constant-voltage DC drive circuits

Different considerations play a role in designing the DC drive circuit of an LED operated under

steady-state conditions. These considerations include the simplicity and cost of the drive circuit, the power efficiency, and the compensation of the temperature dependence of the light intensity emitted from the LED.

A simple drive circuit is a **constant-voltage supply** such as a battery or the rectified AC output of a transformer. There are, however, two drawbacks to constant-voltage drives of LEDs. *First*, the diode current depends exponentially on the voltage, so a small variation in the drive voltage results in a large change in the current. *Secondly*, the threshold voltage of a diode depends on temperature, so any temperature change results in a significant change in current. A constant-voltage operating characteristic of a diode is shown in Fig. 4.16.

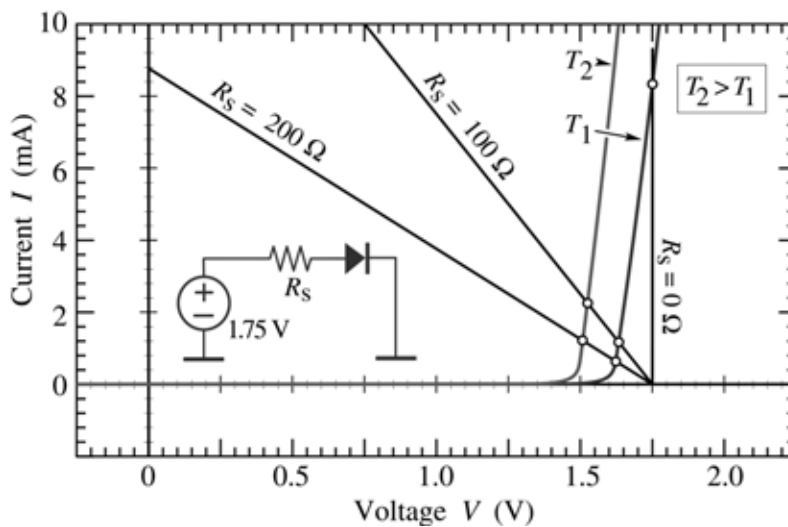


Fig. 4.16. LED drive circuit with series resistance R_s . The intersection between the diode I - V characteristics and the load lines are the points of operation. Small series resistances result in an increased diode current at high temperatures, thus allowing for compensation of a lower LED radiative efficiency.

If a resistor is connected in series with the diode, the strong temperature dependence of the diode current is reduced, as illustrated in Fig. 4.16. The series resistance together with the temperature dependence of the diode determine the temperature coefficient of the diode current.

The series resistance can be used to reduce the temperature dependence of the emission intensity, which occurs when the diode is driven with a constant current. The light intensity of LEDs generally decreases with increasing temperature due to non-radiative recombination. In addition, the threshold voltage decreases with increasing temperature. However, the diode current increases as the temperature increases, as shown in Fig. 4.16. Thus a series resistance can be used to compensate for the emission intensity decrease at elevated temperatures. It should be noted that the overall electrical-to-optical power efficiency drops due to the power consumed in the resistor.

The temperature dependence of LED intensity is an important factor for LEDs used in

outdoor applications. On hot summer days, the temperature and the ambient light intensity are high. Owing to the high temperature, the LED intensity drops. In addition, high brightness is required due to high ambient light levels. This effect can be compensated for by driving the LEDs with a higher current as the temperature increases.

A constant-current drive circuit can consist of a transistor with the LED as a load. A constant-current drive circuit allows one to drive the LED irrespective of the diode threshold voltage and the diode temperature. However, a constant-current drive circuit does not compensate for the decrease of LED emission at elevated temperatures.

References

- Ebeling K. J. “Integrated opto-electronics” Chap. 9 (Springer, Berlin, 1993)
- Emerson D., Abare A., Bergmann M., Slater D., and Edmond J. “Development of deep UV III–N optical sources” 7th International Workshop on Wide-Bandgap III–Nitrides, Richmond VA, March (2002)
- Hunt N. E. J., Schubert E. F., Sivco D. L., Cho A. Y., and Zyzdik G. J. “Power and efficiency limits in single-mirror light-emitting diodes with enhanced intensity” *Electronics Lett.* **28**, 2169 (1992)
- Krames M. R. *et al.* “High-brightness AlGaInP light-emitting diodes” *Proceedings of SPIE* **3938**, 2 (2000)
- Rhoderick E. H. and Williams R. H. “Metal–semiconductor contacts” (Clarendon Press, Oxford, UK, 1988)
- Schubert E. F., Tu L.-W., Zyzdik G. J., Kopf R. F., Benvenuti A., and Pinto M. R. “Elimination of heterojunction band discontinuities by modulation doping” *Appl. Phys. Lett.* **60**, 466 (1992)

5

LED basics: optical properties

5.1 Internal, extraction, external, and power efficiency

The active region of an ideal LED emits *one photon* for *every electron* injected. Each charge quantum-particle (electron) produces one light quantum-particle (photon). Thus the ideal active region of an LED has a *quantum efficiency* of unity. The **internal quantum efficiency** is defined as

$$\eta_{\text{int}} = \frac{\text{\# of photons emitted from active region per second}}{\text{\# of electrons injected into LED per second}} = \frac{P_{\text{int}} / (h\nu)}{I / e} \quad (5.1)$$

where P_{int} is the optical power emitted from the active region and I is the injection current.

Photons emitted by the active region should escape from the LED die. In an ideal LED, all photons emitted by the active region are also emitted into free space. Such an LED has unity *extraction efficiency*. However, in a real LED, not all the power emitted from the active region is emitted into free space. Some photons may never leave the semiconductor die. This is due to several possible loss mechanisms. For example, light emitted by the active region can be reabsorbed in the substrate of the LED, assuming that the substrate is absorbing at the emission wavelength. Light may be incident on a metallic contact surface and be absorbed by the metal. In addition, the phenomenon of *total internal reflection*, also referred to as the *trapped light phenomenon* reduces the ability of the light to escape from the semiconductor. The light **extraction efficiency** is defined as

$$\eta_{\text{extraction}} = \frac{\text{\# of photons emitted into free space per second}}{\text{\# of photons emitted from active region per second}} = \frac{P / (h\nu)}{P_{\text{int}} / (h\nu)} \quad (5.2)$$

where P is the optical power emitted into free space.

The extraction efficiency can be a severe limitation for high-performance LEDs. It is quite difficult to increase the extraction efficiency beyond 50 % without resorting to highly sophisticated and costly device processes.

The *external quantum efficiency* is defined as

$$\eta_{\text{ext}} = \frac{\text{\# of photons emitted into free space per second}}{\text{\# of electrons injected into LED per second}} = \frac{P / (h \nu)}{I / e} = \eta_{\text{int}} \eta_{\text{extraction}} \quad (5.3)$$

The external quantum efficiency gives the ratio of the number of useful light particles to the number of injected charge particles.

The *power efficiency* is defined as

$$\eta_{\text{power}} = \frac{P}{IV} \quad (5.4)$$

where IV is the electrical power provided to the LED. Informally, the power efficiency is also called the “*wallplug efficiency*”.

Exercise. LED efficiency. Consider an LED with a threshold voltage of $V_{\text{th}} = E_g / e = 2 \text{ V}$ with a differential resistance of $R_s = 20 \text{ } \Omega$, so that the I – V characteristic in the forward direction is given by $V = V_{\text{th}} + I R_s$. When the device is operated at 20 mA it emits a light power of 4 mW of energy $h \nu = E_g$. Determine the internal quantum efficiency, the external quantum efficiency, and the power efficiency, assuming that the extraction efficiency is 50 %.

5.2 Emission spectrum

The physical mechanism by which semiconductor LEDs emit light is spontaneous recombination of electron–hole pairs and simultaneous emission of photons. The spontaneous emission process is fundamentally different from the stimulated emission process occurring in semiconductor lasers and super-luminescent LEDs. Spontaneous recombination has certain characteristics that determine the optical properties of LEDs. The properties of spontaneous emission in LEDs will be discussed in this section.

An electron–hole recombination process is illustrated schematically in Fig. 5.1. Electrons in the conduction band and holes in the valence band are assumed to have the parabolic dispersion

relations

$$E = E_C + \frac{(h/2\pi)^2 k^2}{2m_e^*} \quad (\text{for electrons}) \quad (5.5)$$

and

$$E = E_V - \frac{(h/2\pi)^2 k^2}{2m_h^*} \quad (\text{for holes}) \quad (5.6)$$

where m_e^* and m_h^* are the electron and hole effective masses, \hbar is Planck's constant divided by 2π , k is the carrier wave number, and E_V and E_C are the valence and conduction band edges, respectively.

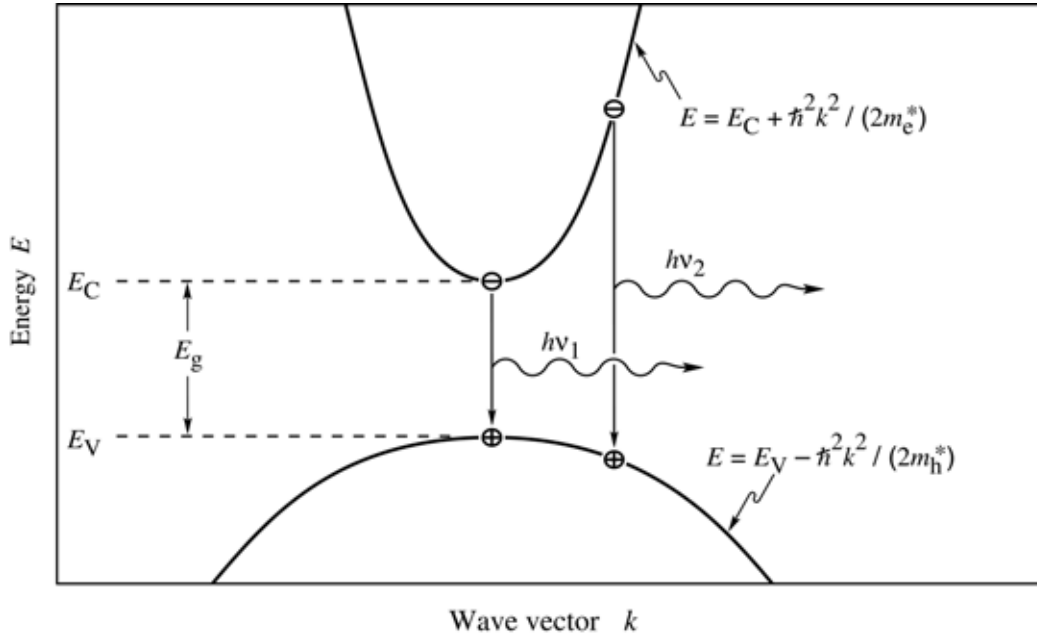


Fig. 5.1. Parabolic electron and hole dispersion relations showing “vertical” electron-hole recombination and photon emission.

The requirement of energy and momentum conservation leads to further insight into the radiative recombination mechanism. It follows from the Boltzmann distribution that electrons and holes have an average kinetic energy of kT . Energy conservation requires that the photon energy is given by the difference between the electron energy, E_e , and the hole energy, E_h , i.e.

$$h\nu = E_e - E_h \approx E_g. \quad (5.7)$$

The photon energy is approximately equal to the bandgap energy, E_g , if the thermal energy is small compared with the bandgap energy $kT \ll E_g$. Thus the desired emission wavelength of an LED can be attained by choosing a semiconductor material with an appropriate bandgap energy. For example, GaAs has a bandgap energy of 1.42 eV at room temperature and thus GaAs LEDs emit at the infrared wavelength of 870 nm.

It is helpful to compare the average carrier momentum with the photon momentum. A carrier with kinetic energy kT and effective mass m^* has the momentum

$$p = m^* v = \sqrt{2m^* \frac{1}{2} m^* v^2} = \sqrt{2m^* kT} . \quad (5.8)$$

The momentum of a photon with energy E_g can be derived from the de Broglie relation

$$p = (h/2\pi) k = h\nu/c = E_g/c . \quad (5.9)$$

Calculation of the carrier momentum (using Eq. 5.8) and the photon momentum (using Eq. 5.9) yields that the carrier momentum is *orders of magnitude larger* than the photon momentum. Therefore the electron momentum cannot change significantly during the transition from the conduction to the valence band. The transitions are therefore “vertical” as shown in Fig. 5.1, i.e. electrons only recombine with holes that have the same momentum or k value.

Using the requirement that electron and hole momenta are the same, the photon energy can be written as the *joint dispersion relation*

$$h\nu = E_C + \frac{(h/2\pi)^2 k^2}{2 m_e^*} - E_V + \frac{(h/2\pi)^2 k^2}{2 m_h^*} = E_g + \frac{(h/2\pi)^2 k^2}{2 m_r^*} \quad (5.10)$$

where m_r^* is the reduced mass given by

$$\frac{1}{m_r^*} = \frac{1}{m_e^*} + \frac{1}{m_h^*} . \quad (5.11)$$

Using the joint dispersion relation, the joint density of states can be calculated and one obtains

$$\rho(E) = \frac{1}{2\pi^2} \left(\frac{2m_r^*}{(h/2\pi)^2} \right)^{3/2} \sqrt{E - E_g} . \quad (5.12)$$

The distribution of carriers in the allowed bands is given by the Boltzmann distribution, i.e.

$$f_B(E) = e^{-E/(kT)} . \quad (5.13)$$

The *emission intensity* as a function of energy is proportional to the product of Eqs. (5.12) and (5.13),

$$I(E) \propto \sqrt{E - E_g} e^{-E/(kT)} \quad (5.14)$$

The lineshape of an LED, as given by Eq. (5.14), is shown in Fig. 5.2. The maximum emission intensity occurs at

$$E = E_g + \frac{1}{2} kT \quad (5.15)$$

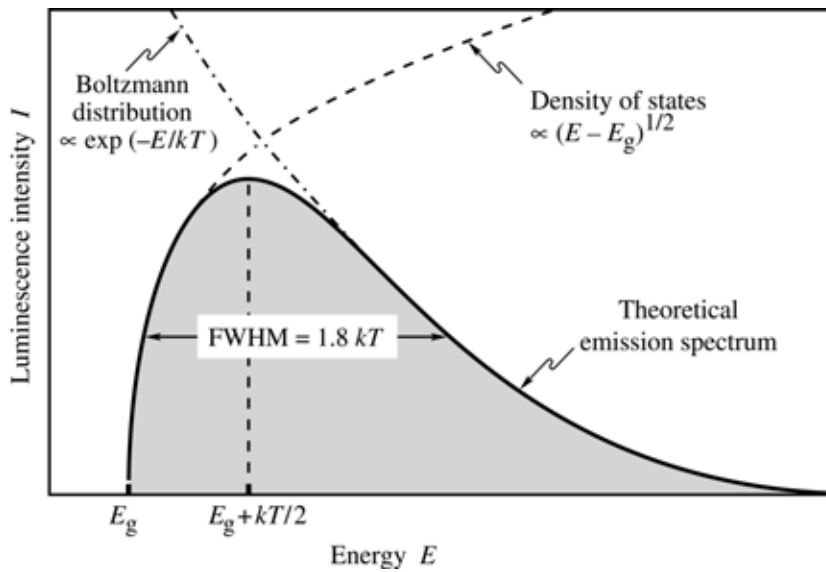


Fig. 5.2. Theoretical emission spectrum of an LED. The full width at half maximum (FWHM) of the emission line is $1.8 kT$.

The full width at half maximum of the emission is

$$\Delta E = 1.8 kT \quad (5.16)$$

For example, the theoretical room-temperature linewidth of a GaAs LED emitting at 870 nm is $\Delta E = 46$ meV or $\Delta\lambda = 28$ nm.

The spectral linewidth of LED emission is important in several respects. *First*, the linewidth of an LED emitting in the visible range is relatively narrow compared with the range of the entire visible spectrum. The LED emission is even narrower than the spectral width of a single color as perceived by the human eye. For example, *red* colors range in wavelength from 625 to 730 nm, which is much wider than the typical emission spectrum of an LED. Therefore, LED emission is perceived by the human eye as *monochromatic*.

Secondly, optical fibers are dispersive, which leads to a range of propagation velocities for a light pulse consisting of a range of wavelengths. The material dispersion in optical fibers limits the “*bit rate \times distance product*” achievable with LEDs.

The spontaneous lifetime of carriers in LEDs in direct-gap semiconductors is of the order of 1 – 100 ns depending on the active region doping concentration (or carrier concentrations) and the material quality. Thus, modulation speeds up to 1 Gbit/s are attainable with LEDs.

5.3 The light escape cone

Light generated inside a semiconductor cannot escape from the semiconductor if it is totally internally reflected at the semiconductor–air interface. If the angle of incidence of a light ray is close to normal incidence, light can escape from the semiconductor. However, total internal reflection occurs for light rays with oblique and grazing-angle incidence. Total internal reflection reduces the external efficiency significantly, in particular for LEDs consisting of high-refractive-index materials.

Assume that the angle of incidence in the semiconductor at the semiconductor–air interface is given by ϕ . Then the angle of incidence of the refracted ray, Φ , can be inferred from Snell’s law

$$\bar{n}_s \sin \phi = \bar{n}_{\text{air}} \sin \Phi \quad (5.17)$$

where \bar{n}_s and \bar{n}_{air} are the refractive indices of the semiconductor and air, respectively. The critical angle for total internal reflection is obtained using $\Phi = 90^\circ$, as illustrated in Fig. 5.3 (a).

Using Snell’s law, one obtains

$$\sin \phi_c = (\bar{n}_{\text{air}} / \bar{n}_s) \sin 90^\circ = \bar{n}_{\text{air}} / \bar{n}_s . \quad (5.18)$$

The refractive indices of semiconductors are usually quite high. For example, GaAs has a refractive index of 3.4. Thus, according to Eq. (5.18), the critical angle for total internal reflection is quite small. In this case, we can use the approximation $\sin \phi_c \approx \phi_c$. The critical angle for total internal reflection is then given by

$$\phi_c \approx \bar{n}_{\text{air}} / \bar{n}_s . \quad (5.19)$$

The angle of total internal reflection defines the **light-escape cone**. Light emitted into the cone can escape from the semiconductor, whereas light emitted outside the cone is subject to total internal reflection.

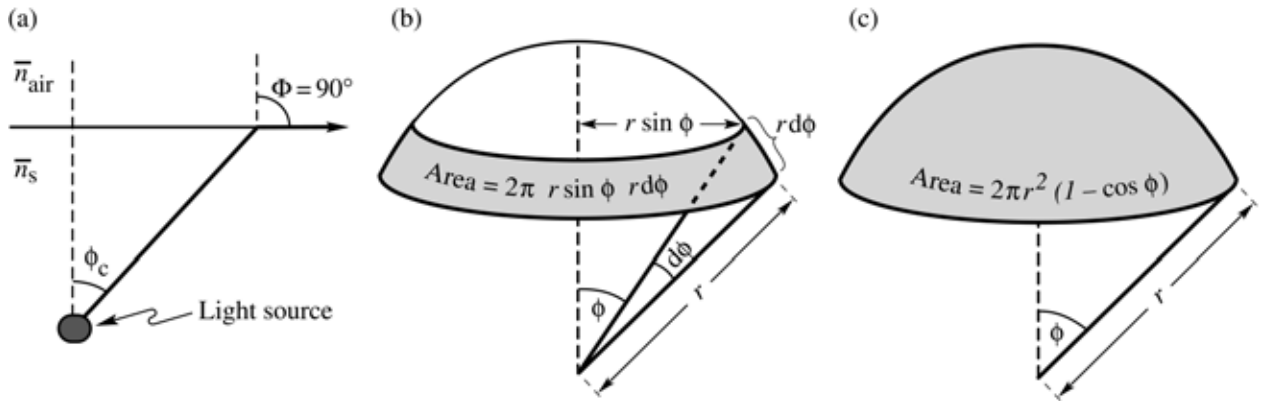


Fig. 5.3. (a) Definition of the escape cone by the critical angle ϕ_c . (b) Area element dA . (c) Area of calotte-shaped section of the sphere defined by radius r and angle ϕ_c .

Next, we calculate the surface area of the spherical cone with radius r in order to determine the total fraction of light that is emitted into the light escape cone. The surface area of the calotte-shaped surface shown in Figs. 5.3 (b) and (c) is given by the integral

$$A = \int dA = \int_{\phi=0}^{\phi_c} 2\pi r \sin \phi \, r \, d\phi = 2\pi r^2 (1 - \cos \phi_c) . \quad (5.20)$$

Let us assume that light is emitted from a point-like source in the semiconductor with a total power of P_{source} . Then the power that can escape from the semiconductor is given by

$$P_{\text{escape}} = P_{\text{source}} \frac{2 \pi r^2 (1 - \cos \phi_c)}{4 \pi r^2} \quad (5.21)$$

where $4 \pi r^2$ is the entire surface area of the sphere with radius r .

The calculation indicates that only a fraction of the light emitted inside a semiconductor can escape from the semiconductor. This fraction is given by

$$\boxed{\frac{P_{\text{escape}}}{P_{\text{source}}} = \frac{1}{2} (1 - \cos \phi_c)} \quad (5.22)$$

Because the critical angle of total internal reflection for high-index materials is relatively small, the cosine term can be expanded into a power series. Neglecting higher-than-second-order terms yields

$$\frac{P_{\text{escape}}}{P_{\text{source}}} \approx \frac{1}{2} \left[1 - \left(1 - \frac{\phi_c^2}{2} \right) \right] = \frac{1}{4} \phi_c^2. \quad (5.23)$$

Using the approximation of Eq. (5.19), one obtains

$$\boxed{\frac{P_{\text{escape}}}{P_{\text{source}}} \approx \frac{1}{4} \frac{\bar{n}_{\text{air}}^2}{\bar{n}_s^2}} \quad (5.24)$$

The escape problem is a significant problem for high-efficiency LEDs. In most semiconductors, the refractive index is quite high (> 2.5) and thus only a few percent of the light generated in the semiconductor can escape from a planar LED. The problem is less significant in semiconductors with a small refractive index and for polymers, which have refractive indices of the order of 1.5.

Exercise: Light escape from planar GaAs, GaN, and polymer LED structures. The refractive indices of GaAs, GaN, and light-emitting polymers are 3.4, 2.5, and 1.5, respectively. Calculate the critical angle of total internal reflection for GaAs, GaN, and for polymers. Also calculate the fraction of light power that can escape from a planar GaAs and GaN semiconductor structures and a polymer LED structure.

What improvement can be attained if a planar GaAs LED is encapsulated in a transparent polymer of refractive index 1.5, if the reflection at the polymer–air interface is neglected?

Solution:

Critical angle for total internal reflection:

GaAs $\phi_c = 17.1^\circ$ GaN $\phi_c = 23.6^\circ$ Polymer $\phi_c = 41.8^\circ$.

Fraction of light that can escape:

GaAs 2.21 % GaN 4.18 % Polymer 12.7 % .

Improvement of the GaAs planar LED due to polymer encapsulation: 232 %.

5.4 The lambertian emission pattern

The index contrast between the light-emitting material and the surrounding material leads to a non-isotropic emission pattern. For high-index light-emitting materials with a planar surface, a lambertian emission pattern is obtained. Figure 5.4 illustrates a point-like light source located a short distance below a semiconductor–air interface. Consider a light ray emitted from the source at an angle ϕ with respect to the surface normal. The light ray is refracted at the semiconductor–air interface and the refracted light ray has an angle Φ with respect to the surface normal. The two angles are related by Snell's law, which, for small angles of ϕ , can be written as

$$\bar{n}_s \phi = \bar{n}_{\text{air}} \sin \Phi . \quad (5.25)$$

Light emitted into the angle $d\phi$ in the semiconductor is emitted into the angle $d\Phi$ in air as shown in Fig. 5.4 (a). Differentiating the equation with respect to Φ and solving the resulting equation for $d\Phi$ yields

$$d\Phi = \frac{\bar{n}_s}{\bar{n}_{\text{air}}} \frac{1}{\cos \Phi} d\phi . \quad (5.26)$$

Power conservation requires that the optical power emitted into the angle $d\phi$ in the semiconductor be equal to the optical power emitted into the angle $d\Phi$ in air, i.e.

$$I_s dA_s = I_{\text{air}} dA_{\text{air}} \quad (5.27)$$

where I_s and I_{air} are the light intensities (measured in units of W/m^2) in the semiconductor and air, respectively. Owing to the cylindrical symmetry of the emission pattern we choose the area element shown in Fig. 5.4 (b). The area element in air is given by

$$dA_{\text{air}} = 2\pi r \sin \Phi r d\Phi . \quad (5.28)$$

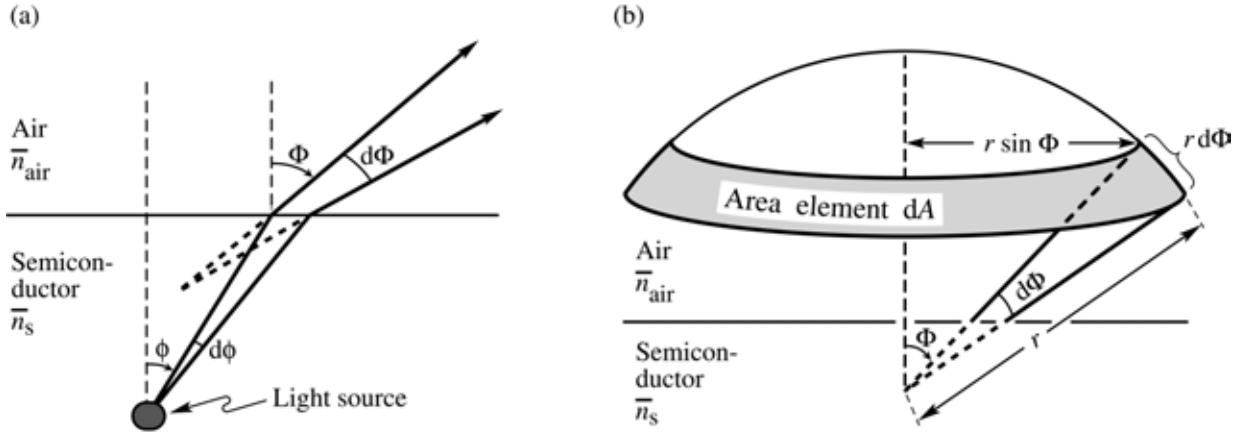


Fig. 5.4. Geometrical model used to derive the Lambertian emission pattern. (a) The light emitted into angle $d\phi$ inside the semiconductor is emitted into the angle $d\Phi$ in air. (b) Illustration of the area element dA of the calotte-shaped section of the sphere.

Using Eqs. (5.26) and (5.27) yields

$$dA_{\text{air}} = 2\pi r^2 \frac{\bar{n}_s^2}{\bar{n}_{\text{air}}^2} \frac{1}{\cos \Phi} \phi d\phi . \quad (5.29)$$

Similarly, the surface element in the semiconductor is given by

$$dA_s = 2\pi r \sin \phi r d\phi \approx 2\pi r^2 \phi d\phi . \quad (5.30)$$

The light intensity in the semiconductor at a distance r from the light source is given by the total source power divided by the surface area of a sphere with radius r , i.e.

$$I_s = \frac{P_{\text{source}}}{4\pi r^2} . \quad (5.31)$$

The light intensity in air can then be inferred from Eqs. (5.27), (5.29), (5.30), and (5.31). One obtains the *lambertian emission pattern* given by

$$I_{\text{air}} = \frac{P_{\text{source}}}{4\pi r^2} \frac{\bar{n}_{\text{air}}^2}{\bar{n}_s^2} \cos \Phi \quad (5.32)$$

The lambertian emission pattern follows a cosine dependence on the angle Φ . The intensity is highest for emission normal to the semiconductor surface, i.e. for $\Phi = 0^\circ$. At an angle of $\Phi = 60^\circ$, the intensity decreases to half of its maximum value. The lambertian emission pattern is shown schematically in Fig. 5.5.

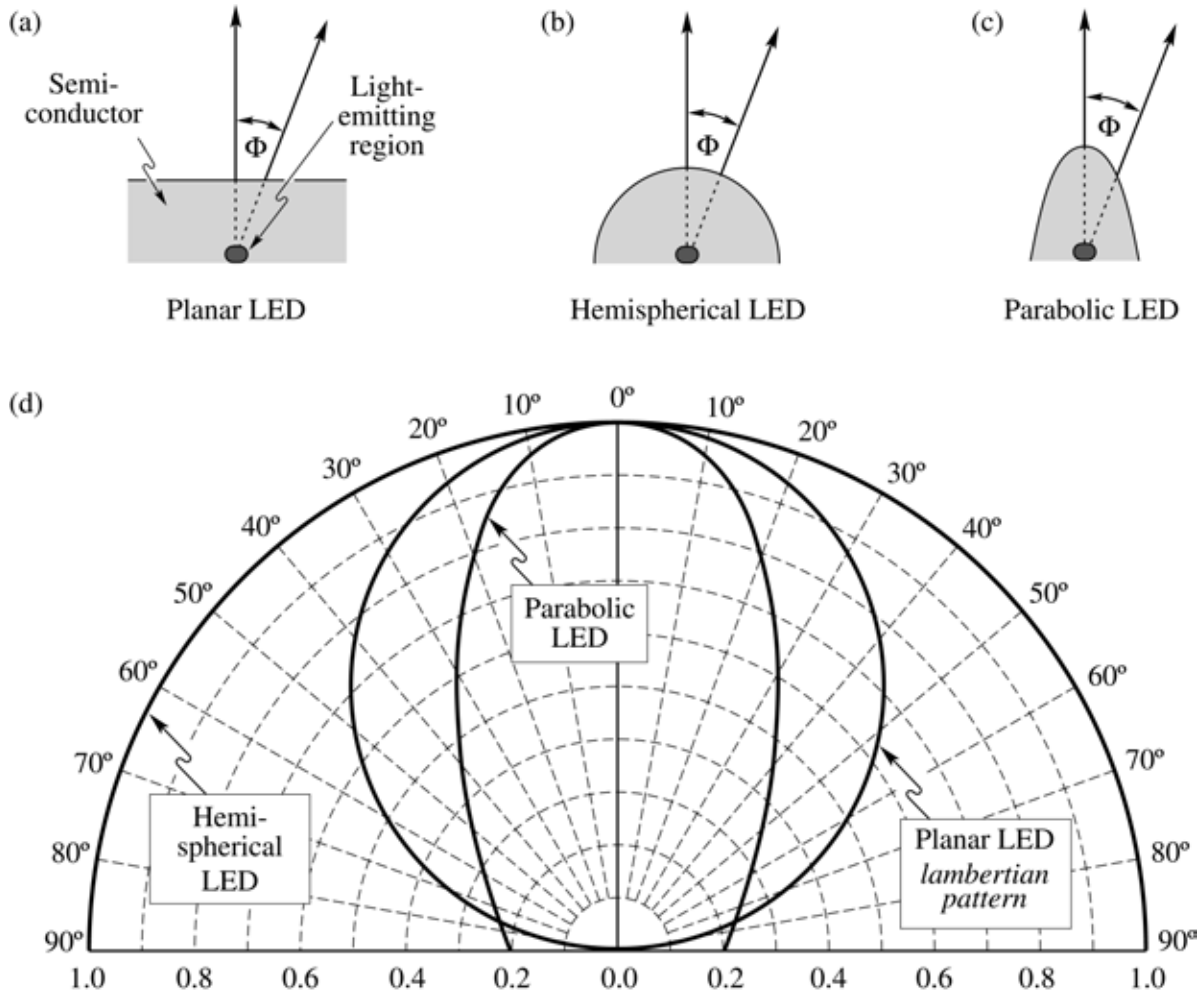


Fig. 5.5. Light-emitting diodes with (a) planar, (b) hemispherical, and (c) parabolic surfaces. (d) Far-field patterns of the different types of LEDs. At an angle of $\Phi = 60^\circ$, the lambertian emission pattern decreases to 50 % of its maximum value occurring at $\Phi = 0^\circ$. The three emission patterns are normalized to unity intensity at $\Phi = 0^\circ$.

Several other surface shapes are also shown in Fig. 5.5. These non-planar surfaces exhibit various emission patterns. An isotropic emission pattern is obtained for hemispherically shaped LEDs, which have the light-emitting region in the center of the sphere. A strongly directed emission pattern can be obtained in LEDs with parabolically shaped surfaces. However, both hemispherical as well as parabolic surfaces are difficult to fabricate.

The total power emitted into air can be calculated by integrating the intensity over the entire hemisphere. The total power is then given by

$$P_{\text{air}} = \int_{\Phi=0^{\circ}}^{90^{\circ}} I_{\text{air}} 2\pi r \sin \Phi \, r d\Phi . \quad (5.33)$$

By using the lambertian emission pattern for I_{air} in Eq. (5.33) and using $\cos \Phi \sin \Phi = (1/2) \sin (2 \Phi)$, the integral can be calculated to yield

$$P_{\text{air}} = \frac{P_{\text{source}}}{4} \frac{\bar{n}_{\text{air}}^2}{\bar{n}_s^2} . \quad (5.34)$$

This result is identical to Eq. (5.24). This is not surprising because the light power that escapes from the semiconductor (P_{escape}) must be identical to the power in air (P_{air}).

In the calculation above, **Fresnel reflection** at the semiconductor–air interface has been neglected. At normal incidence, the Fresnel power transmittance is given by

$$T = 1 - R = 1 - \left(\frac{\bar{n}_s - \bar{n}_{\text{air}}}{\bar{n}_s + \bar{n}_{\text{air}}} \right)^2 = \frac{4 \bar{n}_s \bar{n}_{\text{air}}}{(\bar{n}_s + \bar{n}_{\text{air}})^2} . \quad (5.35)$$

Fresnel reflection losses must be taken into account in a rigorous calculation.

Exercise: LED-to-fiber coupling efficiency. Consider a GaAs LED with a point-like light-emitting region located in close proximity to the planar GaAs LED surface. An optical fiber has an acceptance angle of 12° in air. What fraction of the light emitted by the active region can be coupled into the fiber? Assume a GaAs refractive index of $\bar{n}_s = 3.4$. Neglect Fresnel reflection losses at the semiconductor–air and air–fiber interfaces.

Solution: The acceptance angle in the semiconductor is obtained from Snell's law and is 3.5° .

Thus 0.093 % of the power emitted by the active region can be coupled into the fiber.

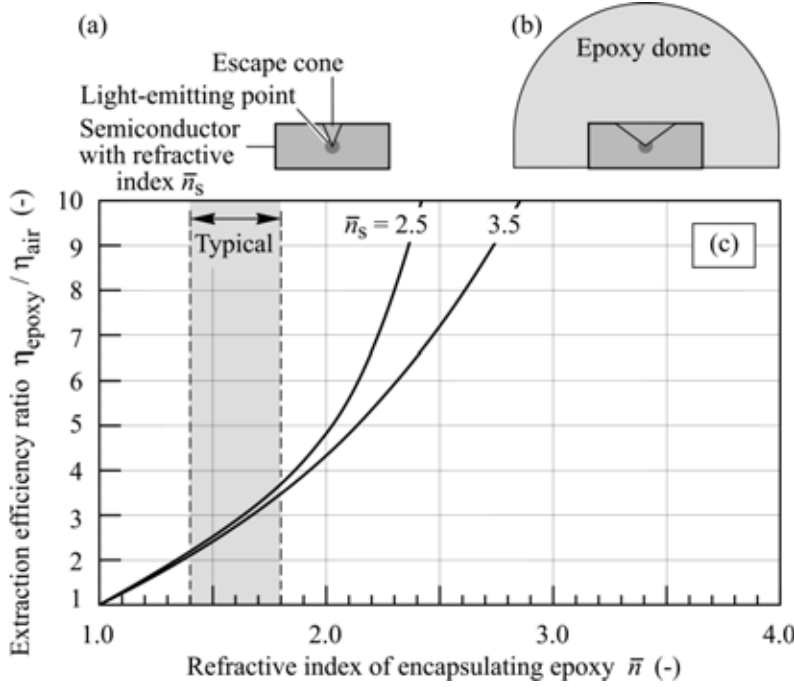


Fig. 5.6. (a) LED without and (b) with dome-shaped epoxy encapsulant. A larger escape angle is obtained for the LED with an epoxy dome. (c) Calculated ratio of light extraction efficiency emitted through the top surface of a planar LED with and without an epoxy dome. The refractive indices of typical epoxies range between 1.4 and 1.8 (adopted from Nuese *et al.*, 1969).

5.5 Epoxy encapsulants

The light extraction efficiency can be enhanced by using dome-shaped encapsulants with a large refractive index. As a result of the encapsulation, the angle of total internal reflection through the top surface of the semiconductor is increased. It follows from Eq. (5.22) that the ratio of extraction efficiency with and without epoxy encapsulant is given by

$$\frac{\eta_{\text{epoxy}}}{\eta_{\text{air}}} = \frac{1 - \cos \phi_{c,\text{epoxy}}}{1 - \cos \phi_{c,\text{air}}} \quad (5.36)$$

where $\phi_{c,\text{epoxy}}$ and $\phi_{c,\text{air}}$ is the critical angle for total internal reflection at the semiconductor–epoxy and semiconductor–air interface, respectively. Figure 5.6 shows the calculated ratio of the extraction efficiency with and without an epoxy dome. Inspection of the figure yields that the efficiency of a typical semiconductor LED increases by a factor of 2 – 3 upon encapsulation with an epoxy having a refractive index of 1.5.

The inset of Fig. 5.6 (b) shows that light is incident at an angle of approximately 90° at the epoxy–air interface due to the dome-shape of the epoxy. Thus, total internal reflection losses do not occur at the epoxy–air interface. Besides improving the external efficiency of an LED, the

encapsulant can also be used as a spherical lens for applications requiring a directed emission pattern. In polymer LEDs, encapsulants increase the extraction efficiency by only a small amount due to the inherently small refractive index of polymers.

5.6 Temperature dependence of emission intensity

The emission intensity of LEDs decreases with increasing temperature. This decrease of the emission intensity is due to several temperature-dependent factors including (i) non-radiative recombination via deep levels, (ii) surface recombination, and (iii) carrier loss over heterostructure barriers.

The temperature dependence of LEDs near room temperature is frequently described by the phenomenological equation

$$I = I|_{0K} \exp(-T/T_1) \quad (5.37)$$

where T_1 is the *characteristic temperature* that describes the temperature dependence of the LED. A *high* characteristic temperature is desirable, which implies a small temperature dependence.

It is interesting to note that both LEDs as well as semiconductor lasers have a distinct temperature dependence of the emission intensity. In LEDs, the decrease is expressed in terms of the “ T_1 equation”. In semiconductor lasers, the threshold current, i.e. the electrical current required for the onset of lasing, increases. In lasers the increase in threshold current is expressed in terms of the well-known T_0 equation. This equation is given by

$$I_{th} = I_{th}|_{0K} \exp(-T/T_0) . \quad (5.38)$$

Note the formal similarity of the “ T_1 equation” (Eq. 5.36) and the “ T_0 equation” (Eq. 5.37). Both equations are purely phenomenological equations intended to describe the experimental results without a strong theoretical framework allowing the derivation of the equations from basic principles.

Experimental results of the temperature dependence of the emission intensity are shown in Fig. 5.7 (adopted from Toyoda Gosei Corporation, 2000). The figure shows the temperature dependence of the emission intensity at a constant current for a blue GaInN/GaN, a green GaInN/GaN, and a red AlGaInP/GaAs LED.

Inspection of Fig. 5.7 reveals that the blue LED has the highest T_1 and the red LED has the lowest T_1 . This is not surprising since the blue LED has the deepest wells so that carrier confinement is very effective in the GaInN/GaN structures.

Exercise. Compensation of the temperature dependence of an LED with a drive circuit.

Consider an LED with a characteristic temperature $T_1 = 100$ K, a turn-on voltage of 1.4 V at 20 °C, a temperature coefficient of the turn-on voltage of 0.5 mV/K, and a differential resistance of 20 Ω for forward voltages larger than the turn-on voltage.

Design a drive circuit consisting of a constant-voltage source and a resistor, which compensates for the temperature dependence of the emission intensity of the LED so that the LED emission intensity is the same at the freezing-point temperature (0 °C) and 60 °C. The LED should draw 20 mA at the freezing-point temperature.

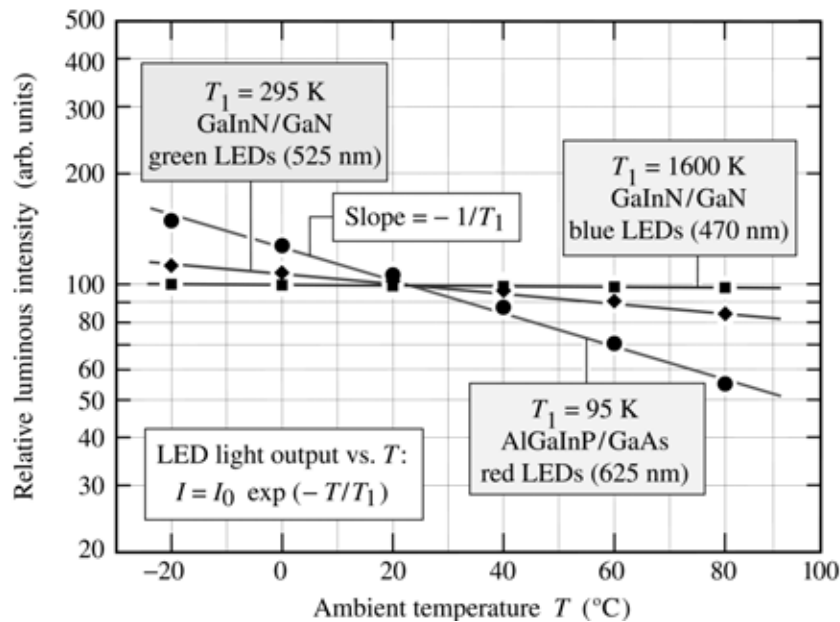


Fig. 5.7. Characteristic temperature T_1 of GaInN/GaN blue, GaInN/GaN green, and AlGaInP/GaAs red LEDs near room temperature (after data from Toyoda Gosei Corp., 2000).

References

Toyoda Gosei Corporation, Japan, General LED catalogue (2000)

6

High internal efficiency LED designs

There are two general possibilities for attaining high internal quantum efficiency. The first possibility is to enhance the radiative recombination probability, and the second possibility is to decrease the non-radiative recombination probability. This can be accomplished in different ways which will be discussed below.

6.1 Double heterostructures

The lifetimes derived from the bimolecular rate equation show that the radiative rate increases with the free carrier concentration for both the low excitation limit as well as the high excitation limit. It is therefore important that the region in which recombination occurs has a high carrier concentration. Double heterostructures are an excellent way to achieve such high carrier concentrations.

A double heterostructure (DH) consists of the active region in which recombination occurs and two confinement layers cladding the active region. A double heterostructure LED structure is shown schematically in Fig. 6.1. The two *cladding* or *confinement layers* have a larger bandgap than the active region. If the bandgap difference between the active and the confinement regions is ΔE_g , then the band discontinuities occurring in the conduction and valence band follow the relation

$$E_g|_{\text{cladding}} - E_g|_{\text{active}} = \Delta E_g = \Delta E_C + \Delta E_V . \quad (1)$$

Both band discontinuities, ΔE_C and ΔE_V , should be much larger than kT in order to avoid carrier escape from the active region into the confinement regions.

The effect of a double heterostructure on the carrier concentration is illustrated schematically in Fig. 6.2. In the case of a homojunction, carriers diffuse to the adjoining side of the junction under forward bias conditions. Minority carriers are distributed over the electron and hole

diffusion lengths as illustrated in Fig. 6.2 (a). In III–V semiconductors, diffusion lengths can be 10 μm or even longer.

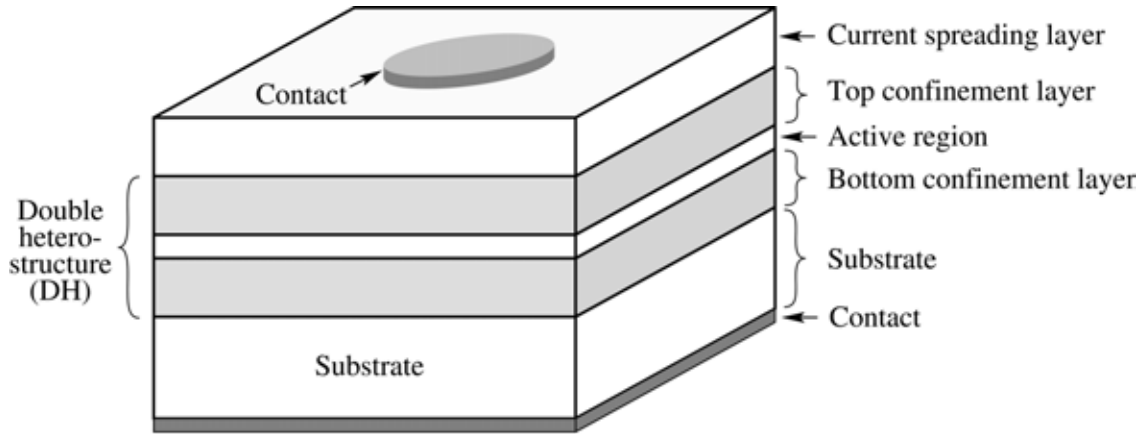


Fig. 6.1. Illustration of a double heterostructure consisting of a bulk or quantum well active region and two confinement layers. The *confinement* layers are frequently called *cladding* layers.

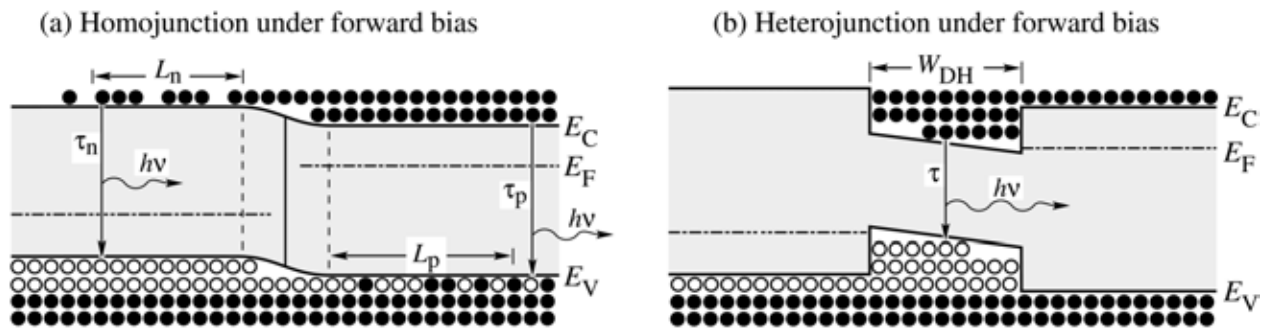


Fig. 6.2. Free carrier distribution in (a) a homojunction and (b) a heterojunction under forward bias conditions. In homojunctions, carriers are distributed over the diffusion length. In heterojunctions, carriers are confined to the well region.

The wide distribution of carriers and the correspondingly low carrier concentration (particularly towards the end of the diffusion tail) can be avoided by the employment of double heterostructures. Carriers are confined to the active region, as shown in Fig. 6.2 (b), as long as the barrier heights are much higher than the thermal energy kT . Today virtually all high-efficiency LEDs use double heterostructure designs.

Double heterostructures are used for *bulk* as well as *quantum well* active regions. Quantum well active regions provide additional carrier confinement to the narrow well regions, which can

further improve the internal quantum efficiency. On the other hand, if a quantum well active region is used, the barriers between the wells will impede the flow of carriers between adjacent wells. Thus the barriers in a multi-quantum well active region need to be sufficiently transparent (low and/or thin barriers) in order to allow for efficient transport between the wells and to avoid the inhomogeneous distribution of carriers within the active region.

The thickness of the active region in a DH LED has a strong influence on the internal quantum efficiency of an LED. Typical active region thicknesses are a few tenths of a micrometer for bulk active regions and even thinner for quantum well active regions. The dependence of device efficiency on the active region thickness is shown in Fig. 6.3 (Sugawara *et al.*, 1992). Inspection of the figure shows that the optimum thickness for an AlGaInP active region is between 0.15 and 0.75 μm .

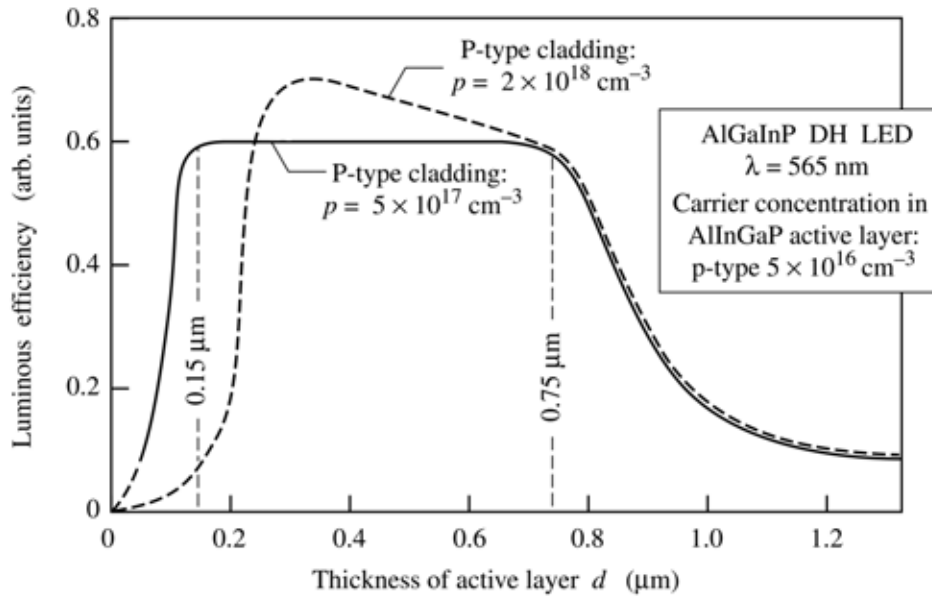


Fig. 6.3. Dependence of the luminous efficiency of an AlGaInP double heterostructure LED emitting at 565 nm on the active layer thickness. The figure reveals an optimum active region thickness of 0.15 – 0.75 μm (after Sugawara *et al.*, 1992).

If the active region in a double heterostructure becomes too thick, e.g. larger than the diffusion length of carriers, the advantage of the double heterostructure is lost and carriers are distributed as they are in homojunctions. On the other hand, if the active region of a double heterostructure becomes too thin, the active region tends to overflow at high injection current levels.

6.2 Doping of active region

Doping of the active region and confinement layers plays a crucial role in the efficiency of double heterostructure (DH) LEDs. The influence of the doping on the internal efficiency is multifaceted. First, we consider the doping of the active region.

The active region of a DH LED must not be heavily doped. Heavy doping with either p-type or n-type dopants would place the p-n junction effectively at the edge of the DH well region, i.e. at the active/confinement interface, thereby promoting carrier spill-over into one of the confinement regions. Diffusion of carriers into the cladding region decreases the radiative efficiency. Thus heavy doping of the active region is rarely done in DH LEDs.

Therefore it is required that the active region is *either* doped at a level *lower* than the doping concentration in the confinement regions *or* left undoped. Typically the doping concentration in intentionally doped active regions is either p-type or n-type in the 10^{16} to low 10^{17} cm^{-3} range. Frequently the active layer is left undoped.

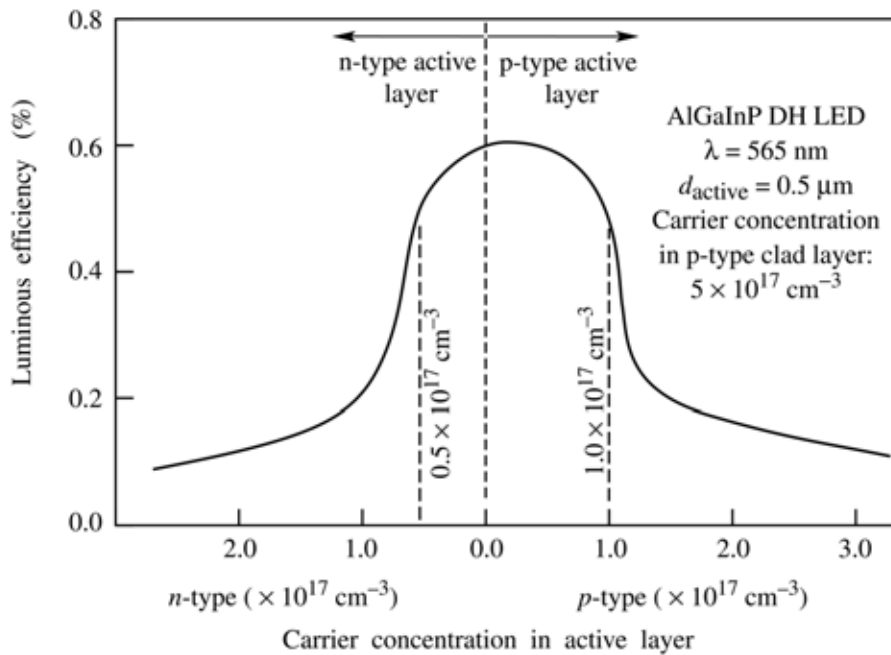


Fig. 6.4. Dependence of the luminous efficiency of an AlGaInP double heterostructure LED emitting at 565 nm on the active layer doping concentration (after Sugawara *et al.*, 1992).

P-type doping of the active region is more common than n-type doping of the active region, due to the generally longer electron-minority-carrier diffusion length compared with the hole-minority-carrier diffusion length. (Note that electrons generally have a higher mobility than holes

in III–V semiconductors.) Thus, p-type doping of the active region ensures a more uniform carrier distribution throughout the active region.

The dependence of device efficiency on the doping concentration of the active region in an AlGaInP DH LED is shown in Fig. 6.4 (Sugawara *et al.*, 1992). Inspection of the figure shows that a high quantum efficiency is obtained only for n-type doping concentrations below $5 \times 10^{16} \text{ cm}^{-3}$, and below $1 \times 10^{17} \text{ cm}^{-3}$ for p-type active layers.

Figure 6.4 also reveals that light p-type doping of the active region is preferable over light n-type doping. Therefore, most active layers of LEDs and lasers are lightly doped with acceptors. Electrons with their larger diffusion length are also more likely to diffuse into the p-type confinement layer if the active region is doped n-type.

Intentional doping of the active region can have advantages as well as disadvantages. The carrier lifetime depends on the concentration of majority carriers. In the low-excitation regime, the radiative carrier lifetime decreases with increasing free carrier concentration, i.e. doping concentration. As a result, the radiative efficiency increases. An example of a material whose luminescence efficiency increases with doping concentration is Be-doped GaAs. It is well known that the radiative efficiency of Be-doped GaAs increases with the Be doping level in the moderate doping concentration range.

On the other hand, dopants may, especially at high concentrations, introduce defects that act as recombination centers. High concentrations of intentional dopants lead to an increased concentration of native defects due to the dependence of the native and non-native defect concentrations on the Fermi level (see, for example, Longini and Greene, 1956; Baraff and Schluter, 1985; Walukiewicz, 1988, 1989, 1994; Neugebauer and Van de Walle, 1999).

The epitaxial growth process may also depend on doping. Doping atoms can act as *surfactants*, i.e. surface-active reagents. For example, a surfactant can increase the surface diffusion coefficient, thereby improving crystal quality. There are many other ways by which surfactants influence the growth process. These processes are generally not understood in great detail. However, the improvement of crystal quality has been found in a number of cases. During the growth of InGaN, for example, a marked improvement of crystal quality has been found upon doping with silicon (Nakamura *et al.*, 1996, 1998).

6.3 P-N junction displacement

The displacement of the p-n junction from its intended location into the cladding layer can be a significant problem in DH LED structures. Usually, the lower confinement layer is n-type, the

upper confinement layer is p-type, and the active region is undoped or lightly doped with n- or p-type dopants. However, if dopant redistribution occurs, the p-n junction can be displaced into one of the confinement layers. The diffusion of dopants can occur during growth and be caused by high growth temperature, a long growth time, or a strongly diffusing dopant. Dopants can redistribute due to diffusion, segregation, and drift.

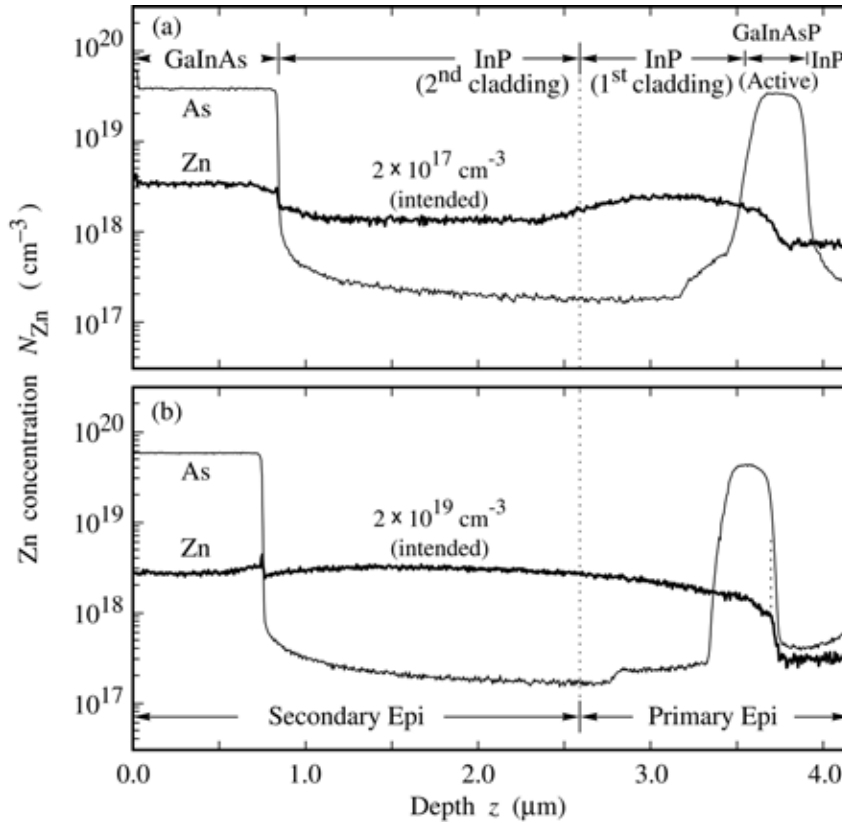


Fig. 6.5. Secondary ion mass spectrometry (SIMS) profile of Zn in a GaInAsP/InP double heterostructure. The structure uses Zn as a p-type dopant. Part (a) shows no p-n junction displacement. Part (b) shows p-n junction displacement caused by high Zn doping of the upper cladding region (after Schubert *et al.*, 1995).

Frequently, acceptors from the top confinement layer diffuse into the active region and also into the lower confinement layer. Impurities such as Zn and Be are small atoms that can easily diffuse through the crystal lattice. In addition, Zn and Be are known to have a strongly concentration-dependent diffusion coefficient. If a certain critical concentration is exceeded, Zn and Be acceptors diffuse very rapidly. As a result, the device will not work well and will not emit light at the intended emission wavelength.

An example of a Zn acceptor profile in a GaInAsP/InP double heterostructure measured by secondary ion mass spectrometry (SIMS) is shown in Fig. 6.5. In Fig. 6.5 (a), the top confinement layer has a moderate intended doping concentration of $2 \times 10^{17} \text{ cm}^{-3}$. The Zn profile indicates that the Zn is mostly confined to the upper confinement layer, even though some Zn has evidently diffused into the active region. In Fig. 6.5 (b), however, the top confinement layer

has a high intended doping concentration of $2 \times 10^{19} \text{ cm}^{-3}$. The profile indicates that Zn has strongly diffused into the active region. The p-n junction is displaced to the edge of the active region. As a result, the device shown in Fig. 6.5 (b) has a much lower quantum efficiency than the device shown in Fig. 6.5 (a).

A model explaining the p-n junction displacement in the GaInAsP/InP DH structure is illustrated in Fig. 6.6 (Schubert *et al.* 1995). It is assumed in this model that the Zn diffusion coefficient increases rapidly above a *critical* concentration N_{critical} . If this concentration is exceeded during growth, Zn will redistribute until the concentration falls below the critical concentration. As a result, Zn can diffuse into and through the active region of the double heterostructure. It is remarkable that the p-n junction displacement can occur, even if the intended Zn concentration in the *confinement region in the vicinity* of the active region is quite low.

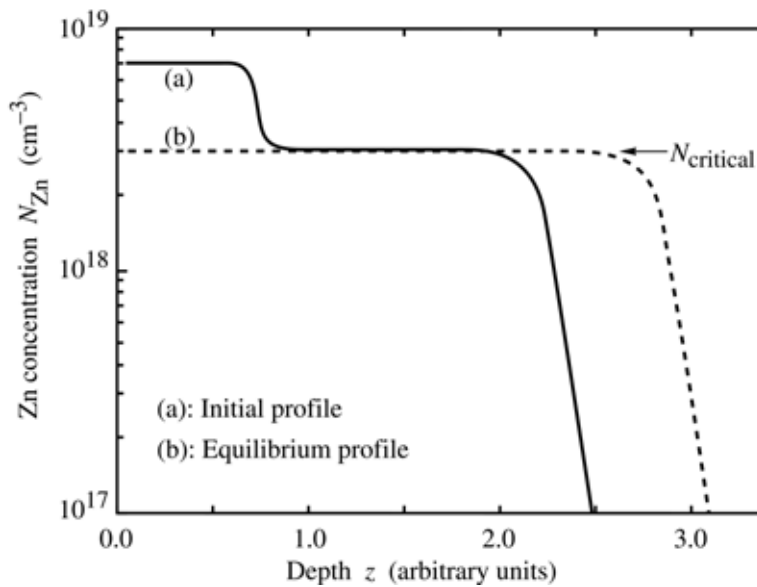


Fig. 6.6. P-N junction displacement process caused by excessive doping of the cladding region. If the acceptor dopant has a highly concentration-dependent diffusion constant and the diffusion constant increases strongly above a critical concentration, N_{critical} , p-n junction displacement occurs in the active region (after Schubert *et al.*, 1995).

6.4 Doping of the confinement regions

Doping of the confinement regions has a strong influence on the efficiency of double heterostructure LEDs. The resistivity of the confinement regions is one factor determining the doping concentration in the confinement layers. The resistivity should be low to avoid resistive heating of the confinement regions.

Another factor is the residual doping concentration in the active region. The active region has a residual doping concentration, even if not intentionally doped. Typical doping concentrations in the active region are in the $10^{15} - 10^{16} \text{ cm}^{-3}$ range. The doping concentration in the

confinement layers must be higher than the doping concentration of the active region to define the location of the p-n junction.

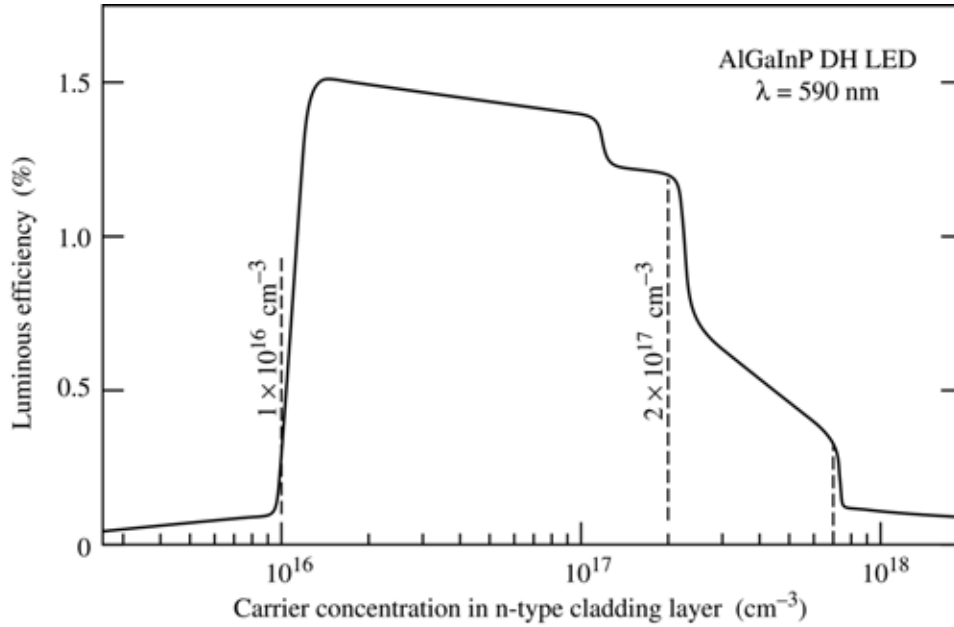


Fig. 6.7. Dependence of the luminous efficiency of an AlGaInP double heterostructure LED emitting at 565 nm on the n-type confinement layer doping concentration (after Sugawara *et al.*, 1992).

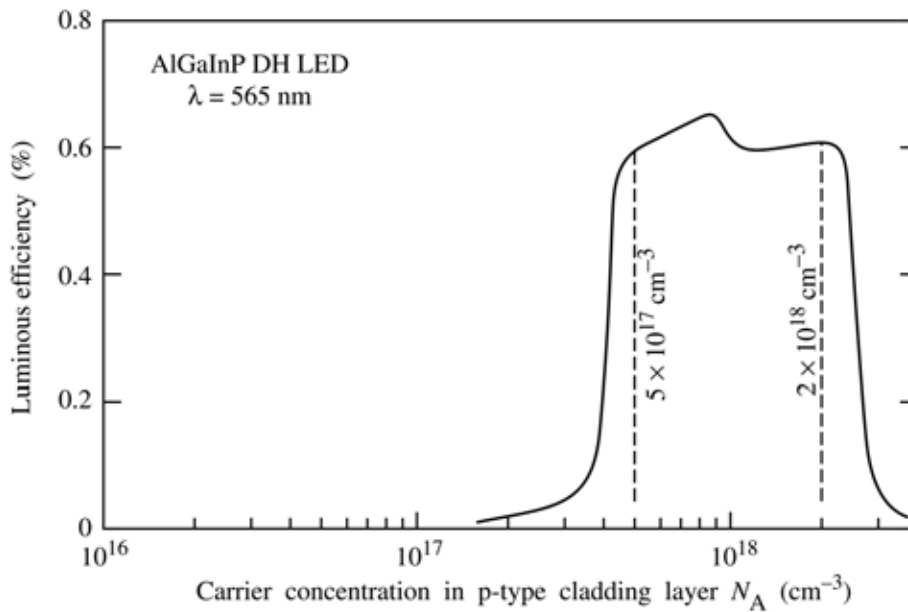


Fig. 6.8. Dependence of the luminous efficiency of an AlGaInP double heterostructure LED emitting at 565 nm on the p-type confinement layer doping concentration (after Sugawara *et al.*, 1992).

The influence of the confinement-layer doping concentration on the internal quantum efficiency was analyzed by Sugawara *et al.* (1992). The results are shown in Figs. 6.7 and 6.8. The figures reveal that there is an optimum doping range for the confinement regions. For n-type confinement regions, the optimum doping concentration ranges from 10^{16} to $2 \times 10^{17} \text{ cm}^{-3}$. For p-type confinement regions, the optimum doping concentration ranges from 5×10^{17} to $2 \times 10^{18} \text{ cm}^{-3}$, clearly higher than in the n-type cladding region. The reason for this marked difference could again lie in the larger diffusion length of electrons than that of holes. A high p-type concentration in the cladding region keeps electrons in the active region and prevents them from diffusing deep into the confinement region.

The carrier leakage out of the active region into the p-type cladding layer in double heterostructure lasers was investigated by Kazarinov and Pinto (1994). It was shown that the electron leakage out of the active region is more severe than the hole leakage. The difference is due to the generally higher diffusion constant of electrons compared with holes.

The band diagram of a double heterostructure under forward bias conditions is shown in Fig. 6.9. The figure illustrates that barriers are formed by the depletion layers at the two confinement–active-region interfaces. Compositional grading of these interfaces can be used to reduce these barriers.

The influence of the cladding layer doping concentration on the radiative efficiency of a double heterostructure laser at threshold is shown in Fig. 6.10 (Kazarinov and Pinto, 1994). Inspection of the figure reveals that the confinement layer doping has a severe influence on the luminous efficiency. Low doping concentrations in the p-type confinement layers facilitates electron escape from the active region, thereby lowering the internal quantum efficiency.

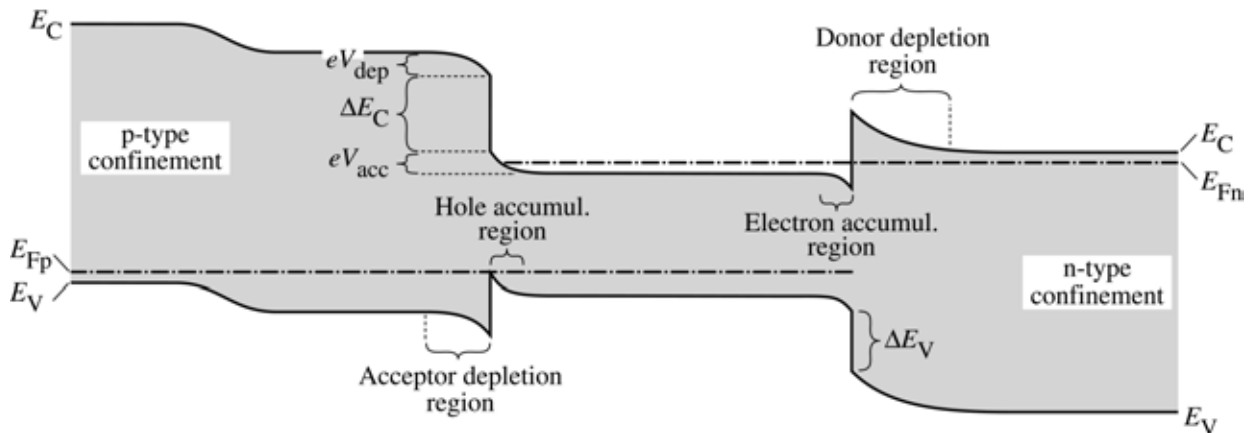


Fig. 6.9. Band diagram of a forward-biased double heterostructure. The p-type confinement layer consists of a lightly doped layer close to the active region and a higher doped layer further away from the active layer (adapted from Kazarinov and Pinto, 1994).

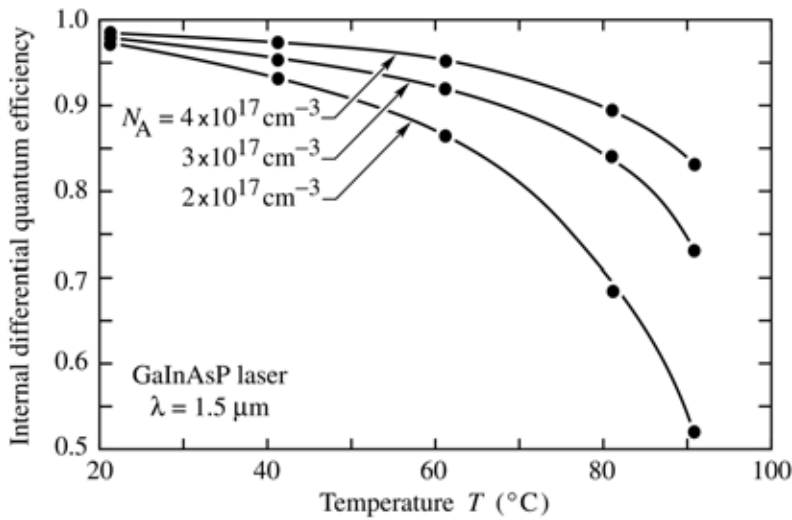


Fig. 6.10. Dependence of the internal differential quantum efficiency (emitted photons per injected electron) on temperature for different p-type doping levels in the cladding layer (after Kazarinov and Pinto, 1994).

6.5 Non-radiative recombination

It is imperative that the material comprising the active region of light-emitting devices is of very high crystal quality. **Deep levels** caused by point defects, unwanted impurities, dislocations, and other defects must have a very low concentration.

Similarly, **surface recombination** must be kept at the lowest possible levels. This can be accomplished by keeping free surfaces several diffusion lengths away from regions in which both electrons and holes reside, i.e. any surfaces must be “out of reach” of the active region.

Mesa-etched LEDs and lasers, in which the mesa etch exposes the active region to air, generally have low internal efficiencies due to surface recombination. Surface recombination also leads to a reduction of the lifetime of LEDs. Surface recombination generates heat at the semiconductor surface which can lead to structural defects such as dark-line defects that further reduce the efficiency of LEDs.

Figure 6.11 shows the light intensity of two mesa-etched and two planar LEDs versus time. Inspection of Fig. 6.11 reveals that (i) the light intensity at $t = 0$ h of the mesa-etched LEDs is slightly lower than that of the planar structure and that (ii) the lifetime of the mesa-etched device is much shorter. In the planar device, electron-hole recombination occurs below the top metal contact far away from the sidewall surfaces of the device. Thus, no intensity reduction or degradation mediated by surface recombination is expected for the planar device.

Note that the presence of surfaces does *not* reduce the radiative efficiency if only *one type of carrier* is present, e.g. near the top contact of the device. Surfaces in such unipolar regions do not have any deleterious effects.

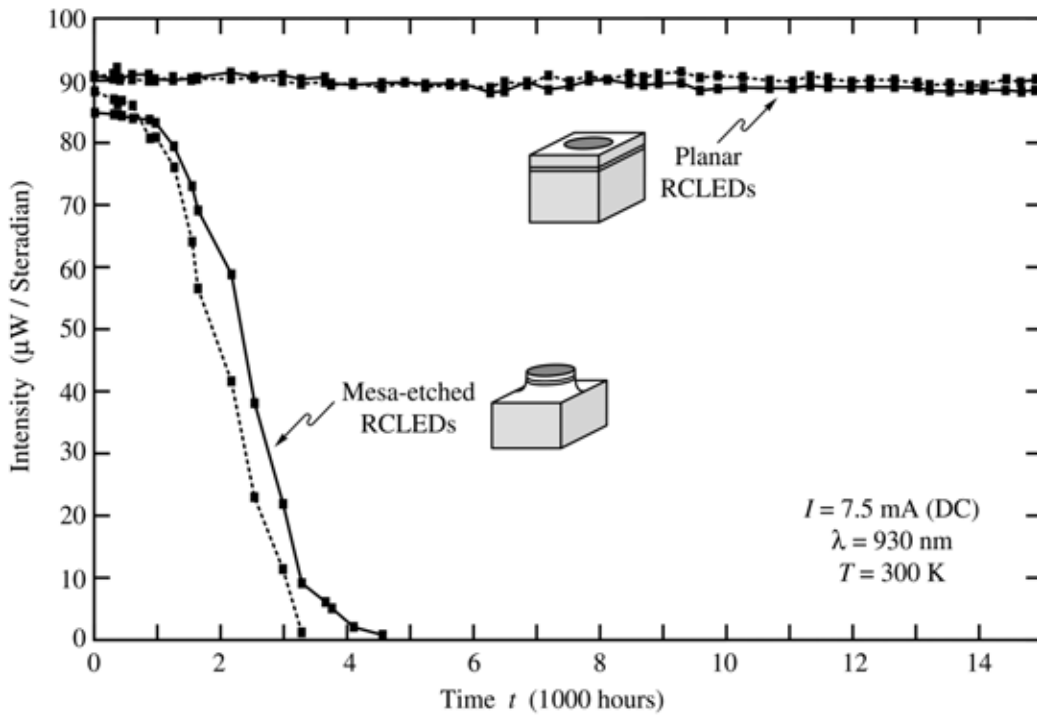


Fig. 6.11. Emission intensity of two mesa-etched LEDs and two planar LEDs versus time (after Schubert and Hunt, 1998)

6.6 Lattice matching

In double heterostructures, the active region material is different from the confinement layer material. However, both materials should have the same crystal structure and lattice constant. If the semiconductors do not have the same lattice constant, defects will occur at or near the interface between the two semiconductors. Figure 6.12 illustrates that “*dangling bonds*” can result as a consequence of mismatched semiconductors.

Inspection of Fig. 6.12 reveals that a *row* of dangling bonds can occur at the interface of two mismatched materials. Such *misfit dislocation lines* are straight lines of extended defects that can be made visible by cathodo-luminescence. Usually a *cross-hatched* pattern can be observed in the micrograph of mismatched structures. The effect of misfit dislocation lines on radiative recombination is shown in Fig. 6.13. The cathodo-luminescence micrograph shown in Fig. 6.13 exhibits a cross-hatch pattern of dark lines. The lines appear darker than the surrounding areas since carriers recombine non-radiatively at these dislocation lines.

Misfit dislocations may not occur directly *at* the interface between the mismatched materials but may start to form *near* the interface. This is because the mismatched crystal grown on top of the semiconductor will initially be strained elastically and assume the same in-plane lattice constant as the underlying substrate. This situation is shown in Fig. 6.14, where a thin layer is

strained so that it has the same in-plane lattice constant as the underlying material. Once the energy needed to strain the lattice exceeds the energy required to form misfit dislocations, the thin film relaxes to its equilibrium lattice constant by forming misfit dislocations. The layer thickness at which misfit dislocations are formed is called the *critical thickness*. It has been calculated by Matthews and Blakeslee (1976). If the layer is thinner than the critical thickness given by the Matthews–Blakeslee law, a thin dislocation-free layer can be grown even if the layers have different lattice constants.

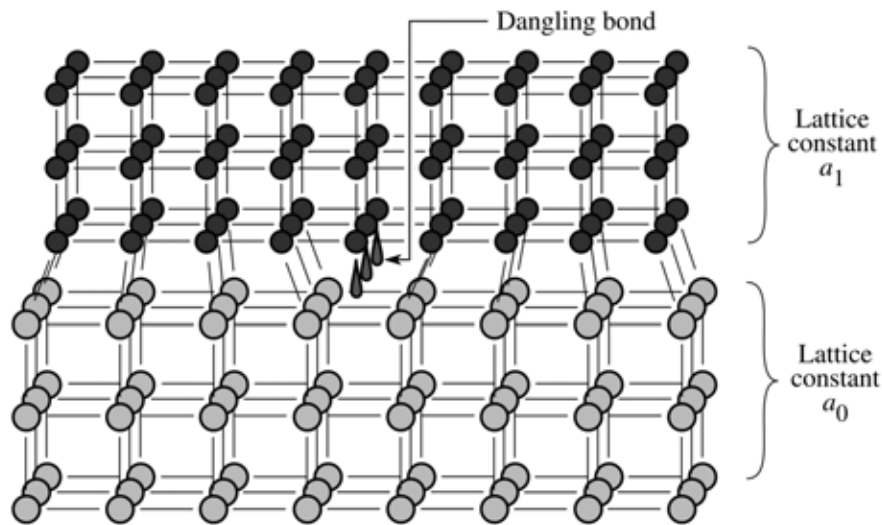


Fig. 6.12. Illustration of two crystals with mismatched lattice constant resulting in dislocations at or near the interface between the two semiconductors.

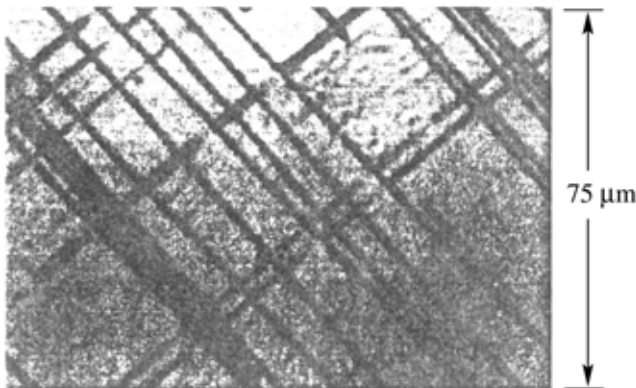


Fig. 6.13. Cathodo-luminescence image of a $0.35\ \mu\text{m}$ thick $\text{Ga}_{0.95}\text{In}_{0.05}\text{As}$ layer grown on a GaAs substrate. The dark lines forming a cross-hatch pattern are due to misfit dislocations (after Fitzgerald *et al.*, 1989).

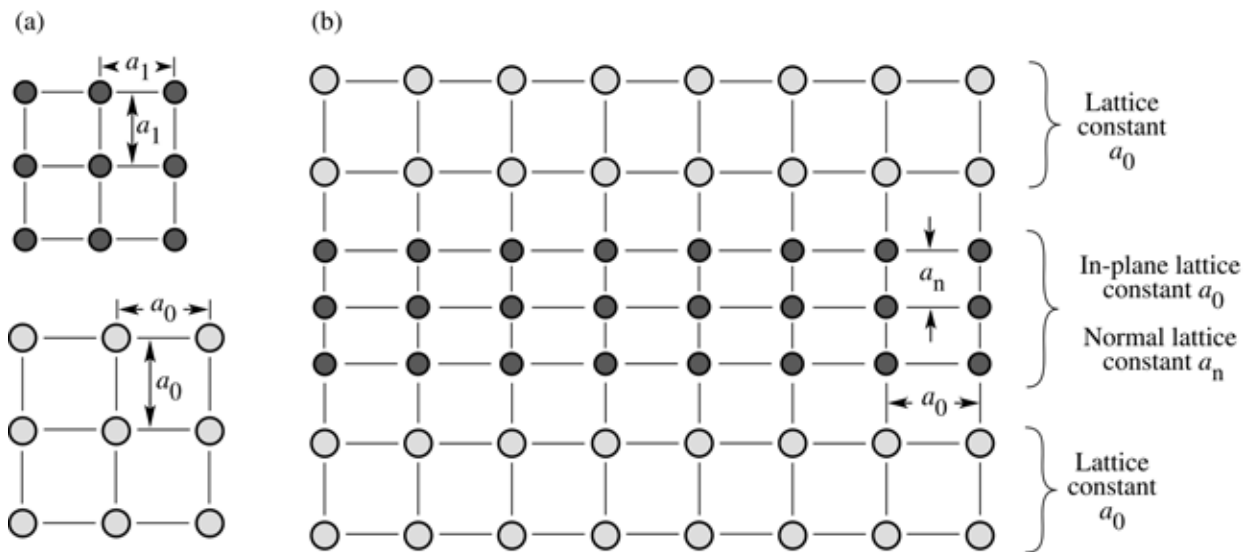


Fig. 6.14. (a) Cubic-symmetry crystals with equilibrium lattice constant a_1 and a_0 . (b) Thin, coherently strained crystal layer with equilibrium lattice constant a_1 sandwiched between two semiconductors with equilibrium lattice constant a_0 . The coherently strained layer assumes an in-plane lattice constant a_0 and a normal lattice constant a_n .

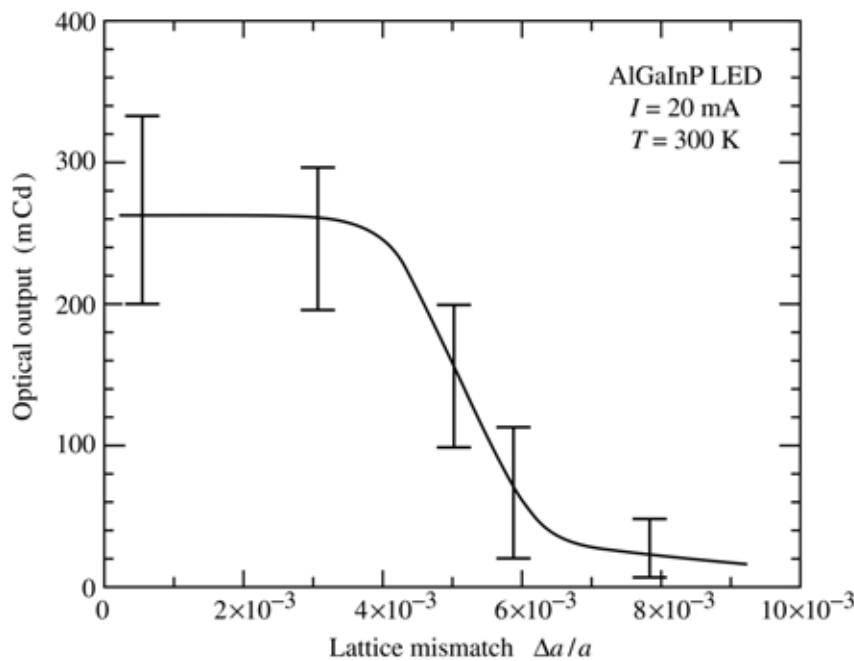


Fig. 6.15. Optical output intensity of an AlGaInP LED driven with an injection current of 20 mA versus the lattice mismatch between the AlInGaP active region and the GaAs substrate (after Watanabe and Usui, 1987).

The density of misfit dislocation lines per unit length is proportional to the lattice mismatch. Consequently, the efficiency of LEDs is expected to drop as the lattice mismatched is increased. Figure 6.15 shows the decrease of the optical intensity of an AlGaInP LED grown on a GaAs

substrate. The AlGaInP material used for high-brightness red LEDs, is matched to the GaAs substrate. Figure 6.15 reveals that the light output drops strongly as the mismatch, defined as $\Delta a / a$, exceeds 3×10^{-3} .

Red GaAsP LEDs are grown on GaAs substrates and the active layers are mismatched to the substrate. As a result, red GaAsP LEDs are low cost but also low efficiency. The homojunction GaAsP LEDs grown on GaAs substrates are the lowest-cost red devices available.

Whereas the GaAs and InP material family is strongly affected by surface recombination and lattice mismatch, the GaN material family is less so. One of the reasons for the insensitivity of the GaN material family to dislocation defects is the lower electrical activity that dislocations in these materials appear to exhibit. Another reason is the smaller diffusion length of carriers in the GaN material family. If the mean distance between dislocations is larger than the diffusion length, in particular the hole diffusion length, non-radiative recombination at these dislocations will not be severe. Another model explaining the high efficiency of GaInN is based on compositional fluctuations of the ternary alloy, which localize carriers preventing them from diffusing towards dislocation lines.

References

- Baraff G. A. and Schluter M. "Electronic structure, total energies, and abundances of the elementary point defects in GaAs" *Phys. Rev. Lett.* **55**, 1327 (1985)
- Fitzgerald E. A., Watson G. P., Proano R. E., Ast D. G., Kirchner P. D., Pettit G. D., and Woodall J. M. "Nucleation mechanism and the elimination of misfit dislocations at mismatched interfaces by reduction of growth area" *J. Appl. Phys.* **65**, 2220 (1989)
- Kazarinov R. F. and Pinto M. R. "Carrier transport in laser heterostructures" *IEEE J. Quantum Electronics* **30**, 49 (1994)
- Longini R. L. and Greene R. F. "Ionization interaction between impurities in semiconductors and insulators" *Phys. Rev.* **102**, 992 (1956)
- Matthews J. W. and Blakeslee A. E. "Defects in epitaxial multilayers. III. Preparation of almost perfect multilayers" *J. Crystal Growth* **32**, 265 (1976)
- Nakamura S., Mukai T., and Iwasa N, "Light-emitting GaN-based compound semiconductor device" US Patent 5,578,839 (1996)
- Nakamura S., Mukai T., and Iwasa N, "Light-emitting GaN-based compound semiconductor device" US Patent 5,747,832 (1998)
- Neugebauer J. and Van de Walle C. G. "Chemical trends for acceptor impurities in GaN"

- J. Appl. Phys.* **85**, 3003 (1999)
- Sugawara H., Ishikawa M., Kokubun Y., Nishikawa Y., Naritsuka S., Itaya K., Hatakoshi G., Suzuki M., “Semiconductor light emitting device” US Patent 5,153,889, issued Oct. 6 (1992)
- Schubert E. F., Downey S. W., Pinzone C. and Emerson A. B. “Evidence of very strong inter-epitaxial-layer diffusion in Zn doped GaInPAs/InP structures” *Applied Physics A* **60**, 525 (1995)
- Schubert E. F. and Hunt N. E. J. “15,000 hours stable operation of resonant-cavity light-emitting diodes” *Applied Physics A* **66**, 319 (1998)
- Walukiewicz, W. “Fermi level dependent native defect formation: consequences for metal–semiconductor and semiconductor–semiconductor interfaces” *J. Vac. Sci. Technol. B*, **6**, 1257 (1988)
- Walukiewicz, W. “Amphoteric native defects in semiconductors” *Appl. Phys. Lett.* **54**, 2094 (1989)
- Walukiewicz, W. “Defect formation and diffusion in heavily doped semiconductors” *Phys. Rev. B* **50**, 5221 (1994)
- Watanabe H., Usui A. “Light emitting diode” US Patent 4,680,602, issued July 14 (1987)

High extraction efficiency structures

Owing to the high refractive index of semiconductors, light incident on a planar semiconductor–air interface is totally internally reflected, if the angle of incidence is sufficiently large. Snell’s law gives the critical angle of total internal reflection. As a result of total internal reflection, light can be “trapped” inside the semiconductor. Light trapped in the semiconductor will eventually be reabsorbed by a defect, the substrate, active region, or another absorbing layer.

If the light is absorbed by the substrate, the electron–hole pair will most likely recombine non-radiatively due to the inherently low efficiency of substrates. If the light is absorbed by the active region, the electron–hole pair may re-emit a photon or recombine non-radiatively. For active regions with internal quantum efficiencies of less than 100 %, a reabsorption event by the active region reduces the efficiency of the LED. The absorption of light by the metallic contacts is another mechanism that lowers the extraction efficiency.

The *external quantum efficiency* of an LED is the product of the *internal quantum efficiency*, η_{int} , and the *extraction efficiency*, $\eta_{\text{extraction}}$, i.e.

$$\eta_{\text{ext}} = \eta_{\text{int}} \eta_{\text{extraction}} \cdot \quad (7.1)$$

The extraction efficiency thus plays an important role in increasing the power efficiency of LEDs.

7.1 Absorption of below-bandgap light in semiconductors

To obtain a high light extraction efficiency and avoid absorption of light, all semiconductor layers other than the active region should have a larger bandgap energy than the photon energy. This can be done in different ways, for example by using double heterostructures, window layers, and other structures that will be discussed below. In this section, we discuss the absorption of light if the energy of the light is *below* the energy gap of the semiconductor.

Naively, one would assume that a semiconductor can absorb light only if the photon energy is higher than the bandgap energy and that the semiconductor is transparent for photon energies below the bandgap. However, semiconductors do absorb below-bandgap light, although with a much lower absorption coefficient.

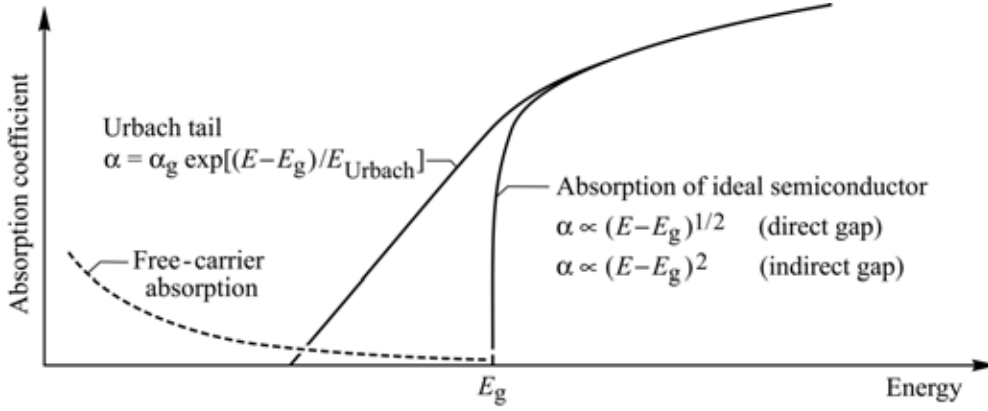


Fig. 7.1. Absorption coefficient of a semiconductor with bandgap E_g versus energy. The “Urbach tail” dominates absorption near but below the bandgap. Absorption further below the bandgap is dominated by free-carrier absorption.

The absorption coefficient versus energy of an idealized direct-gap semiconductor and a real semiconductor is shown schematically in Fig. 7.1. In an idealized semiconductor at low temperatures, the absorption coefficient versus energy is given by (Pankove, 1971)

$$\alpha \propto (E - E_g)^{1/2} \quad (\text{direct gap}) \quad (7.2a)$$

$$\alpha \propto (E - E_g)^2 \quad (\text{indirect gap}) . \quad (7.2b)$$

An idealized semiconductor has a *zero* band-to-band absorption coefficient at the bandgap energy ($E = E_g$). The absorption strength in a *real* semiconductor, for below-bandgap light, can be expressed in terms of an exponentially decaying absorption strength. In this absorption tail, called the **Urbach tail**, the absorption coefficient versus energy is given by

$$\alpha = \alpha_g \exp \left[(E - E_g) / E_{\text{Urbach}} \right] \quad (7.2c)$$

where α_g is the experimentally determined absorption coefficient at the bandgap energy and E_{Urbach} is the characteristic energy (here called the **Urbach energy**), which determines how rapidly the absorption coefficient decreases for below-bandgap energies.

Urbach (1953) measured the absorption tail for different temperatures and showed that the *Urbach energy* is approximately kT , the thermal energy. The temperature dependence of the Urbach tail led Knox (1963) to the conclusion that the below-bandgap transitions are phonon-assisted transitions. Thus, the Urbach energy is given by

$$E_{\text{Urbach}} = kT . \quad (7.3)$$

The Urbach tail can be caused by mechanisms other than phonon-assisted absorption. Any mechanism introducing a potential fluctuation will lead to local variations of the semiconductor band edges. As a result of these fluctuations, the bandgap energy fluctuates as well, and below-bandgap transitions can occur.

The most common potential fluctuations are fluctuations caused by random dopant distribution and local variations of the chemical composition of a ternary or quaternary alloy semiconductor.

Potential fluctuations caused by random dopant distribution can be calculated using Poisson statistics. The magnitude of these fluctuations is given by (see, for example, Schubert *et al.*, 1997)

$$\Delta E_{\text{Urbach}} = \frac{2e^2}{3\epsilon} \sqrt{\left(N_D^+ + N_A^-\right) \frac{r_s}{3\pi}} e^{-3/4} \quad (7.4)$$

where r_s is the screening radius.

Potential fluctuations caused by random compositional fluctuations can be calculated using binomial statistics. The magnitude of these fluctuations is given by (see, for example, Schubert *et al.*, 1984)

$$\Delta E_{\text{alloy}} = \frac{dE_g}{dx} \left[\frac{x(1-x)}{4a_0^{-3}V_{\text{exc}}} \right]^{1/2} \quad (7.5)$$

where x is the alloy composition of the ternary semiconductor alloy, a_0 is the semiconductor lattice constant, and V_{exc} is the excitonic volume of the electron-hole pair.

It depends on the specific case which of the different physical effects dominates the Urbach tail. Generally, binary semiconductors such as GaP or GaAs have a smaller tail than alloys such as AlGaAs or GaAsP. Furthermore, low-doped semiconductors have a smaller Urbach tail than heavily doped semiconductors.

For energies sufficiently lower than the bandgap energy, absorption due to the Urbach tail is negligibly small and **free-carrier absorption** becomes the dominant absorption mechanism. As the name suggests, a free carrier is excited to a higher energy by absorption of a photon. The absorption transition must conserve momentum. This momentum change is provided by acoustic phonons, optical phonons, or by impurity scattering.

Free-carrier absorption is proportional to the free-carrier concentration, since a free carrier is required for an absorption event. Theoretical considerations in terms of the classical Drude free-electron model further show that free-carrier absorption increases as the square of the incident wavelength (Pankove, 1971). Thus the proportionalities

$$\alpha_{fc} \propto n \lambda^2 \quad \text{and} \quad \alpha_{fc} \propto p \lambda^2 \quad (7.6)$$

are valid for n-type and p-type semiconductors, respectively. Theoretical considerations in terms of a quantum mechanical treatment show that the absorption coefficient is proportional to $\lambda^{3/2}$, $\lambda^{5/2}$, and $\lambda^{7/2}$ depending upon whether acoustic phonon scattering, optical phonon scattering, or ionized impurity scattering is involved in the momentum conservation process, respectively (Swaminathan and Macrander, 1991).

In n-type and p-type GaAs, the room-temperature free-carrier absorption coefficient near the bandgap energy ($\lambda \approx 950$ nm) can be expressed as (Casey and Panish, 1978)

$$\alpha_{fc} = 3 \text{ cm}^{-1} \frac{n}{10^{18} \text{ cm}^{-3}} + 7 \text{ cm}^{-1} \frac{p}{10^{18} \text{ cm}^{-3}} . \quad (7.7)$$

Inspection of the equation indicates that the free-carrier absorption coefficient can be of the order of 10 cm^{-1} at high carrier concentrations. Approximate values for the free-carrier absorption coefficient in several compound semiconductors are given in Table 7.1.

In LEDs, free-carrier absorption can affect the intensity of waveguided modes radiating out the side of the chip. Free-carrier absorption also plays a role in LEDs with transparent semiconductor substrates. Such transparent substrates have a typical thickness greater than $100 \mu\text{m}$. If the doping concentration of a transparent substrate is high, free-carrier absorption will reduce the light-output power. If it is low, the substrate becomes resistive. Thus a compromise needs to be made between the different doping requirements of a transparent substrate. For thin layers, such as confinement layers, free-carrier absorption effects are negligibly small if the optical path length within the layer is short.

Material	Wavelength	Electron concentration	α_{fc}
GaN	1.0 μm	$1 \times 10^{18} \text{ cm}^{-3}$	40 cm^{-1} (a, e)
GaP	1.0 μm	$1 \times 10^{18} \text{ cm}^{-3}$	22 cm^{-1} (b, e)
GaAs	1.0 μm	$1 \times 10^{18} \text{ cm}^{-3}$	3.0 cm^{-1} (c)
InP	1.0 μm	$1.1 \times 10^{18} \text{ cm}^{-3}$	2.5 cm^{-1} (d)

Table 7.1. Free-carrier absorption coefficient (α_{fc}) of n-type semiconductors. (a) After Ioffe (2002). (b) After Wiley and DiDomenico (1970). (c) After Casey and Panish (1978). (d) After Kim and Bonner (1983) and Walukiewicz *et al.* (1980). (e) Data are extrapolated using the proportionality $\alpha_{fc} \propto n \lambda^2$.

7.2 Double heterostructures

Virtually all LED structures employ double heterostructures. They consist of two *confinement* layers and an *active* region. The band diagram of a double heterostructure is shown in Fig. 7.2. The active region has a smaller bandgap energy than the two confinement regions. As a result, the confinement regions are *transparent* to the light emitted by the active region. Since the confinement regions are relatively thin, they can be considered, for all practical purposes, to be totally transparent.

Reabsorption of light by the active region in the current-injected area below the top contact can also be neglected. The active region is, under normal injection conditions, injected with high current densities so that the electron and hole quasi-Fermi levels rise into the bands, as illustrated in Fig. 7.2. As a result, the active region is practically transparent for near-bandgap emission under high injection conditions.

Note, however, that the active region is in equilibrium sufficiently far away from the current-injected region. These regions are not current injected and thus will absorb near-bandgap light emitted by the active region. To reduce optical losses by absorption, the active region should have a high internal quantum efficiency to make re-emission of absorbed photons likely.

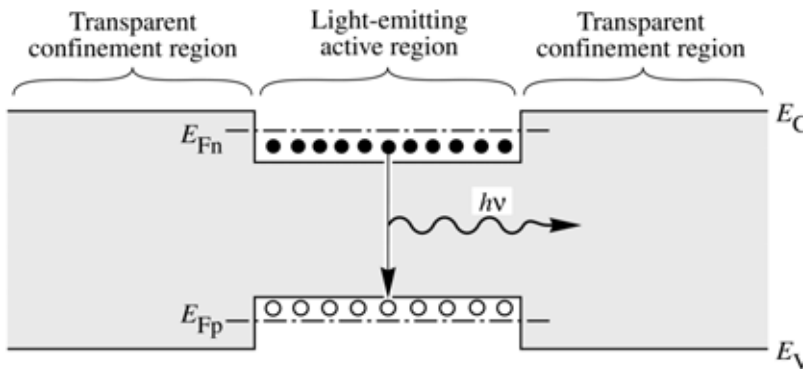


Fig. 7.2. Double heterostructure with optically transparent confinement regions. Reabsorption in the active region is unlikely due to the high carrier concentration in the active region and the resulting Burstein–Moss shift of the absorption edge.

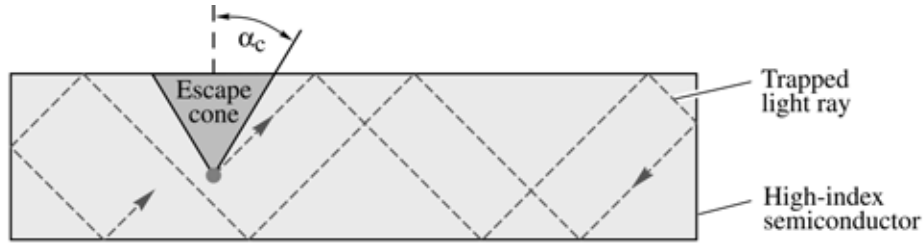


Fig. 7.3. “Trapped light” in a rectangular-parallelpiped-shaped semiconductor unable to escape for emission angles greater than α_c due to total internal reflection.

7.3 Shaping of LED dies

One of the most important problems facing high-efficiency LEDs is the occurrence of **trapped light** within a high-index semiconductor. The occurrence of trapped light is illustrated in Fig. 7.3. A light ray emitted by the active region will be subject to total internal reflection, as predicted by Snell’s law. In the high-index approximation, the angle of total internal reflection is given by

$$\alpha_c = \bar{n}_s^{-1} \quad (7.8)$$

where \bar{n}_s is the semiconductor refractive index and the critical angle α_c is given in radians. For high-index semiconductors, the critical angle is quite small. For example, for a refractive index of 3.3, the critical angle for total internal reflection is only 17° . Thus most of the light emitted by the active region is trapped inside the semiconductor. The trapped light is most likely to be absorbed by the thick substrate. Once absorbed, the electron–hole pair is likely to recombine non-radiatively due to the comparatively low quality and efficiency of the substrate.

The light-escape problem has been known since the infancy of LED technology in the 1960s. It has also been known that the geometrical shape of the LED die plays a critical role. The optimum LED would be spherical in shape with a point-like light-emitting region in the center of the LED. Such a spherical LED is shown in Fig. 7.4 (a). Light emanating from the point-like active region is incident at a normal angle at the semiconductor–air interface. As a result, total internal reflection does not occur in such LEDs. Note, however, that the light is still subject to Fresnel reflection at the interface unless the sphere is coated with an anti-reflection coating.

LEDs with a hemispherical dome-like structure (Carr and Pittman, 1963) as well as other shapes, e.g. inverted, truncated cones (Franklin and Newman, 1964; Loebner, 1973) have been demonstrated to improve extraction efficiency over conventional designs, i.e. rectangular

parallelepiped chips. However, the practical utility of such devices has not been realized primarily due to the high cost associated with shaping of individual LED dies.

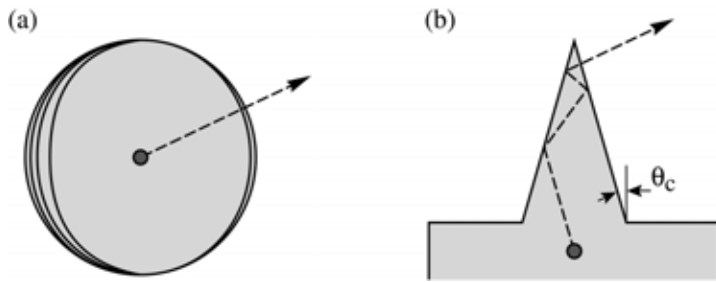


Fig. 7.4. Schematic illustration of different geometric shapes for LEDs with perfect extraction efficiency. (a) Spherical LED with a point-like light-emitting region at the center of the sphere. (b) A cone-shaped LED.

Unfortunately, spherical LEDs with a point-like light source in the center of the LED are somewhat impractical devices. Semiconductor fabrication technology is, in view of the flat substrates used in epitaxial growth, a *planar* technology. Thus spherical LEDs are difficult to fabricate using conventional planar technology.

Another interesting LED structure is a cone-shaped structure, as shown schematically in Fig. 7.4 (b). Light rays are emanating from the active region at or below the base of the cone. The light rays incident at the cone–air boundary are either transmitted through the semiconductor–air interface or guided by the cone. The guided rays undergo multiple reflections. As light rays undergo multiple reflections, they will form a progressively increasing angle of incidence at the semiconductor–air interface. As a result, light guided by the cone will eventually have near-normal incidence and escape from the cone. Although an interesting concept, cone-shaped LEDs are difficult to fabricate and manufacture.

The most common LED structure has the shape of a rectangular parallelepiped as shown in Fig. 7.5 (a). Such LED dies are fabricated by cleaving the wafer along its natural cleaving planes. The LEDs have a total of *six* escape cones, two of them perpendicular to the wafer surface, and four of them parallel to wafer surface. The bottom escape cone will be absorbed by the substrate if the substrate has a lower bandgap than the active region. The four in-plane escape cones will be at least partially absorbed by the substrate. Light in the top escape cone will be obstructed by the top contact, unless a thick current-spreading layer is employed. Thus the simple rectangular parallelepiped LED is clearly a structure with low extraction efficiency. However, a substantial advantage of such LEDs is the low manufacturing cost.

An LED with a cylindrical shape is shown in Fig. 7.5 (b). The cylindrical LED has the advantage of higher extraction efficiency compared with a cube-shaped LED. An escape ring, as

shown in Fig. 7.5 (b), replaces the four in-plane escape cones of the rectangular LED, which results in a substantial improvement of the extraction efficiency. The cylindrical-shaped LED requires one more processing step (etching step) compared with rectangularly shaped LEDs.

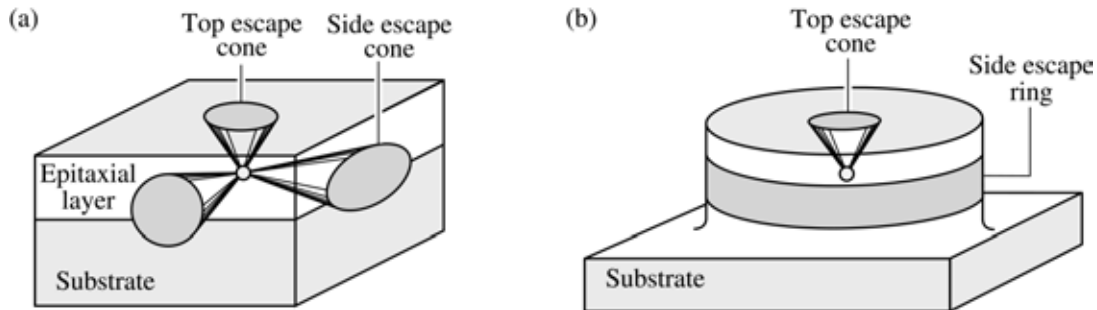


Fig. 7.5. Illustration of different geometric shapes of LEDs. (a) Rectangular parallelepipedal LED die with a total of six escape cones. (b) Cylindrical LED die with a top escape cone and a side escape ring.

The *truncated inverted pyramid* (TIP) LED is an example of an LED that employs advanced LED die shaping (Krames *et al.*, 1999). The schematic structure and a photograph of an AlGaInP/GaP TIP LED are shown in Fig. 7.6. Ray traces indicated in the figure show that light rays entering the base of the pyramids can escape from the semiconductor after undergoing one or multiple internal reflections. The TIP geometry reduces the mean photon path-length within the crystal, and thus reduces the effect of internal loss mechanisms.

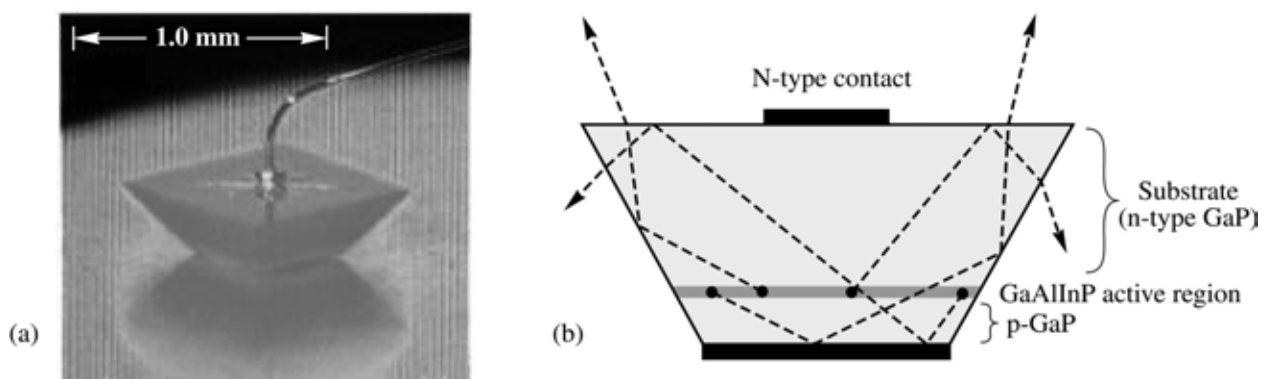


Fig. 7.6. Truncated inverted pyramid (TIP) AlGaInP/GaP LED. (a) LED driven by an electrical injection current. (b) Schematic diagram of the LED illustrating the enhanced light extraction efficiency (after Krames *et al.*, 1999).

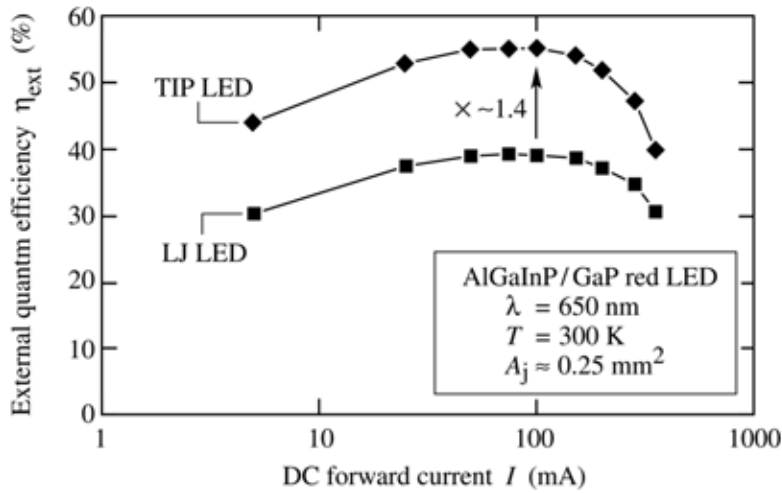


Fig. 7.7. External efficiency vs. forward current for red-emitting (650 nm) truncated inverted pyramid (TIP) LEDs and large-junction (LJ) LEDs mounted in power-lamp packages. The TIP LED exhibits a 1.4 times improvement in extraction efficiency compared with the LJ device, resulting in a peak external quantum efficiency of 55 % at 100 mA (after Krames *et al.*, 1999).

The geometric shape of the LED is chosen in such a way that trapping of light is minimized. Ray tracing computer models are employed to maximize the escape probability from the semiconductor. The TIP LED is a high-power LED with a large p-n junction area of $500 \mu\text{m} \times 500 \mu\text{m}$. The luminous efficiency of TIP LEDs exceeds 100 lm/W and is one of the highest ever achieved with LEDs.

The TIP LED performance versus injection current is shown in Fig. 7.7 (Krames *et al.*, 1999). A peak luminous efficiency of 102 lm/W was measured for orange-spectrum ($\lambda \approx 610 \text{ nm}$) devices at an injection current of 100 mA. This luminous efficiency exceeds that of most fluorescent (50 – 104 lm/W) and all metal-halide (68 – 95 lm/W) lamps. In the amber color regime, the TIP LED provides a photometric efficiency of 68 lm/W ($\lambda \approx 598 \text{ nm}$). This efficiency is comparable to the source efficiency of 50 W high-pressure sodium discharge lamps. A peak external quantum efficiency of 55 % was measured for red-emitting ($\lambda \approx 650 \text{ nm}$) TIP LEDs. Under pulsed operation (1 % duty cycle), an efficiency of 60.9 % was achieved (data not shown), which sets a lower bound on the extraction efficiency of these devices.

The rectangular parallelepiped and cylindrical LED structures can be fabricated with wafer-scale processing steps. More advanced structures, such as the TIP LED, are also fabricated by wafer-scale processes, namely a sawing process. The manufacturing cost of LEDs requiring die-level processing steps is much higher compared with LEDs manufactured with wafer-scale processes.

Other efficient ways to increase the light extraction efficiency include the use of *roughened* or *textured semiconductor surfaces* (see, for example, Schnitzer *et al.*, 1993; Windisch *et al.*, 1999, 2000, 2001, 2002) and the use of tapered output couplers (Schmid *et al.*, 2000, 2001, 2002). For infrared GaAs-based devices, external quantum efficiencies near 50 % have been

demonstrated with surface-textured LEDs and devices having tapered output couplers. A detailed discussion of properties and fabrication of microstructured surfaces was given by Sinzinger and Jahns (1999).

7.4 Current-spreading layer

In LEDs with thin top confinement layers, the current is injected into the active region mostly under the top electrode. Thus, light is generated under an opaque metal electrode. This results in a low extraction efficiency. This problem can be avoided with a **current-spreading layer** that spreads the current under the top electrode to regions not covered by the opaque top electrode.

The *current-spreading layer* is synonymous with the **window layer**. The term window layer is occasionally used to indicate the *transparent character* of this layer and its ability to enhance the extraction efficiency.

The usefulness of current-spreading layers was realized during the infancy of LEDs. Nuese *et al.* (1969) demonstrated a substantial improvement of the optical output power in GaAsP LEDs by employing a current-spreading or window layer. The window layer is the top semiconductor layer located between the upper cladding layer and the top ohmic contact. The effect of a current-spreading layer is illustrated in Fig. 7.8. Light is emitted only around the perimeter of the top contact for LEDs without a current-spreading layer, as shown in Fig. 7.8 (a). The addition of a current-spreading layer results in more uniform and brighter surface emission as shown in Fig. 7.8 (b).

Nuese *et al.* (1969) demonstrated current-spreading layers composed of the ternary GaAsP and the binary GaP and discussed the requirements of the current-spreading layer. These requirements include low resistivity and large thickness for current-spreading, and transparency to minimize absorption losses. To reduce absorption losses, Nuese *et al.* (1969) employed a high P mole fraction in the $\text{GaAs}_{1-x}\text{P}_x$ current-spreading layer, namely $0.45 < x \leq 1.0$, higher than the P mole fraction in the $\text{GaAs}_{1-x}\text{P}_x$ active region, where $x = 0.45$. Thus the bandgap energy of the current-spreading layer is higher than the bandgap of the active region. Although Nuese *et al.* discussed the properties of the current-spreading layer qualitatively, they did not provide a quantitative theoretical framework of current spreading. The theoretical foundation of current-spreading layers in devices with linear contact geometry was given by Thompson, as discussed below. The use of a current-spreading layer was adopted in most top-emitting LED designs, including AlGaAs LEDs (Nishizawa *et al.*, 1983; Moyer 1988), GaP LEDs (Groves *et al.*, 1977, 1978a, 1978b), and AlGaInP LEDs (Kuo *et al.*, 1990; Sugawara *et al.*, 1991, 1992a, 1992b).

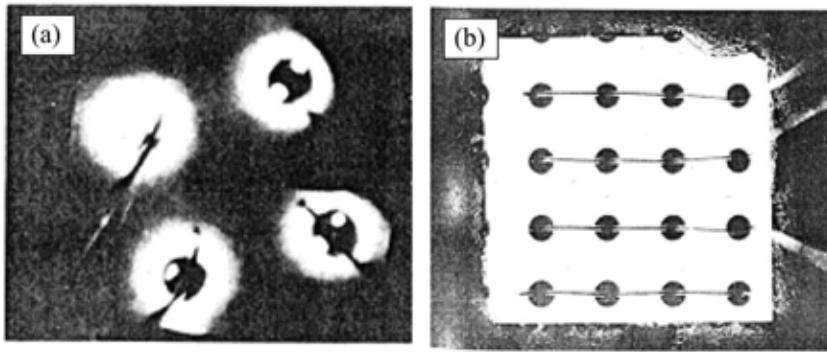


Fig. 7.8. Effect of the current-spreading layer on LED output. (a) Top view of an LED without a current-spreading layer. Emission occurs only near the perimeter of the contact. (b) Top view of an LED with a current-spreading layer (after Nuese *et al.*, 1969).

The effect of the current-spreading layer is illustrated schematically in Fig. 7.9. Without a current-spreading layer, the current-injected area of the active region is limited to approximately the contact size, as indicated in Fig. 7.9 (a). The addition of the current-spreading layer results in a larger current-injected area, as shown in Fig. 7.9 (b).

Current-spreading layers are predominantly employed in top-emitting LEDs. Two different approaches for AlGaInP visible LEDs, grown on GaAs substrates, are shown in Figs. 7.9 (c) and (d). A GaP current-spreading layer was reported by Kuo *et al.* (1990) and Fletcher *et al.*, (1991a, 1991b). GaP has a bandgap of $E_{g,\text{GaP}} = 2.26$ eV and is thus transparent for red, orange, yellow, and part of the green spectrum. LEDs with emission wavelengths as short as about 550 nm have been fabricated. GaP, as a binary compound semiconductor, is very transparent for energies below the bandgap, i.e. the Urbach tail energy of GaP is small. Furthermore GaP is an indirect-gap semiconductor, which is inherently less absorbing compared with direct-gap semiconductors. Thus very little light is absorbed even in thick GaP current-spreading layers.

However, GaP is lattice mismatched to the underlying epitaxial layers. The lower confinement, active, and upper confinement layers are lattice-matched to the GaAs substrate. Since GaP has a lattice constant that is about 3.6 % smaller than that of GaAs, a high density of threading dislocations and stacking faults are expected at the upper-cladding-layer-to-GaP interface. One could assume that these dislocations, which will act as non-radiative recombination centers, will not degrade the internal quantum efficiency of the LED because they are located at the confinement–window layer interface and in the window layer, far away from the active region. However, if the dislocations propagated downward towards the active region during device operation, the efficiency and the reliability of the LEDs would be affected. The issues associated with the confinement–window interface have apparently been resolved since AlGaInP/GaAs LEDs with GaP window layers have excellent reliability and efficiency.

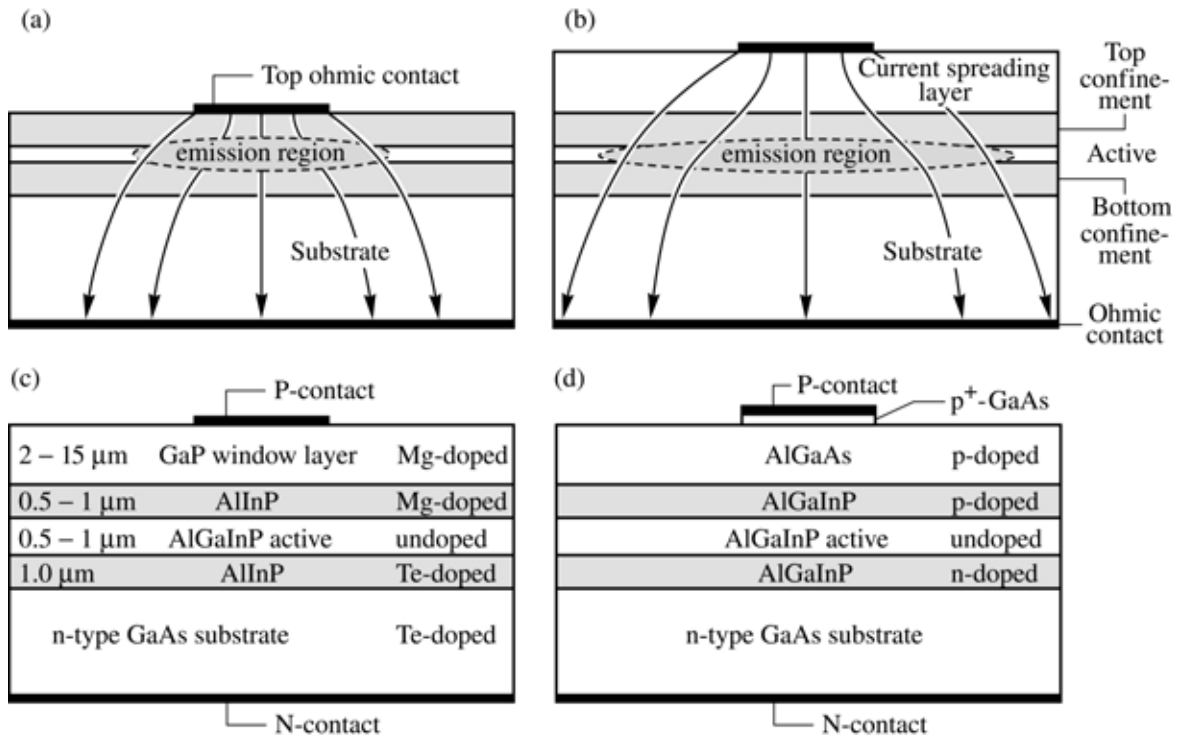


Fig. 7.9. Current-spreading structures in high-brightness AlGaInP LEDs. Illustration of the effect of a current-spreading layer for LEDs (a) without and (b) with a spreading layer on the light extraction efficiency. (c) GaP current-spreading structure (Fletcher *et al.*, 1991a, 1991b). (d) AlGaAs current-spreading structure (Sugawara *et al.*, 1992a, 1992b).

An alternative approach for increasing the extraction efficiency in AlGaInP/GaAs LEDs uses AlGaAs current-spreading layers (Sugawara *et al.*, 1991, 1992a, 1992b). $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is lattice-matched to GaAs for all chemical compositions $0 \leq x \leq 1$. AlAs has a bandgap energy of $E_{g,\text{AlAs}} = 2.9$ eV. For $x > 0.45$, $\text{Al}_x\text{Ga}_{1-x}\text{As}$ becomes an indirect semiconductor. The absorption coefficient of indirect semiconductors is much smaller than that of direct-gap semiconductors. The AlGaAs current-spreading layer is lattice matched to the underlying confinement layer and thus misfit dislocations, as in the case of GaP spreading layers, do not arise for AlGaAs current-spreading layers. However, the absorption of light in the AlGaAs layers is higher than the absorption in the GaP spreading layers. The AlGaAs is a ternary alloy and fluctuations of the cation concentration (Al and Ga) leads to a local variation of the bandgap energy. The compositional fluctuations lead to an absorption tail for energies below the AlGaAs bandgap so that AlGaAs has a larger Urbach energy than GaP.

It is well known that Al-containing compounds are difficult to grow by organo-metallic vapor-phase epitaxy (OMVPE), the common epitaxial crystal growth technique for LEDs. Aluminum is a very reactive element and OMVPE cleanliness is essential. Even minor leaks in

the growth system will result in the degradation of Al-containing films. This applies, in particular, to compounds containing a high percentage of Al such as AlAs. The optical properties of AlGaAs current-spreading layers are therefore usually inferior to the properties of GaP spreading layers. The electrical properties of AlAs or AlGaAs with high Al content are also inferior to GaP. Despite these difficulties, viable AlGaInP LEDs with AlGaAs current-spreading layers have been developed and are commercially available.

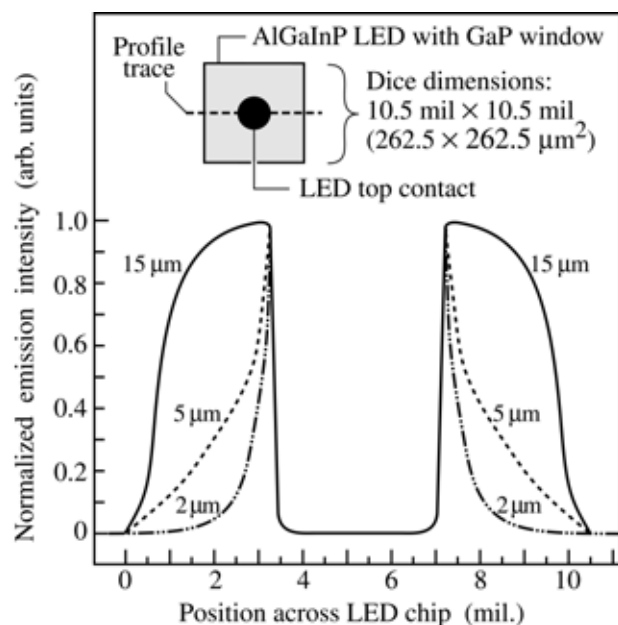


Fig. 7.10. The effect of GaP window thickness on current spreading is illustrated by surface light emission intensity profiles for three different AlGaInP LED chips with window layer thicknesses of 2, 5, and 15 μm . The profile is indicated by the dashed line in the inset. The dip in the middle of the profiles is due to the opaque ohmic contact pad. A microscope fitted with a video camera was used in the measurements (after Fletcher *et al.*, 1991a).

The increase in light extraction efficiency is illustrated in Fig. 7.10 for p-type GaP current-spreading layer with a resistivity of $0.05 \Omega \text{ cm}$ and thicknesses ranging from 2 to 15 μm (Fletcher *et al.*, 1991a). The data shown in Fig. 7.10 were obtained using a near-field microscope and a video analyzer. A single-line scan was taken across the chip, including the center p-type contact pad to obtain the light intensity profile, as shown in the inset. Since the intensity is directly proportional to the current density at any given point, the current-spreading characteristics are obtained. For a window thickness of 2 μm , current spreading is limited. As the window layer thickness is increased to 15 μm , the current spreads well beyond the contact, reaching almost the edge of the chip. An even larger thickness of the window layer would spread the current to the edges of the chip. Such strong current spreading is not desirable due to surface recombination.

The effect of current spreading on the efficiency of an AlGaInP/GaAs LED with a GaP current-spreading layer is shown in Fig. 7.11. For as sufficiently thick window layer the

extraction efficiency is increased by a factor of approximately 8. The comparison of pulsed with direct current (DC) measurements shown in Fig. 7.11 shows that the efficiency drop occurring at high currents is caused by heating of the device.

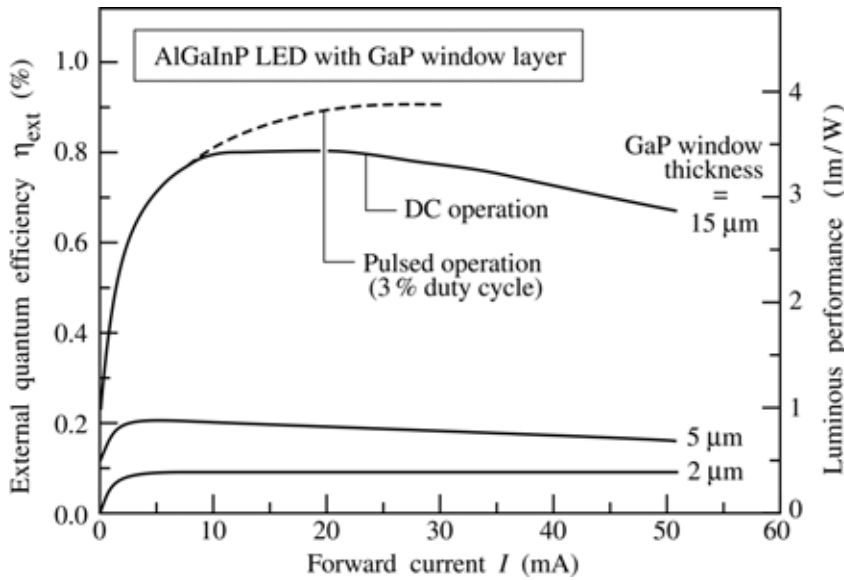


Fig. 7.11. Bare chip external quantum efficiency (photons/electron) and luminous performance (lm/W) versus forward current for AlGaInP LEDs with GaP window layer thicknesses of 2, 5, and 15 μm . Solid curves are under DC conditions. The dashed curve is under pulsed condition using 400 ns pulses and a 3 % duty cycle. Heating is essentially eliminated in this case (after Fletcher *et al.*, 1991a).

The optimum thickness range of current-spreading layers in AlGaInP/GaAs LEDs with $\text{Al}_{0.70}\text{Ga}_{0.30}\text{As}$ current-spreading layers was investigated by Sugawara *et al.* (1991, 1992a, 1992b). The p-type doping concentration of the $\text{Al}_{0.70}\text{Ga}_{0.30}\text{As}$ spreading layer was $3 \times 10^{18} \text{ cm}^{-3}$. The luminous efficiency of the LED versus current-spreading layer thickness is shown in Fig. 7.12. Inspection of the figure reveals that the optimum thickness of the current-spreading layer is between 5 and 30 μm . For a current-spreading layer thickness of 15 μm , the efficiency of the device increases by a factor of 30 compared with a device with no current-spreading layer at all. The optimum doping concentration in the p-type current-spreading layer was found in the low 10^{18} cm^{-3} range.

The disadvantage of no or a very thin current-spreading layer is that most of the light is generated under the opaque metal contact pad, thereby hindering the escape of light from the LED die. A very thick window layer is equally disadvantageous. *First*, a thick window layer spreads the current all the way to the edge of the LED dies, which leads to increased surface recombination and thus lowers the efficiency of the LED. *Secondly*, light absorption increases with the thickness of the window layer due to the absorption of below-bandgap light in the window layer. *Thirdly*, a thick window layer will increase the ohmic resistance of the device thereby lowering the overall efficiency. *Fourthly*, the long growth times required for thick

current-spreading layers may result in the diffusion of dopants from the confinement layers into the active region, thereby lowering the internal quantum efficiency.

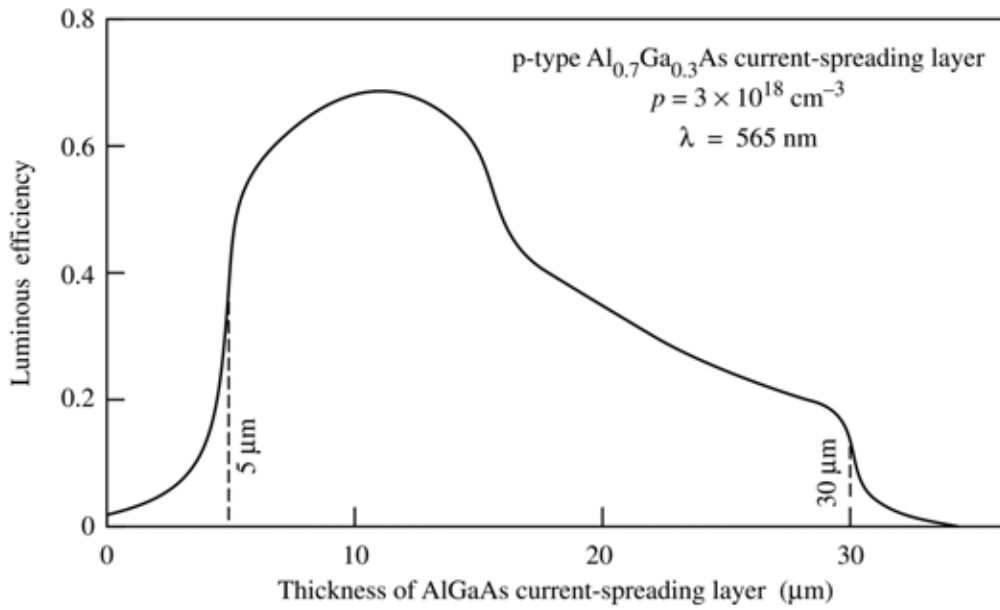


Fig. 7.12. Efficiency of an AlInGaP LED emitting in the red (565 nm) versus the thickness of an Al_{0.70}Ga_{0.30}As current spreading layer (after Sugawara *et al.*, 1992a).

Current spreading is an important issue in many LED materials, in particular in those materials that possess low conductivity. Current spreading in the top p-type layer of GaN/GaInN LEDs is very weak due to the high resistivity of the p-type top cladding layer. The hole mobility in III–V nitrides is typically $1 - 20 \text{ cm}^2/\text{V s}$ and the hole concentrations are in the 10^{17} cm^{-3} range, resulting in resistivities $> 1 \text{ } \Omega \text{ cm}$. Jeon *et al.* (2001) demonstrated an LED with a **tunnel junction** adjoining the p-type confinement region above the active region. An n-type layer on top of the tunnel junction allows for lateral current spreading under the top contact. Owing to the employment of the tunnel junction, the LED has *two* n-type but *no* p-type ohmic contact.

7.5 Theory of current spreading

The theory of current spreading under a **linear stripe top contact** geometry has been reported by Thompson (1980). Such a stripe-like geometry is typical for semiconductor lasers. Figure 7.13 (a) shows the schematic cross section of a stripe-geometry semiconductor laser. The laser has a current-spreading layer located above the p-n junction. Because of the symmetry of the laser, only the right half of the laser is shown, so that the left edge of the contact shown in the

diagram is actually the center of the laser stripe. The model assumes a constant potential and current density (J_0) under the metal ($x < r_c$). The potential throughout the substrate is assumed to be constant. The current density $J(x)$ extending away from the contact is given by

$$J(x) = \frac{2 J_0}{\left[(x - r_c) / L_s + \sqrt{2} \right]^2} \quad (x \geq r_c) \quad (7.9)$$

where L_s is the current spreading length given by

$$L_s = \sqrt{\frac{t n_{\text{ideal}} k T}{\rho J_0 e}} \quad (7.10)$$

where ρ is the resistivity of the current spreading-layer, t is the thickness of the current-spreading layer, and n_{ideal} is the diode ideality factor. The diode ideality factor has typical values of $1.05 < n_{\text{ideal}} < 1.35$.

We next develop a theoretical model that can be applied to *linear stripe* (see Fig. 7.13 (a)) as well as *circular contact* (see Fig. 7.13 (b)) shapes. We first consider the ***linear stripe contact geometry***. We assume that the current at the edge of the spreading region ($x = r_c + L_s$) is a factor of e^{-1} lower than under the metal contact. Then the voltage drop across the junction at the edge of the current-spreading layer is $n_{\text{ideal}} kT / e$ lower than under the metal contact. This voltage drops within the current spreading region. The resistance of the current spreading region along the lateral direction per unit stripe length dy is given by

$$R = \rho \frac{L_s}{t dy} . \quad (7.11)$$

The current flowing vertically through the junction in the current spreading region is given by

$$I = J_0 L_s dy . \quad (7.12)$$

Using Ohm's law, one obtains

$$\rho \frac{L_s}{t dy} J_0 L_s dy = \frac{n_{\text{ideal}} kT}{e} . \quad (7.13)$$

Solving this equation for t yields

$$t = \rho L_s^2 J_0 \frac{e}{n_{\text{ideal}} kT} \quad (7.14)$$

Comparison of this equation with Eq. (7.10) yield that the two equations are identical. Equation (7.14) allows one to calculate the required current-spreading layer thickness t for a given resistivity of this layer and the desired current spreading length L_s .

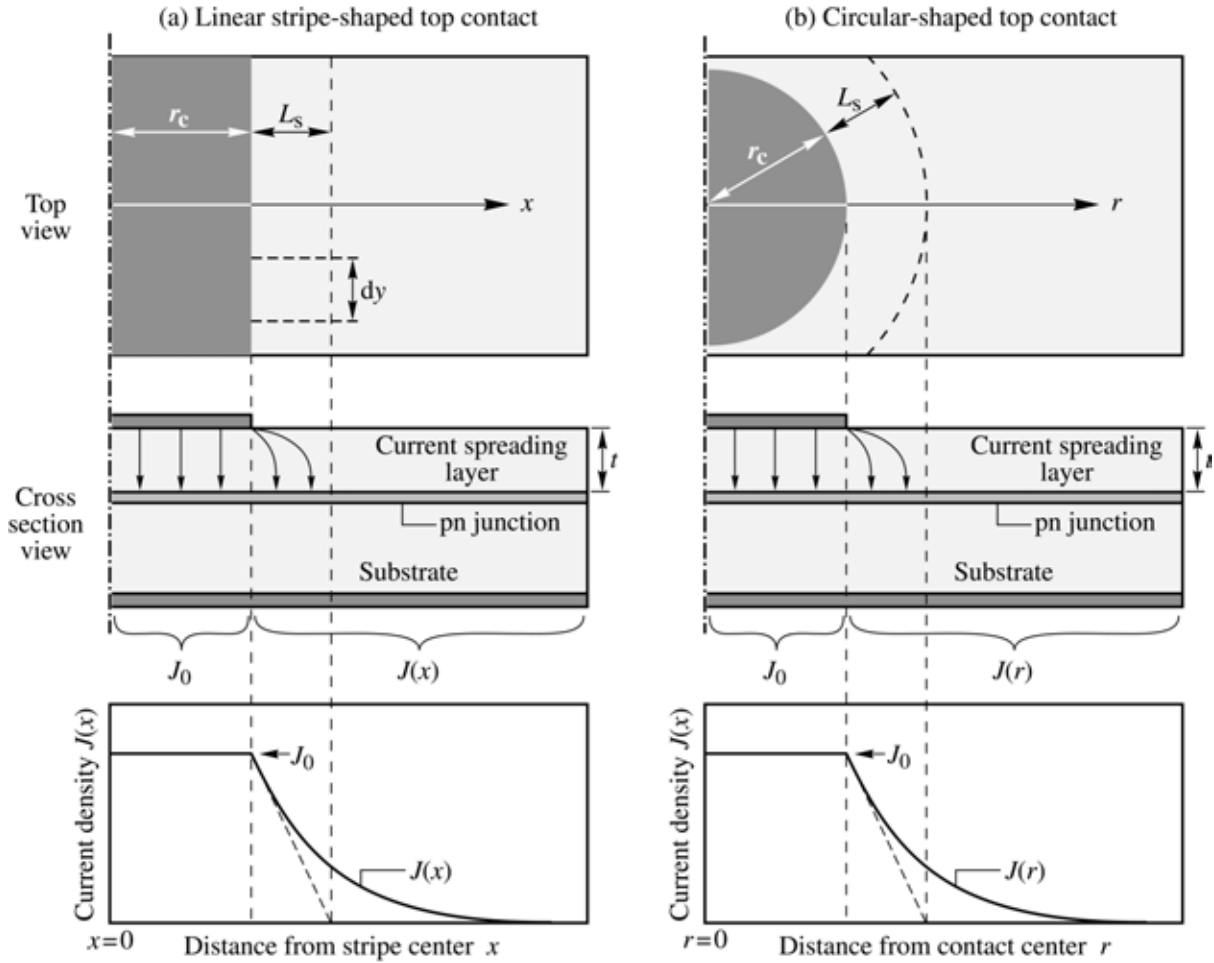


Fig. 7.13. Schematic illustration of current spreading in structures with different top contact geometries. (a) Linear stripe contact geometry. (b) Circular contact geometry.

We next consider the **circular contact geometry** shown in Fig. 7.13 (b). The circular geometry is relevant to LEDs with a circular top contact. Proceeding in a similar way, we write the lateral resistance from the edge of the contact to the edge of the current spreading region. This resistance is given by

$$R = \int_{r_c}^{r_c+L_s} \rho \frac{1}{A} dr = \int_{r_c}^{r_c+L_s} \rho \frac{1}{t 2\pi r} dr = \frac{\rho}{2\pi t} \ln\left(1 + \frac{L_s}{r_c}\right). \quad (7.15)$$

The current flowing vertically through the junction in the current spreading region is given by

$$I = J_0 \left[\pi (L_s + r_c)^2 - \pi r_c^2 \right] = J_0 \pi L_s (L_s + 2r_c). \quad (7.16)$$

Using Ohm's law, one obtains

$$\frac{\rho}{2\pi t} \ln\left(1 + \frac{L_s}{r_c}\right) J_0 \pi L_s (L_s + 2r_c) = \frac{n_{\text{ideal}} kT}{e}. \quad (7.17)$$

Solving this equation for t yields

$$t = \rho L_s \left(r_c + \frac{L_s}{2} \right) \ln\left(1 + \frac{L_s}{r_c}\right) \left(J_0 \frac{e}{n_{\text{ideal}} kT} \right) \quad (7.18)$$

Equation (7.18) allows one to calculate the required current-spreading layer thickness t for a given resistivity of this layer and the desired current spreading length L_s . Note that for large values of r_c , we can simplify Eq. (7.18) using the approximation $\ln(1+x) \approx x$. Thus, in the limit of large values of r_c ($r_c \rightarrow \infty$), Eq. (7.18) and Eq. (7.14) are identical, as expected.

7.6 Current crowding in LEDs on insulating substrates

Current crowding also occurs in mesa-structure LEDs grown on insulating substrates. This type of LED includes GaInN/GaN LEDs grown on sapphire substrates. In these LEDs, the p-type contact is usually located on the top of the mesa, and the n-type contact is located on an n-type buffer layer at the bottom of the mesa. As a result, the current tends to *crowd* at the edge of the mesa contact adjoining the n-type contact.

A lateral p-side-up mesa LED grown on an insulating substrate is shown in Fig. 7.14 (a). It is intuitively clear that the p-n junction current *crowds* near the edge of the mesa as indicated in the figure. An equivalent circuit model is shown in Fig. 7.14 (b) and includes the p-type contact resistance and the resistances of the n-type and p-type cladding layers. The p-n junction is approximated by an ideal diode. The circuit model also shows several nodes separated by a distance dx .

Assuming that the p-type metal contact has the same electrostatic potential at every point, application of Kirchhoff's current law to two adjacent nodes yields

$$\frac{d^2V}{dx^2} = \frac{\rho_n}{t_n} J_0 \left[\exp\left(\frac{eV_j}{kT}\right) - 1 \right]. \quad (7.19)$$

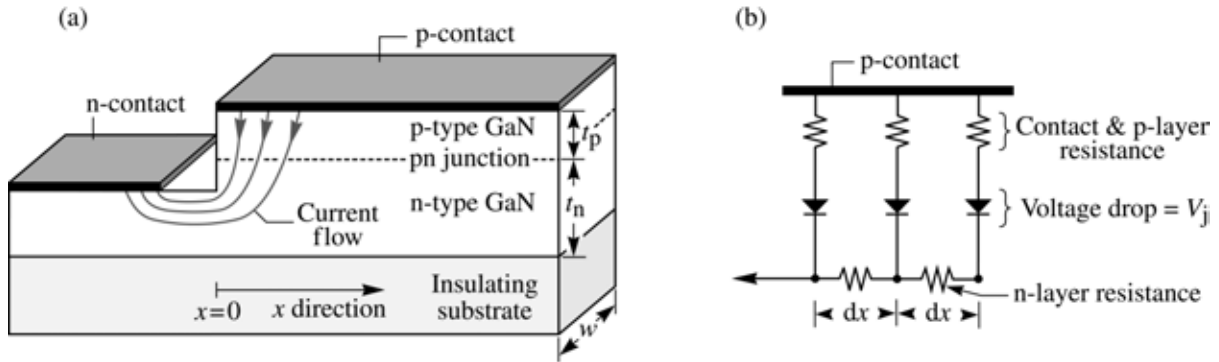


Fig. 7.14. (a) Current crowding in a mesa-structure GaN-based LED grown on an insulating substrate. (b) Equivalent circuit consisting of n-type and p-type layer resistances, p-type contact resistance, and ideal diodes representing the p-n junction.

In the case of zero or negligible resistance of the p-type layer, it is $dV = dV_j$. In this case, Eq. (7.19) can be easily solved and an analytic solution was given by Thompson (1980) who calculated the spreading length in p-n junction diodes grown on conductive substrates. In Thompson's study, the material resistivity of the top p-type cladding layer was considered, but the lower n-type cladding layer resistivity was neglected. However, in GaN/InGaN LEDs with the p-type layer on top and the n-type layer underneath, the resistive n-type layer causes current crowding and cannot be neglected. Furthermore, p-type resistances are so high that they cannot be neglected either. As will be shown in the following calculation, both types of material resistances play peculiar roles in the current crowding problem.

Next, we take into account the resistance of the n-type layer, p-type layer, and p-type contact resistance. The voltage drop across the p-n junction and the p-type resistors is given by

$$V = R_v I_0 \left[\exp\left(\frac{eV_j}{kT}\right) - 1 \right] + V_j, \quad (7.20)$$

where R_v ('vertical resistance') is the sum of the p-type layer resistance and the p-type contact resistance of the area element $w \, dx$, that is

$$R_v = \rho_p \frac{t_p}{w dx} + \rho_c \frac{1}{w dx}, \quad (7.21)$$

where ρ_p is the resistivity of the p-type layer and ρ_c is the p-type specific contact resistance. Forming the second derivative of V with respect to x in Eq. (7.20) and inserting the result into Eq. (7.19) yields the differential equation

$$\frac{e}{kT} (\rho_c + \rho_p t_p) J_0 \exp\left(\frac{eV_j}{kT}\right) \left[\frac{d^2 V_j}{dx^2} + \frac{e}{kT} \left(\frac{dV_j}{dx} \right)^2 \right] + \frac{d^2 V_j}{dx^2} = \frac{\rho_n}{t_n} J_0 \left[\exp\left(\frac{eV_j}{kT}\right) - 1 \right]. \quad (7.22)$$

In order to solve the differential equation, we restrict ourselves to the forward-bias operation of the diode. In this case, the junction voltage is much larger than kT/e , that is

$$V_j \gg kT/e \quad \text{and} \quad \exp(eV_j/kT) \gg 1. \quad (7.23)$$

Furthermore, we assume that the voltage drop across the p-type series resistance and contact resistance is much larger than kT/e

$$(\rho_c + \rho_p t_p) J_0 \exp(eV_j/kT) \gg kT/e. \quad (7.24)$$

This condition applies to typical GaN/GaInN LEDs. Using the approximations of Eqs. (7.23) and (7.24), Eq. (7.22) can be simplified to

$$\frac{d^2 V_j}{dx^2} + \frac{e}{kT} \left(\frac{dV_j}{dx} \right)^2 = \frac{\rho_n}{(\rho_c + \rho_p t_p) t_n} \frac{kT}{e}. \quad (7.25)$$

Solving Eq. (7.25) for V_j yields $V_j(x) = V_j(0) - (kT/e)(x/L_s)$. Inserting V_j into the equation $J = J_0 \exp(eV_j/kT)$ yields the solution of the differential equation as

$$\boxed{J(x) = J(0) \exp(-x/L_s)} \quad (7.26)$$

where $J(0)$ is the current density at the p-type mesa edge and L_s is denoted as the **current spreading length**, that is, as the length where the current density has dropped to the $1/e$ value of the current density at the edge, so that $J(L_s)/J(0) = 1/e$. The current spreading length is given by

$$\boxed{L_s = \sqrt{(\rho_c + \rho_p t_p) t_n / \rho_n}} \quad (7.27)$$

Equation (7.27) shows that the current distribution depends on epitaxial layer thicknesses and material properties. A thick low-resistivity n-type buffer layer is needed to ensure that current crowding is minimized. Equation (7.27) also illustrates a somewhat surprising result; namely that the *decrease* of p-type specific contact resistance or p-type layer resistivity *enhances* the current crowding effect. For *low* p-type contact and confinement resistances, strong current crowding results, unless the n-type buffer layer is very conductive so that t_n / ρ_n is very large. In GaN/GaInN devices, the sum of p-type contact and p-type layer resistances can be larger than the n-type cladding resistance, especially if t_n is small.

An experimental result on the current crowding effect in GaInN/GaN LEDs grown on a sapphire substrate is shown in Fig. 7.15 (Guo and Schubert, 2001). A micrograph of the optical emission from a GaInN LED is shown in Fig. 7.15 (a). The picture was taken from the sapphire substrate side of the LED and shows the intensity of blue light emission. The micrograph clearly reveals that the emission intensity decreases with increasing distance from the mesa edge. Figure 7.15 (b) shows the experimental intensity as a function of the distance from the mesa edge. Also shown is a theoretical fit to the experimental data using the exponential decrease in current density derived above. The experimental and the theoretical data exhibit very good agreement if a current spreading length of $550 \mu\text{m}$ is used in the calculation.

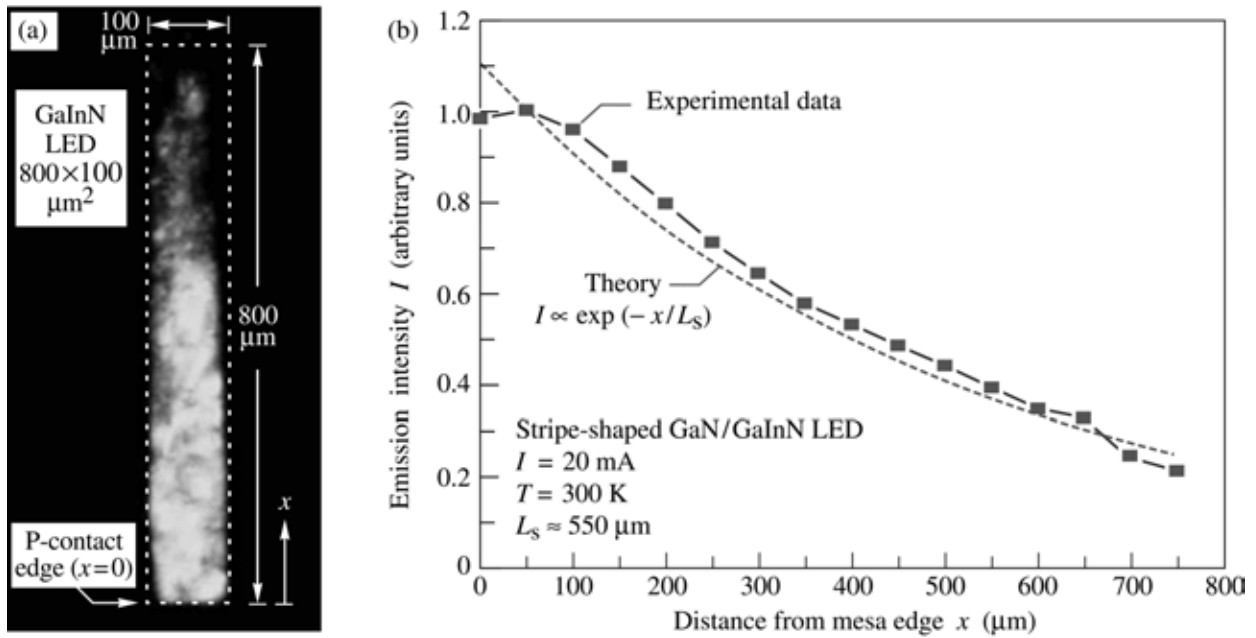


Fig. 7.15. (a) Micrograph of optical emission from mesa-structure GaInN/GaN LED grown on an insulating sapphire substrate. The LED has a stripe-shaped $800 \mu\text{m} \times 100 \mu\text{m}$ p-type contact. (b) Theoretical and experimental emission intensity versus the distance from the mesa edge (after Guo and Schubert, 2001).

High contact resistances and high p-GaN resistivity are not desirable for high-power devices since these resistances generate heat. On the other hand, these resistances alleviate the current crowding effect. Note that with the expected future improvement of the contact and p-type doping in GaN devices, and larger device and contact sizes, *current crowding becomes increasingly severe* unless novel contact geometries are introduced to alleviate the problem. Such novel contact geometries can include inter-digitated structures with p-type finger widths of less than L_s . For device dimensions much smaller than L_s , the current crowding effect can be neglected.

7.7 Cross-shaped contacts and other contact geometries

Different requirements need to be satisfied by the top contact. In regular LEDs, the top contact provides a pad for the bonding wire. The pad is usually circular with a typical diameter of $100\text{ }\mu\text{m}$. The top contact pad also provides a low-resistance ohmic contact to the current-spreading layer.

Typical top contact geometries are shown in Fig. 7.16. The simplest geometry is just a circular contact pad, as shown in Fig. 7.16 (a). A cross-shaped contact, as shown in Fig. 7.16 (b), provides a more uniform current distribution over the entire area of the active region. Note, however, that no or little current should flow to the *edges* of the LED die to avoid surface recombination.

For large-area LEDs, a simple circular pad or a cross-shaped top contact are insufficient for uniform current distribution. In such large-area devices, patterns that include a ring, as shown in Fig. 7.16 (c), are better suited to providing uniform current distribution.

The area of the top contact is kept small so that the light emanating from the active region is not hindered by the opaque contact. However, the contact resistance scales with the contact area so that the top contact area cannot be scaled down arbitrarily.

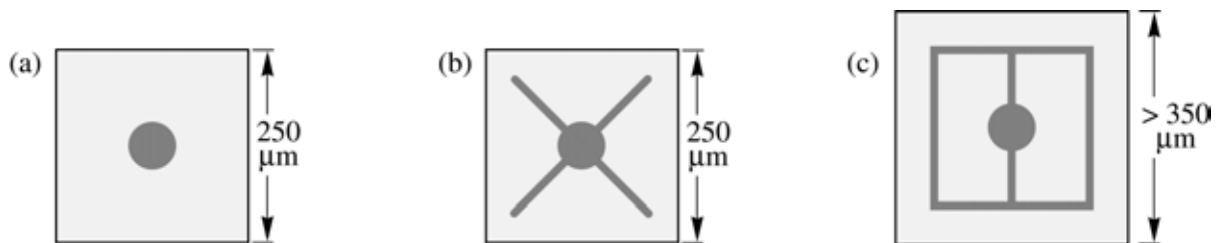


Fig. 7.16. Top view on an LED die with (a) a circular contact also serving as a bond pad and (b) a cross-shaped contact with a circular bond pad. (c) Typical contact geometry used for larger LED dies.

7.8 Transparent substrate technology

Visible $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ LEDs with typical operating wavelengths of 560 – 660 nm are grown lattice matched on GaAs substrates. Since the energy gap of GaAs is $E_g = 1.424$ eV ($\lambda_g = 870$ nm) at room temperature, GaAs substrates are absorbing at these emission wavelengths. As a result, the light emitted towards the substrate will be absorbed by the thick GaAs substrate. Thus, the extraction efficiency of AlGaInP/GaAs LEDs grown on GaAs substrates is low.

The extraction efficiency of AlGaInP LEDs can be increased substantially by removal of the GaAs substrate and bonding of the epitaxial layer to a GaP substrate (Kish *et al.*, 1994). GaP is an indirect semiconductor with an energy gap of $E_g = 2.24$ eV ($\lambda_g = 553$ nm). Thus, GaP does not absorb light with $\lambda > 553$ nm emitted by the AlGaInP active region.

The fabrication process of the AlGaInP LEDs wafer-bonded to a GaP substrate is illustrated schematically in Fig. 7.17. The AlGaInP double heterostructure (DH) is initially grown by OMVPE on a GaAs substrate. Subsequently, a thick GaP window layer (~ 50 μm) is grown on the DH by chloride VPE, a growth technique that allows for the low-cost growth of thick epitaxial layers. The GaAs substrate is then removed using a wet chemical selective etching process (Adachi and Oe, 1983; Kish *et al.*, 1994). During the removal of the GaAs substrate, the thick GaP window layer serves as a temporary mechanical support for the DH. The DH with the GaP window is then wafer-bonded to a GaP substrate. The DH with the GaP window is then wafer-bonded to a GaP substrate.

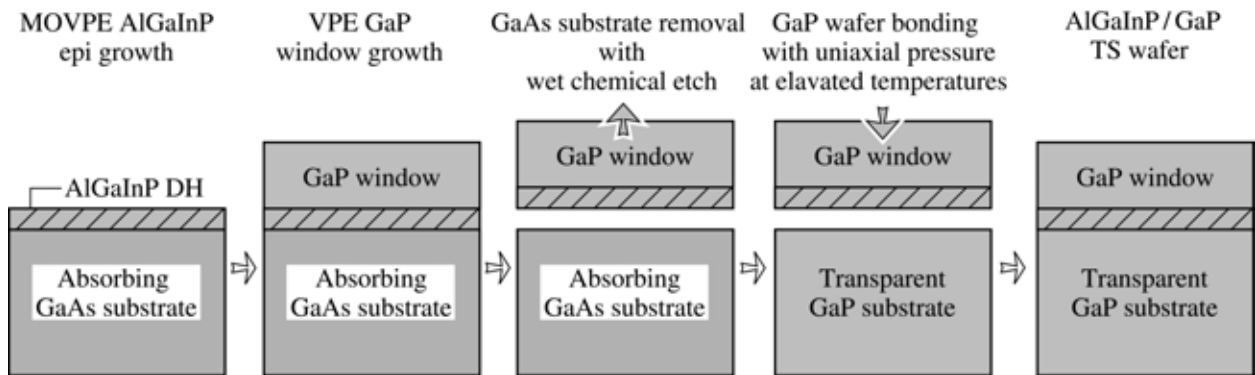


Fig. 7.17. Schematic fabrication process for wafer-bonded transparent substrate (TS) AlGaInP/GaP LEDs. After the selective removal of the original GaAs substrate, elevated temperature and uniaxial pressure are applied, resulting in the formation of a single TS LED wafer (adopted from Kish *et al.*, 1994).

Wafer bonding processes require a high degree of cleanliness, the lack of any particles

between the wafers, and the absence of native surface oxides on both wafers. Frequently, the gap between two wafers is filled with a contact liquid. Rotating the wafers at a high rate, spins out the contact liquid. Kish *et al.* (1995) and Hoefler *et al.* (1996) reported an AlGaInP-to-GaP wafer bonding process, suitable for 50 mm (2 inch) GaP substrates. Uniaxial pressure and elevated temperatures (750 – 1000 °C) are used in this process (Hoefler *et al.*, 1996). Kish *et al.* (1995) showed that the achievement of low-resistance ohmic conduction across wafer-bonded interfaces is critically dependent upon the crystallographic alignment of the bonded wafer surfaces, irrespective of the lattice mismatch between the surfaces. Furthermore, Kish *et al.* (1995) showed that the crystallographic surface orientation of the bonded surfaces must be nominally matched while simultaneously maintaining rotational alignment of the wafers. Low diode forward voltages of 2.2 V for AlGaInP/GaP LEDs are routinely achieved under high-volume manufacturing conditions with the process. The reliability of wafer-bonded LEDs is comparable to monolithic AlGaInP/GaAs LEDs. Usually the technical details of wafer bonding processes are proprietary and not known to the general public.

The forward voltage is a critical figure of merit for wafer-bonded p-n junction devices. A low voltage indicates true semiconductor-to-semiconductor chemical bonding and an absence of interfacial oxide layers. The forward current–voltage characteristic of commercial absorbing-substrate (AS) and transparent-substrate (TS) AlGaInP LEDs is shown in Fig. 7.18. Inspection of the figure reveals that TS LEDs have a higher forward voltage and series resistance compared with AS devices.

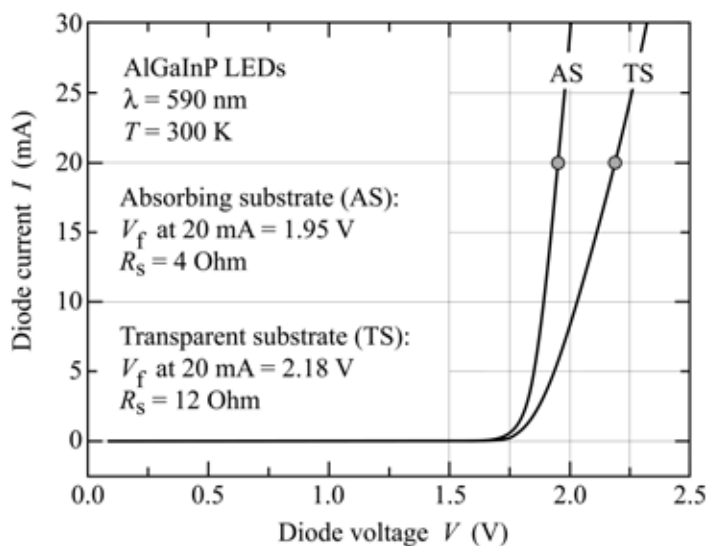


Fig. 7.18. Current–voltage characteristic, forward voltage, and series resistance of absorbing-substrate (GaAs) and transparent-substrate (GaP) LEDs with AlGaInP active regions.

The higher forward voltage in TS devices is probably due to either ohmic losses occurring at

the bonded interface or in the GaP substrate. The n-type doping concentration in the GaP substrate is kept moderately low, to minimize free-carrier absorption.

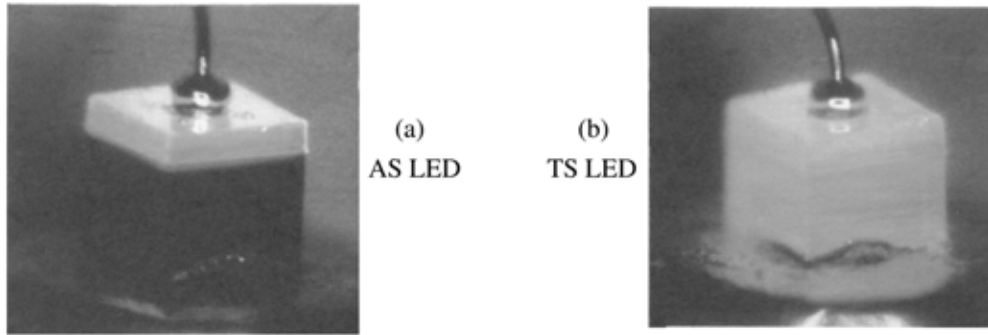


Fig. 7.19. (a) Amber AlGaInP LED with a GaP window layer and an absorbing GaAs substrate (AS). (b) Amber AlGaInP LED with a GaP window layer and a transparent GaP substrate (TS) fabricated by a wafer bonding. Conductive Ag-loaded die-attach epoxy can be seen at the bottom of the TS LED (after Kish and Fletcher, 1997).

Figure 7.19 compares AlGaInP/GaAs AS LEDs with AlGaInP/GaP TS LEDs emitting in the amber wavelength range. The substrate appears dark for the AS LED, in contrast to the TS LED. Transparent substrate AlGaInP/GaP LEDs have a factor of 1.5 – 3.0 higher external efficiency compared with AS AlGaInP/GaAs LEDs.

7.9 Anti-reflection optical coatings

Anti-reflection (AR) coatings are frequently used in communication LEDs to reduce the Fresnel reflection at the semiconductor–air interface. For normal incidence, the intensity reflection coefficient for normal incidence is given by

$$R = \frac{(\bar{n}_s - \bar{n}_{\text{air}})^2}{(\bar{n}_s + \bar{n}_{\text{air}})^2} \quad (7.28)$$

where \bar{n}_s and \bar{n}_{air} is the refractive index of the semiconductor and of air, respectively.

For normal incidence, Fresnel reflection at the semiconductor–air interface can be reduced to zero, if an AR coating cladding the semiconductor has the following parameters:

$$\text{Thickness: } \lambda/4 = \lambda_0 / (4 \bar{n}_{\text{AR}}) \quad \text{Refractive index: } \bar{n}_{\text{AR}} = \sqrt{\bar{n}_s \bar{n}_{\text{air}}} . \quad (7.29)$$

An AR coating with optimum thickness and refractive index is shown in Fig. 7.20. The refractive indices and the transparency ranges of different dielectric materials suitable for AR coatings are given in Table 7.2.

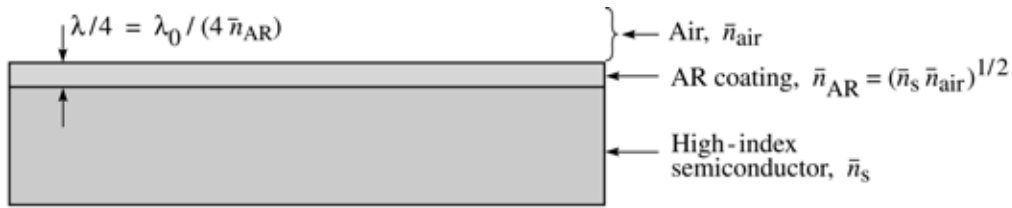


Fig. 7.20. Illustration of optimum thickness and refractive index of an anti-reflection (AR) coating.

Dielectric material	Refractive index	Transparency range
SiO ₂ (silica)	1.45	> 0.15 μm
Al ₂ O ₃ (alumina)	1.76	> 0.15 μm
TiO ₂ (titania)	2.50	> 0.35 μm
Si ₃ N ₄ (silicon nitride)	2.00	> 0.25 μm
ZnS (zinc sulphide)	2.29	> 0.34 μm
CaF ₂ (calcium fluoride)	1.43	> 0.12 μm

Table 7.2. Refractive index and transparency range of common dielectrics suitable as anti-reflection coatings (after Palik, 1998).

7.10 Epoxy dome

Virtually all LEDs, with the exception of communication LEDs used for fiber optic applications, are encapsulated with an optically transparent epoxy. The encapsulation occurs when the epoxy is in its liquid state. The epoxy is subsequently cured and becomes solid.

The refractive index contrast between the semiconductor and air is reduced by epoxy. The epoxy has a typical refractive index of 1.5, similar to silica. A *lower* index contrast at the semiconductor surface *increases* the angle of total internal reflection thereby enlarging the LED escape cone and extraction efficiency.

Furthermore, the epoxy usually has the shape of a hemisphere, as shown in Fig. 7.21 (a), so that the angle of incidence at the epoxy–air interface is always normal to the epoxy surface. As a result, total internal reflection does not occur at the epoxy–air interface.

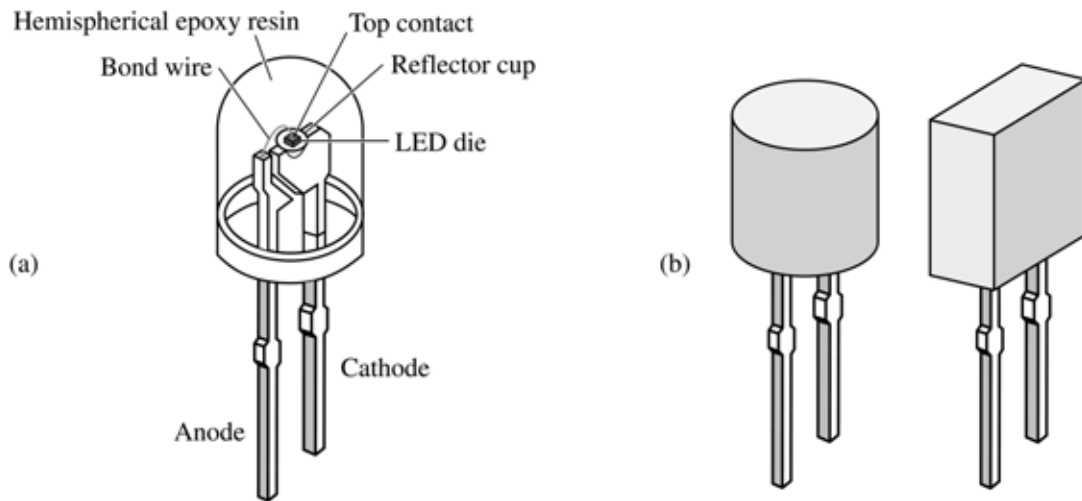


Fig. 7.21. Typical packages of LEDs. (a) LED with hemispherical epoxy dome. (b) LEDs with cylindrical and rectangular epoxy packages.

There are also many different types of LEDs that do *not* have a hemispherical shape for the epoxy encapsulant. Some LEDs have a rectangular or cylindrical shape with a planar front surface. Examples of such shapes are shown in Fig. 7.21 (b). Planar-surface LEDs are frequently used under circumstances where the intended viewing angle is close to normal incidence or where the LED is intended to blend in with a planar surface.

Epoxy domes provide protection against unwanted mechanical shock and chemicals. The epoxy also stabilizes the LED die and bonding wire. Finally the epoxy provides mechanical stability to the two metal leads of the LED and holds them in place.

The epoxy remains transparent for the longer-emission wavelengths and does not show any signs of degradation over a period of many years for long-wavelength visible and IR LEDs. However, it has been reported that the epoxy degrades and loses transparency in LEDs emitting at shorter wavelengths, i.e. in the blue, violet, and UV (Barton *et al.*, 1998).

7.11 Distributed Bragg reflectors

For LED structures on light-absorbing substrates, about 50 % of the light emitted by the active region is absorbed by the substrate. This represents a substantial loss. The absorption of light in the substrate can be avoided by placing a reflector between the substrate and the LED active layers. Light emanating from the active region towards the substrate will then be reflected and can escape from the semiconductor through the top surface.

Distributed Bragg reflectors (DBRs) are well suited for inclusion between the substrate and the active layers. The schematic LED structure with a DBR is shown in Fig. 7.22. LEDs with

DBRs were first demonstrated by Kato *et al.* (1991) in the AlGaAs/GaAs material system for LEDs emitting in the IR.

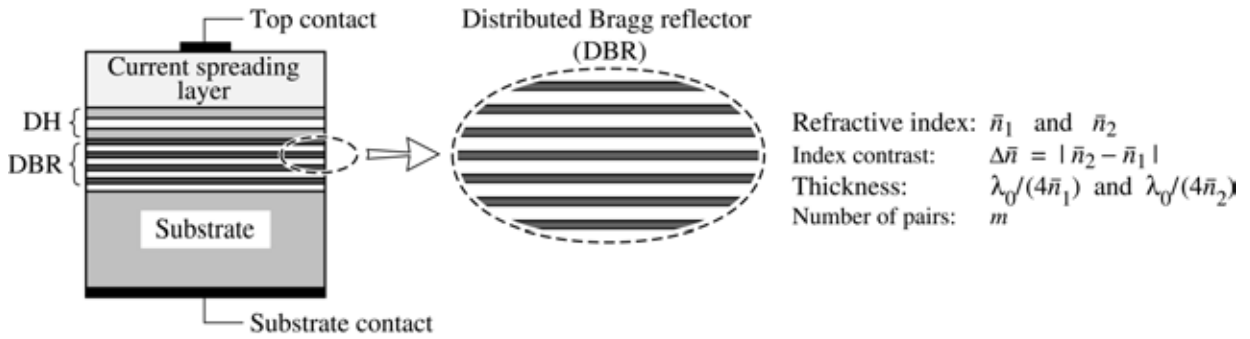


Fig. 7.22. LED with a distributed Bragg reflector (DBR) located between the substrate and the lower confinement layer.

A DBR is a multi-layer reflector consisting of typically 5 – 50 pairs of two materials with different refractive indices. As a result of the difference in refractive index, Fresnel reflection occurs at each of the interfaces. Usually the refractive index difference between the two materials is small so that the magnitude of the Fresnel reflection at one interface is also quite small. However, DBRs consist of many interfaces. More importantly, the thicknesses of the two materials is chosen in such a way, that *all reflected waves* are in *constructive interference*. For normal incidence, this condition is fulfilled when both materials have a thickness of a *quarter wavelength* of the reflected light, i.e.

$$t_{l,h} = \lambda_{l,h} / 4 = \lambda_0 / (4 \bar{n}_{l,h}) \quad (\text{normal incidence}) \quad (7.30)$$

where λ_0 is the vacuum Bragg wavelength of the light, $t_{l,h}$ is the thickness of the low-index (l) and high-index (h) material, and $\bar{n}_{l,h}$ is the refractive index of the low-index (l) and high-index (h) material. (The thickness $t_{l,h}$ given by Eq. (7.30) can also be an odd integer multiple of the value given in the equation.)

For an oblique angle of incidence, the wave vector can be separated into a parallel and a normal component. As in the normal-incident case, the thickness of the DBR layers must be a quarter wavelength for the wave vector component normal to the DBR layers. For an oblique angle of incidence $\Theta_{l,h}$, the optimum thicknesses for high reflectivity are given by

$$t_{l,h} = \lambda_{l,h} / (4 \cos \Theta_{l,h}) = \lambda_0 / (4 \bar{n}_{l,h} \cos \Theta_{l,h}) \quad (\text{oblique incidence}) . \quad (7.31)$$

(Again, the thickness $t_{1,h}$ given in Eq. (7.31) can also be an odd integer multiple of the value given by the equation.) For DBRs with sufficiently many quarter-wave pairs, reflectivities near 100 % can be obtained.

A DBR must fulfill several additional conditions. *First*, since a DH is grown on top of the DBR, the DBR must be lattice matched to the DH in order to avoid misfit dislocations. *Secondly*, to attain high-reflectivity DBRs, the DBR needs to be transparent at the wavelength of operation unless the DBR has a high-index contrast. High index-contrast DBRs (e.g. Si/SiO₂) yield high reflectivity, even if one of the materials is absorbing at the wavelength of interest. *Thirdly*, if the DBR is in the current path, the DBR must be conductive.

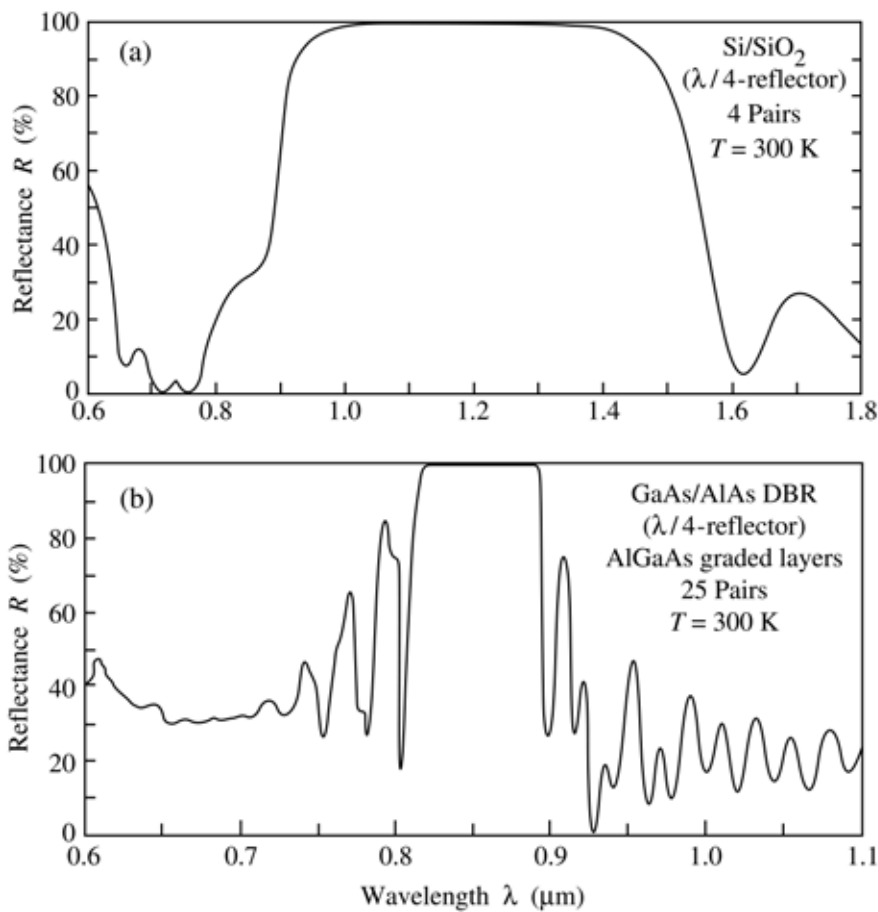


Fig. 7.23. Reflectance of two distributed Bragg reflectors (DBRs) versus wavelength. (a) four-pair Si/SiO₂ reflector with high index contrast. (b) 25-pair AlAs/GaAs reflector. The high-index-contrast DBR only needs four pairs to attain high reflectivity. Note that the stop band of the high-index-contrast DBR is wider compared with the low-contrast DBR.

The reflectance of a Si/SiO₂ and an AlAs/GaAs DBR versus wavelength are shown in Fig. 7.23. The Bragg wavelength is located in the center of the high-reflectivity band or **stop band**. Inspection of the figure reveals that (i) the reflectivity of high-contrast DBRs (Si/SiO₂) is much higher than the reflectivity of low-index contrast DBRs (AlAs/GaAs) for the same number

of quarter-wave pairs and that (ii) the width of the stop band of the high-index difference DBR is much wider than the stop band width of the low-contrast DBR.

Properties of DBRs can be calculated by the matrix method (Born and Wolf, 1989). The properties of DBRs have been analyzed in detail by Coldren and Corzine (1995), Yariv (1989), and Björk *et al.* (1995). Here, only a brief summary will be given.

Consider a distributed Bragg reflector consisting of m pairs of two dielectric, lossless materials with refractive index \bar{n}_l and \bar{n}_h , where the subscripts 'l' and 'h' denote the 'low-refractive-index' and the 'high-refractive-index' material, respectively. The thickness of the two layers is assumed to be a quarter wave, i.e. $L_l = \lambda_{\text{Bragg}} / (4\bar{n}_l)$, and $L_h = \lambda_{\text{Bragg}} / (4\bar{n}_h)$. The period of the DBR is $L_l + L_h$. The reflectivity of a single interface is given by Fresnel's equation for normal incidence

$$r = \frac{\bar{n}_h - \bar{n}_l}{\bar{n}_h + \bar{n}_l}. \quad (7.32)$$

Multiple reflections at the interfaces of the DBR and constructive interference of the multiple reflected waves increase the reflectivity with increasing numbers of pairs. The reflectivity has a maximum at the Bragg wavelength λ_{Bragg} . The reflectivity at the Bragg wavelength of a DBR with m quarter-wave pairs is given by (Coldren and Corzine, 1995)

$$R_{\text{DBR}} = |r_{\text{DBR}}|^2 = \left[\frac{1 - (\bar{n}_l / \bar{n}_h)^{2m}}{1 + (\bar{n}_l / \bar{n}_h)^{2m}} \right]^2. \quad (7.33)$$

The stop band of a DBR depends on the difference in refractive index of the two constituent materials, $\bar{n}_h - \bar{n}_l = \Delta\bar{n}$. The spectral width of the stop band is given by (Yariv, 1989)

$$\Delta\lambda_{\text{stop band}} = \frac{2 \lambda_{\text{Bragg}} \Delta\bar{n}}{\bar{n}_{\text{eff}}} \quad (7.34)$$

where \bar{n}_{eff} is the effective refractive index of the DBR. For efficient operation of the LED, the stop band should be wider than the emission spectrum of the active region.

The effective refractive index of the DBR can be calculated by requiring the same optical path length normal to the layers for the DBR and an effective medium. The effective refractive index is then given by

$$\bar{n}_{\text{eff}} = 2 \left(\frac{1}{\bar{n}_l} + \frac{1}{\bar{n}_h} \right)^{-1}. \quad (7.35)$$

The optical wave penetrates into the DBR only by a finite number of quarter-wave pairs. That is, a finite number out of the total number of quarter-wave pairs are effectively reflecting the wave. The effective number of pairs “seen” by the wave electric field is given by (Coldren and Corzine, 1995)

$$m_{\text{eff}} \approx \frac{1}{2} \frac{\bar{n}_h + \bar{n}_l}{\bar{n}_h - \bar{n}_l} \tanh \left(2 m \frac{\bar{n}_h - \bar{n}_l}{\bar{n}_h + \bar{n}_l} \right). \quad (7.36)$$

For thick DBRs ($m \rightarrow \infty$), the tanh function approaches unity and one obtains

$$m_{\text{eff}} \approx \frac{1}{2} \frac{\bar{n}_h + \bar{n}_l}{\bar{n}_h - \bar{n}_l}. \quad (7.37)$$

At the Bragg wavelength ($\lambda = \lambda_{\text{Bragg}}$), the phase change of the reflected wave is zero. In the vicinity of the Bragg wavelength ($\lambda \approx \lambda_{\text{Bragg}}$), the phase of the reflected wave changes *linearly* with wavelength.

The reflectivity of the DBR depends on the polar angle of incidence and on the wavelength. Although an analytic result can be obtained for the reflectivity of a DBR at the Bragg wavelength for normal incidence, the reflectivity of a DBR at an arbitrary wavelength and an arbitrary angle of incidence can only be calculated numerically. If light from an isotropic source is reflected by a DBR, the reflected intensity can be obtained by integration over all angles. The reflectance at a certain wavelength λ is then given by

$$R(\lambda) = \frac{\int_0^{\pi/2} R(\lambda, \Theta) 2\pi \sin \Theta \, d\Theta}{\int_0^{\pi/2} 2\pi \sin \Theta \, d\Theta} = \frac{1}{2\pi} \int_0^{\pi/2} R(\lambda, \Theta) 2\pi \sin \Theta \, d\Theta. \quad (7.38)$$

The total light intensity reflected by the DBR is calculated by

$$I_r = \int_{\lambda} I_i(\lambda) R(\lambda) \, d\lambda \quad (7.39)$$

where $I_i(\lambda)$ is the emission intensity spectrum of the active region incident on the DBR and I_r is

the intensity reflected by the DBR. For isotropic emitters such as the active region of LEDs, it can be assumed that the emission spectrum incident on the DBR is independent of the emission angle.

An efficient DBR will be optimized in such a way that it maximizes the intensity of the reflected light. In addition, the escape of the light reflected by the DBR from the LED die must be taken into account. Maximizing the LED extraction efficiency using a DBR is a complicated problem that cannot be solved analytically. *Ray tracing computer software* is used to maximize the extraction efficiency in LED structures with DBRs.

A priori, it is not clear that the use of conventional DBRs provides the best extraction efficiency. DBRs with layer thicknesses thinner or thicker than a quarter-wave should also be considered for maximization of the extraction efficiency. Such variable-thickness DBRs have a *lower reflectivity* but a *wider stop-band width* compared with quarter-wave DBRs. For active regions with broad emission spectra, such variable period DBRs can be advantageous.

Ideally, the layers comprising the DBR are transparent. Layers of **transparent DBRs** have negligible absorption losses. However, transparent materials may not always be available so that **absorbing** materials must be used. Such **absorbing DBRs** have a maximum reflectivity of less than 100 %, even if an infinite number of pairs were used.

An example of a partially absorbing Si/SiO₂ DBR is shown in Fig. 7.23. Silicon absorbs light for $\lambda < 1.1 \mu\text{m}$, i.e. for $h\nu > E_g$. However, the results shown in the figure demonstrate that very high reflectivities can be attained at $1.0 \mu\text{m}$, where Si is absorbing. This is due to the high index contrast between Si and SiO₂.

Transparent and absorbing DBRs lattice matched to GaAs are used in the AlGaInP/GaAs material system. This material system is used for high-efficiency visible LEDs emitting at $\lambda > 550 \text{ nm}$ (green, yellow, amber, orange, and red). The properties of transparent and absorbing DBRs used in this material system have been compiled by Kish and Fletcher (1997) and are summarized in Table 7.3. Inspection of the table reveals that the absorbing Al_{0.5}In_{0.5}P/GaAs DBRs have the advantage of a high refractive index contrast. However, the absorbing nature of the DBR imposes an upper limit on the maximum reflectivity. High-contrast DBRs have a *wider* stop-band width. The transparent Al_{0.5}In_{0.5}P/(AlGa)_{0.5}In_{0.5}P DBRs have the advantage of negligible optical losses. However, many pairs are needed to attain high reflectivity and the stop-band width is narrower as in high index-contrast DBRs.

Material system	Bragg wavelength	\bar{n}_{low}	\bar{n}_{high}	$\Delta\bar{n}$	Transparency range
$\text{Al}_{0.5}\text{In}_{0.5}\text{P}/\text{GaAs}$	590 nm	3.13	3.90	0.87	> 870 nm (lossy)
$\text{Al}_{0.5}\text{In}_{0.5}\text{P}/\text{Ga}_{0.5}\text{In}_{0.5}\text{P}$	590 nm	3.13	3.74	0.61	> 649 nm (lossy)
$\text{Al}_{0.5}\text{In}_{0.5}\text{P}/(\text{Al}_{0.3}\text{Ga}_{0.7})_{0.5}\text{In}_{0.5}\text{P}$	615 nm	3.08	3.45	0.37	> 592 nm
$\text{Al}_{0.5}\text{In}_{0.5}\text{P}/(\text{Al}_{0.4}\text{Ga}_{0.6})_{0.5}\text{In}_{0.5}\text{P}$	590 nm	3.13	3.47	0.34	> 576 nm
$\text{Al}_{0.5}\text{In}_{0.5}\text{P}/(\text{Al}_{0.2}\text{Ga}_{0.5})_{0.5}\text{In}_{0.5}\text{P}$	570 nm	3.15	3.46	0.31	> 560 nm
AlAs/GaAs	900 nm	2.97	3.54	0.57	> 870 nm
SiO_2/Si	1300 nm	1.46	3.51	2.05	> 1106 nm

Table 7.3. Properties of distributed Bragg reflector (DBR) materials used for visible and infrared LED applications. The DBRs marked as ‘lossy’ are absorbing at the Bragg wavelength (data after Adachi, 1990; Adachi *et al.*, 1994; Kish and Fletcher, 1997; Babic *et al.*, 1999; Palik, 1998).

In practice, transparent layers are used at and near the top (epitaxial side) of the DBR whereas absorbing layers are used towards the bottom (substrate side) of the DBR. For DBRs used in manufactured AlGaInP/GaAs LEDs, each of the layers is different and optimized to keep the number of pairs low, the absorption low, and the reflectivity spectrum broad (Streubel, 2000).

Table 7.3 also shows properties of the AlAs/GaAs and the SiO_2/Si material systems. The SiO_2/Si material system is an example of a high index-contrast system. However, it cannot be used for current conduction due to the insulating nature of the SiO_2 . The AlAs/GaAs material system is used in resonant-cavity LEDs and vertical-cavity surface-emitting lasers emitting in the range 880 – 980 nm.

A DBR that is resonant with the peak emission wavelength, is not necessarily the optimum reflector for an absorbing substrate (AS) LED. Although a DBR has a high reflectivity for normal incidence, it rapidly decreases for off-normal angles of incidence. Since the steradian increases with angle according to a *sine* function, it is advantageous to shift the normal-incidence resonance wavelength of the DBR to wavelengths longer than the peak emission wavelength.

Different strategies have been employed to optimize DBRs. Chiou *et al.* (2000) used a composite DBR in an AlGaInP LED, that is, two types of DBRs stacked on top of each other, namely a non-absorbing $(\text{Al}_{0.4}\text{Ga}_{0.6})_{0.5}\text{In}_{0.5}\text{P}/\text{Al}_{0.5}\text{In}_{0.5}\text{P}$ DBR resonant at the peak emission wavelength of 590 nm and an additional high-contrast absorbing AlAs/GaAs DBR. This DBR was located below the DBR made of phosphides. The arsenide DBR was resonant at a wavelength about 10 % longer than the peak emission wavelength in order to reflect light

incident at off-normal angles. The authors found a substantial improvement in light output with the composite DBR.

Non-periodic DBRs have a wider stop band and thereby a high reflectivity over a wider range of angles. To find optimum non-periodic DBR structures, numerical simulations have been conducted (see, for example, Li *et al.*, 1999).

7.12 Current-blocking layers

In conventional DH LEDs with small top contacts and large backside contacts, the current injected by the top contact enters the active region essentially under the top contact. The extraction of light generated in the active region is thus strongly hindered by the opaque metal contact. One possibility to alleviate this problem is the use of a thick current spreading-layer. Another possibility is the use of a current-blocking layer. This layer blocks the current from entering the active region below the top contact. The current is deflected away from the top contact, thus allowing for much higher extraction efficiency.

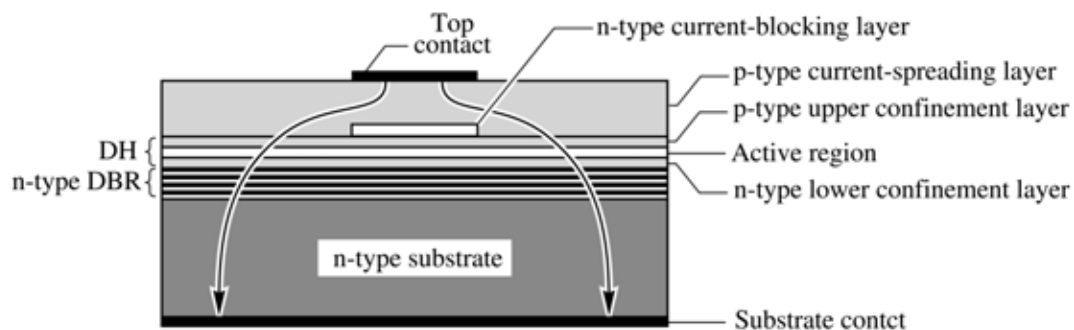


Fig. 7.24. LED with an n-type current-blocking layer located on the upper confinement layer. Light emission occurs in the regions not covered by the opaque top ohmic contact. The LED is fabricated by *epitaxial regrowth*. After growth of the current-blocking layer, the wafer is taken out of the growth system for etching. The wafer is then re-introduced into the epitaxial system for growth of the current-spreading layer.

The schematic structure of an LED employing a current-blocking layer is shown in Fig. 7.24. The blocking layer is located on top of the upper confinement layer and has approximately the same size as the top metal contact. The current-blocking layer has n-type conductivity and is embedded in material with p-type conductivity. Owing to the p-n junction surrounding the current-blocking layer, the current flows around the current-blocking layer as indicated in Fig. 7.24.

Current-blocking layers can be fabricated by *epitaxial regrowth*. In this process, the DH and the thin n-type current-blocking layer are grown over the entire wafer surface. Subsequently, the wafer is taken out of the growth system for etching. The regions to be etched are defined by photolithography. The entire blocking layer is etched away *except* the region where the top ohmic contact is going to be located, as shown in Fig. 7.24. Frequently, the current-blocking-layer etch is *selective* so that it does not etch into the upper confinement layer. Subsequently, the wafer is returned to the growth system for resumption of epitaxial growth, i.e. for the growth of the current-spreading layer.

Regrowth processes are expensive due to the reduction of the device and wafer yield that usually accompanies regrowth processes. Cleaning of the wafer surface after completion of the first growth and the etching is critical. Defects occurring at the regrowth interface can lead to a reduction in yield. Therefore, processes requiring epitaxial regrowth are more expensive and are not suited for low-cost devices such as visible LEDs. However, for more expensive devices such as communication LEDs, epitaxial regrowth processes are used.

In AlGaInP LEDs, n-type GaAs has been used as the current-blocking layer. The n-type GaAs layer is located on top of the AlGaInP upper confinement layer. The GaAs current-blocking layer is lattice matched to the underlying AlGaInP confinement layer. Selective wet chemical etches are available that etch the GaAs but do not etch the AlGaInP (Adachi and Oe, 1983).

Vertical-cavity surface-emitting lasers (VCSELs) also employ current-blocking layers to funnel the current to the active region located between the mirrors of the laser. However, oxygen or hydrogen implantation rather than epitaxial regrowth is used in VCSELs to form current-blocking layers. The implantation depth is limited so that the lateral resistance can become substantial in large-area contacts.

7.13 Reflective and transparent contacts

Metallic ohmic contacts are practically opaque for thicknesses > 50 nm. Thus, light incident on the top or bottom contact of an LED will not be transmitted through the contact.

Annealing and alloying forms low-resistance ohmic contacts. Typical annealing temperatures for alloyed contacts are between 375 and 450 °C for III–V arsenides and III–V phosphides and up to 600 °C for III–V nitrides. During the annealing process, the metal surface changes from a smooth to a rough appearance and a concomitant decrease in the optical reflectivity results.

In contrast, non-alloyed contacts are just deposited on the semiconductor without annealing.

Highly doped semiconductor surface layers are needed to obtain good ohmic I – V characteristics for such non-alloyed ohmic contacts. However, even for highly doped semiconductors, the contact resistance of non-alloyed contacts is usually higher than for alloyed contacts.

In LEDs with transparent substrates, e.g. AlGaInP LEDs on GaP, light emanating from the active region is incident on the substrate contact. To increase the reflectance of the backside, an ohmic contact geometry covering only a small fraction of the substrate surface can be used, such as a multiple-stripe or a ring-shaped contact. Using Ag-loaded conductive epoxy to attach the LED die to the package provides a high-reflectivity material in the regions not covered by the ohmic contact.

The die-attach epoxy can also serve as a reflector in LEDs grown on a transparent material, e.g. GaInN LEDs grown on sapphire substrates. Conductive Ag-loaded epoxy has a high conductivity as well as high reflectance. Such a highly reflective epoxy can increase the extraction efficiency in LEDs grown on transparent substrates.

Very thin metal contacts are semitransparent, i.e. a small percentage of the incident light is transmitted through the metal. Most metal contacts have a transmittance of approximately 50 % at a metal film thickness of 5 – 10 nm. The exact value of the transmittance needs to be calculated by taking into account the real and imaginary part of the refractive index (see, for example, Palik, 1998). However, very thin metallic contacts may form an island structure rather than a single continuous film. Furthermore, the electrical resistance of thin metal films can be large, in particular if an island structure is formed.

There are ohmic contacts that are transparent to visible light. These contacts include indium-tin oxide, frequently referred to as ITO (Ray *et al.*, 1983; Shen *et al.*, 1998; Margalith *et al.*, 1999; Mergel *et al.*, 2000; Shin *et al.*, 2001). The materials can be considered as a tin oxide semiconductor that is doped with indium. Indium substitutes for Sn and therefore acts as an acceptor. Generally the specific contact resistance of ITO contacts is higher than the contact resistance of alloyed metal contacts.

7.14 Flip-chip packaging

For LEDs with two top contacts, such as GaInN/GaN LEDs grown on sapphire substrates, both regular packaging (epi-side up) and flip-chip packaging is used. Flip-chip packaging using **solder-bump bonding** is a more expensive packaging process compared with regular packaging where the LED top contact pads are contacted by wire bonding. An advantage of flip-chip

packaging of GaInN/GaN LEDs is that the metal pads are not hindering the extraction of light radiating from the active region.

In epi-side up packaged GaInN LEDs, a large-area p-type top contact results in a uniform current distribution in the active region. However, such a large top contact will hinder the extraction of light. This fundamental trade-off can be avoided by flip-chip packaging. This is especially important for power devices.

References

- Adachi S. "Properties of Gallium Arsenide" *EMIS Datareview Ser.*, **2**, 513 INSPEC (IEEE, New York, 1990)
- Adachi S., Kato H., Moki A., and Ohtsuka K. "Refractive index of $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ quaternary alloys" *J. Appl. Phys.* **75**, 478 (1994)
- Adachi S. and Oe K. "Chemical etching characteristics of (001) GaAs" *J. Electrochem. Soc.* **130**, 2427 (1983)
- Babic D. I., Piprek L., and Bowers J. E. in "Vertical-cavity surface-emitting lasers" edited by C. W. Wilmsen, H. Temkin, and L. A. Coldren (Cambridge University Press, Cambridge, 1999)
- Barton D. L., Osinski M., Perlin P., Helms C. J., and Berg N. H. "Life tests and failure mechanisms of GaN/AlGaIn/InGaIn light-emitting diodes" *Proceedings of SPIE* **3279**, 17 (1998)
- Björk G., Yamamoto Y., and Heitmann H. "Spontaneous emission control in semiconductor microcavities" in "Confined electrons and photons" edited by Burstein E. and Weisbuch C. (Plenum Press, New York, 1995)
- Born M. and Wolf E. "Principles of optics" 6th edition (Pergamon Press, New York, 1989)
- Carr W. N. and Pittman G. E. "One-Watt GaAs p-n junction infrared source" *Appl. Phys. Lett.* **3**, 173 (1963)
- Casey Jr. H. C. and Panish M. B. "Heterostructure lasers. Part A: Fundamental principles" pp. 46, 47, and 175 (Academic Press, San Diego, 1978)
- Chiou S. W., Lee C. P., Huang C. K., and Chen C. W. "Wide angle distributed Bragg reflectors for 590 nm amber AlGaInP light-emitting diodes" *J. Appl. Phys.* **87**, 2052 (2000)
- Coldren L. A. and Corzine S. W. "Diode lasers and photonics integrated circuits" (John Wiley and Sons, New York, 1995)

- Fletcher R. M., Kuo C. P., Osentowski T. D., Huang K. H., and Craford M. G. "The growth and properties of high performance AlInGaP emitters using lattice mismatched GaP window layers" *J. Electronic Materials* **20**, 1125 (1991a)
- Fletcher R. M., Kuo C. P., Osentowski T. D., and Robbins V. M. "Light-emitting diode with an electrically conductive window" US Patent 5,008,718 (1991b)
- Franklin A. R. and Newman R. "Shaped electroluminescent GaAs diodes" *J. Appl. Phys.* **35**, 1153 (1964)
- Groves W. O. and Epstein A. S. "Epitaxial deposition of III–V compounds containing isoelectronic impurities" US patent 4,001,056 (1977)
- Groves W. O., Herzog A. H., and Craford M. G. "Process for the preparation of electroluminescent III–V materials containing isoelectronic impurities" US patent Re. 29,648 (1978a)
- Groves W. O., Herzog A. H., and Craford M. G. "GaAsP electroluminescent device doped with isoelectronic impurities" US patent Re. 29,845 (1978b)
- Guo X. and Schubert E. F. "Current crowding and optical saturation effects in GaInN/GaN light-emitting diodes grown on insulating substrates" *Appl. Phys. Lett.* **78**, 3337 (2001)
- Hoefer G. E., Vanderwater D. A., DeFever D. C., Kish F. A., Camras M. D., Steranka F. M., and Tan I.-H. "Wafer bonding of 50-mm diameter GaP to AlGaInP–GaP light-emitting diode wafers" *Appl. Phys. Lett.* **69**, 803 (1996)
- Ioffe Institute (Saint Petersburg, Russia) database on compound semiconductors available at < <http://www.ioffe.rssi.ru/SVA/NSM/Semicond/> > (2002)
- Jeon S.-R., Song Y.-H., Jang H.-J., Yang G. M., Hwang S. W., and Son S. J. "Lateral current spreading in GaN-based light-emitting diodes utilizing tunnel contact junctions" *Appl. Phys. Lett.* **78**, 3265 (2001)
- Kato T., Susawa H., Hirotani M., Saka T., Ohashi Y., Shichi E., and Shibata S. "GaAs/GaAlAs surface emitting IR LED with Bragg reflector grown by MOCVD" *J. Cryst. Growth* **107**, 832 (1991)
- Kim O. K. and Bonner W. A. "Infrared reflectance and absorption of n-type InP" *Journal of Electronic Materials* **12**, 827 (1983)
- Kish F. A. and Fletcher R. M. "AlGaInP light-emitting diodes" in "High brightness light-emitting diodes" edited by G. B. Stringfellow and M. G. Craford *Semiconductors and Semimetals* Vol. **48** (Academic, San Diego, 1997)

- Kish F. A., Steranka F. M., DeFevre D. C., Vanderwater D. A., Park K. G., Kuo C. P., Osentowski T. D., Peanasky M. J., Yu J. G., Fletcher R. M., Steigerwald D. A., Craford M. G., and Robbins V. M. "Very high-efficiency semiconductor wafer-bonded transparent-substrate $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}/\text{GaP}$ light-emitting diodes" *Appl. Phys. Lett.* **64**, 2839 (1994)
- Kish F. A., Vanderwater D. A., Peanasky M. J., Ludowise M. J., Hummel S. G., and Rosner S. J. "Low-resistance ohmic conduction across compound semiconductor wafer-bonded interfaces" *Appl. Phys. Lett.* **67**, 2060 (1995)
- Knox R. S. "Theory of excitons" (Academic, New York, 1963)
- Krames M. R. Ochiai-Holcomb M., Höfler G. E., Carter-Coman C., Chen E. I., Tan I.-H., Grillo P., Gardner N. F., Chui H. C., Huang J.-W., Stockman S. A., Kish F. A., Craford M. G., Tan T. S., Kocot C. P., Hueschen M., Posselt J., Loh B., Sasser G., and Collins D. "High-power truncated-inverted-pyramid $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}/\text{GaP}$ light-emitting diodes exhibiting > 50 % external quantum efficiency" *Appl. Phys. Lett.* **75**, 2365 (1999)
- Kuo C. P., Fletcher R. M., Osentowski T. D., Lardizabal M. C., Craford M. G., and Robbins V. M. "High performance AlGaInP visible light emitting diodes" *Appl. Phys. Lett.* **57**, 2937 (1990)
- Li H., Gu G., Chen H., and Zhu S. "Disordered dielectric high reflectors with broadband from visible to infrared" *Appl. Phys. Lett.* **74**, 3260 (1999)
- Loebner E. E. "The future of electroluminescent solids in display applications" *Proceedings of the IEEE* **61**, 837 (1973)
- Margalith T., Buchinsky O., Cohen D. A., Abare A. C., Hansen M., DenBaars S. P., and Coldren L. A. "Indium tin oxide contacts to gallium nitride optoelectronic devices" *Appl. Phys. Lett.* **74**, 3930 (1999)
- Mergel D., Stass W., Ehl G., and Barthel D. "Oxygen incorporation in thin films of $\text{In}_2\text{O}_3:\text{Sn}$ prepared by radio frequency sputtering" *J. Appl. Phys.* **88**, 2437 (2000)
- Moyer C. D. "Photon recycling light emitting diode" US patent 4,775,876 (1988)
- Nishizawa J., Koike M., and Jin C. C. "Efficiency of GaAlAs heterostructure red light-emitting diodes" *J. Appl. Phys.* **54**, 2807 (1983)
- Nuese C. J., Tietjen J. J., Gannon J. J., and Gossenberger H. F. "Optimization of electroluminescent efficiencies for vapor-grown GaAsP diodes" *J. Electrochem Soc.: Solid State Science* **116**, 248 (1969)
- Palik E. D. "Handbook of optical constants of solids" (Academic Press, San Diego, 1998)

- Pankove J. I. "Optical processes in semiconductors" p. 75 and section on Urbach tail (Dover, New York, 1971)
- Ray S., Banerjee R., Basu N., Batabyal A. K., and Barua A. K. "Properties of tin doped indium oxide thin films prepared by magnetron sputtering" *J. Appl. Phys.* **54**, 3497 (1983)
- Schmid W., Eberhard F., Jager R., King R., Joos J., and Ebeling K. "45 % quantum-efficiency light-emitting diodes with radial outcoupling taper" *Proceedings SPIE* **3938**, 90 (2000)
- Schmid W., Scherer M., Jager R., Strauss P., Streubel K., and Ebeling K. "Efficient light-emitting diodes with radial outcoupling taper at 980 and 630 nm emission wavelength" *Proceedings SPIE* **4278**, 109 (2001)
- Schmid W., Scherer M., Karnutsch C., Plohl A., Wegleiter W., Schad S., Neubert B., and Streubel K. "High-efficiency red and infrared light-emitting diodes using radial outcoupling taper" *IEEE J. Select. Topics Quant. Electronics* **8**, 256 (2002)
- Schnitzer I., Yablonovitch E., Caneau C., Gmitter T. J., and Scherer A. "30 % external quantum efficiency from surface-textured, thin-film light-emitting diodes" *Appl. Phys. Lett.* **63**, 2174 (1993)
- Schubert E. F., Goebel E. O., Horikoshi Y., Ploog K., and Queisser H. J. "Alloy broadening in photoluminescence spectra of AlGaAs" *Phys. Rev. B* **30**, 813 (1984)
- Schubert E. F., Goepfert I. D., Grieshaber W., and Redwing J. M. "Optical properties of Si-doped GaN" *Appl. Phys. Lett.* **71**, 921 (1997)
- Shin J. H., Shin S. H., and Park J. I., and Kim H. H. "Properties of dc magnetron sputtered indium tin oxide films on polymeric substrates at room temperature" *J. Appl. Phys.* **89**, 5199 (2001)
- Sheu J. K., Su Y. K., Chi G. C., Jou M. J., and Chang C. M. "Effects of thermal annealing on the indium tin oxide Schottky contacts of n-GaN" *Appl. Phys. Lett.* **72**, 3317 (1998)
- Sinzinger S. and Jahns J. "Microoptics" (Wiley-VCH, New York, 1999)
- Streubel K., personal communication (2000)
- Sugawara H., Ishakawa M., and Hatakoshi G. "High-efficiency InGaAlP/GaAs visible light-emitting diodes" *Appl. Phys. Lett.* **58**, 1010 (1991)
- Sugawara H., Ishakawa M., Kokubun Y., Nishikawa Y., Naritsuka S., Itaya K., Hatakoshi G., Suzuki M. "Semiconductor light-emitting device" United States Patent 5,153,889, issued Oct. 6 (1992a)
- Sugawara H., Itaya K., Nozaki H., and Hatakoshi G. "High-brightness InGaAlP green light-emitting diodes" *Appl. Phys. Lett.* **61**, 1775 (1992b)

- Swaminathan V. and Macrander A. T. “Materials aspects of GaAs and InP based structures” (Prentice Hall, Englewood Cliffs, 1991)
- Thompson G. H. B. “Physics of semiconductor laser devices” (John Wiley and Sons, New York, 1980)
- Urbach F. “The long-wavelength edge of photographic sensitivity of the electronic absorption of solids” *Phys. Rev.* **92**, 1324 (1953)
- Walukiewicz W., Lagowski J., Jastrzebski L., Rava P., Lichtensteiger M., Gatos C. H., and Gatos H. C. “Electron mobility and free-carrier absorption in InP; determination of the compensation ratio” *J. Appl. Phys.* **51**, 2659 (1980)
- Weily J. D. and DiDomenico Jr. M. “Free-carrier absorption in n-type GaP” *Phys. Rev. B* **1**, 1655 (1970)
- Windisch R., Schoberth S., Meinlschmidt S., Kiesel P., Knobloch A., Heremans P., Dutta B., Borghs G., and Doehler G. H. “Light propagation through textured surfaces” *J. Opt. A: Pure Appl. Opt.* **1**, 512 (1999)
- Windisch R., Dutta B., Kuijk M., Knobloch A., Meinlschmidt S., Schoberth S., Kiesel P., Borghs G., Doehler G. H., and Heremans P. “40 % efficient thin-film surface textured light-emitting diodes by optimization of natural lithography” *IEEE Trans. Electron Dev.* **47**, 1492 (2000)
- Windisch R., Rooman C., Kuijk M., Borghs G., and Heremans P. “Impact of texture-enhanced transmission on high-efficiency surface-textured light-emitting diodes” *Appl. Phys. Lett.* **79**, 2315 (2001)
- Windisch R., Rooman C., Dutta B., Knobloch A., Borghs G., Doehler G. H., and Heremans P. “Light-extraction mechanisms in high-efficiency surface-textured light-emitting diodes” *IEEE J. Selected Topics Quant. Electronics* **8**, 248 (2002)
- Yariv A. “Quantum electronics” 3rd edition (John Wiley and Sons, New York, 1989)

8

Visible-spectrum LEDs

Originally, LEDs were exclusively used for low-brightness applications such as indicator lamps. In these applications, the efficiency and the overall optical power of the LED are *not* of primary importance. However, in more recent applications, for example traffic light applications, the light emitted by LEDs must be seen even in bright sunlight and from a considerable distance. LEDs with high efficiency and brightness are required for such applications.

In this chapter, low-brightness as well as high-brightness LEDs are discussed. GaAsP and nitrogen-doped GaAsP LEDs are suitable only for low-brightness applications. AlGaAs LEDs are suitable for low- as well as high-brightness applications. AlGaInP and GaInN LEDs are used in high-brightness applications.

8.1 The GaAsP, GaP, GaAsP:N, and GaP:N material systems

The $\text{GaAs}_{1-x}\text{P}_x$ and $\text{GaAs}_{1-x}\text{P}_x\text{:N}$ material system is used for emission in the red, orange, yellow, and green wavelength range. The GaAsP system is lattice mismatched to GaAs substrates, resulting in a relatively low internal quantum efficiency. As a result these LEDs are suitable for low-brightness applications only.

$\text{GaAs}_{1-x}\text{P}_x$ was one of the first material systems used for visible-spectrum LEDs (Holonyak and Bevacqua 1962; Holonyak *et al.* 1963, 1966; Pilkuhn and Rupprecht 1965; Nuese *et al.* 1966; Wolfe *et al.*, 1965). In the early 1960s, GaAs substrates were already available. Bulk growth of GaAs substrates was initiated in the 1950s and epitaxial growth by LPE and VPE started in the 1960s. As phosphorus is added to GaAs, the ternary alloy $\text{GaAs}_{1-x}\text{P}_x$, or briefly GaAsP is formed. The addition of phosphorus increases the bandgap of GaAs, which emits in the infrared at 870 nm. The visible wavelength range starts at about 750 nm, so that a small amount of phosphorus is sufficient to attain visible-spectrum light emitters. Note, however, that the sensitivity of the human eye is low at the edges of the visible spectrum.

A significant problem with GaAsP LEDs is the lattice mismatch between the GaAs substrate and the GaAsP epitaxial layer. A large mismatch exists between GaAs and GaP (about 3.6 %) so

that many misfit dislocations occur when the critical thickness of GaAsP on GaAs is exceeded. As a result, the luminescence efficiency decreases substantially in GaAsP with increasing phosphorus content. GaAsP LEDs are therefore useful for low-brightness applications only.

It has been realized early in the GaAsP work that the lattice mismatch between the GaAs substrate and the GaAsP epilayer reduces the radiative efficiency. It was also found that the radiative efficiency of the active p-n junction layer strongly depends on the growth conditions and, in particular, on the thickness of the GaAsP buffer layer (Nuese *et al.* 1969). A thick buffer layer reduces the dislocation density by annihilation of misfit dislocations. However, the dislocation density does not approach the low dislocation density of GaAs substrates, so that even with thick GaAsP buffer layers the dislocation density is substantial.

The band structure of GaAs, GaAsP, and GaP is shown schematically in Fig. 8.1. The figure shows that GaAsP is a direct-gap semiconductor for low phosphorus mole fractions. Beyond the direct-indirect crossover occurring at phosphorus mole fractions of about 45 – 50 %, the semiconductor becomes indirect and the radiative efficiency drops rapidly (Holonyak *et al.*, 1963, 1966). GaP is an indirect-gap semiconductor and therefore is unsuitable as an efficient LED material.

GaAsP and GaP LEDs are frequently doped with isoelectronic impurities such as nitrogen (Grimmeiss and Scholz, 1964; Logan *et al.*, 1967a, 1967b, 1971; Craford *et al.*, 1972; Groves and Epstein 1977; Groves *et al.* 1978a, 1978b). The isoelectronic impurities form an optically active level within the forbidden gap of the semiconductor so that carriers recombine radiatively via the nitrogen levels, as indicated in Fig. 8.1.

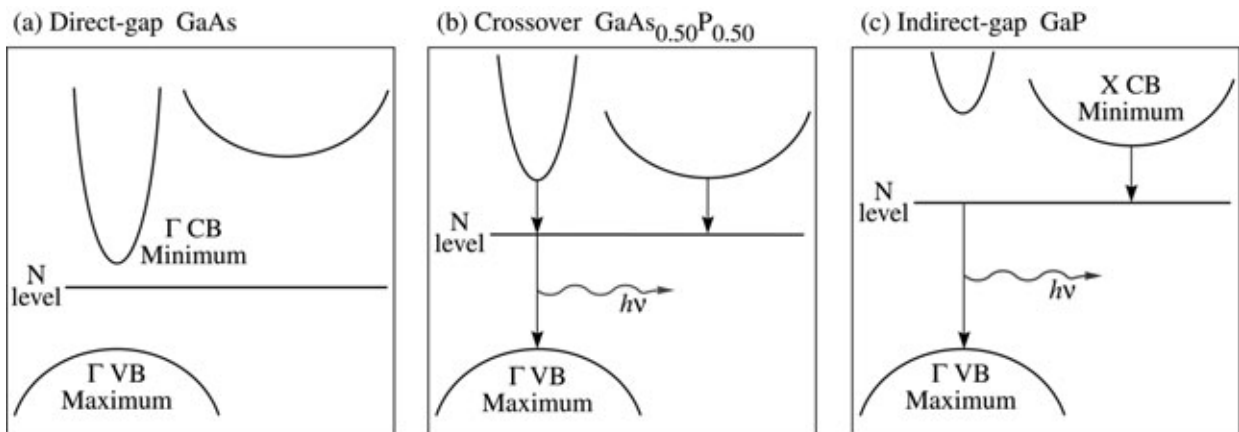


Fig. 8.1. Schematic band structure of GaAs, GaAsP, and GaP. Also shown is the nitrogen level. At a P mole fraction of about 45 – 50 %, the direct – indirect crossover occurs.

Isoelectronically doped LEDs are also interesting from a fundamental point of view. They are one of the first practical applications of Heisenberg's uncertainty principle. Isoelectronic impurities have an electronic wave function that is strongly localized in position space (small Δx). Therefore, the wave function is delocalized in momentum space (large Δp). Since the level is delocalized in momentum space, two vertical transitions can occur via the isoelectronic trap, with one of them being radiative. Physically speaking, the change in momentum occurring when an electron makes a transition from the indirect X valley of the conduction band to the central Γ valley of the valence band, the momentum change is absorbed by the isoelectronic impurity atom.

The emission wavelength of undoped and nitrogen-doped GaAsP is shown in Fig. 8.2 (Craford *et al.* 1972). The emission energy of GaAsP and GaP doped with the isoelectronic impurity nitrogen is below the bandgap of the semiconductor. Figure 8.2 illustrates that the emission energy is about 50 – 150 meV below the bandgap of the semiconductor. As a result, reabsorption effects are much less likely in nitrogen-doped structures compared with LEDs based on band-edge emission. This is a substantial advantage of LEDs doped with isoelectronic impurities.

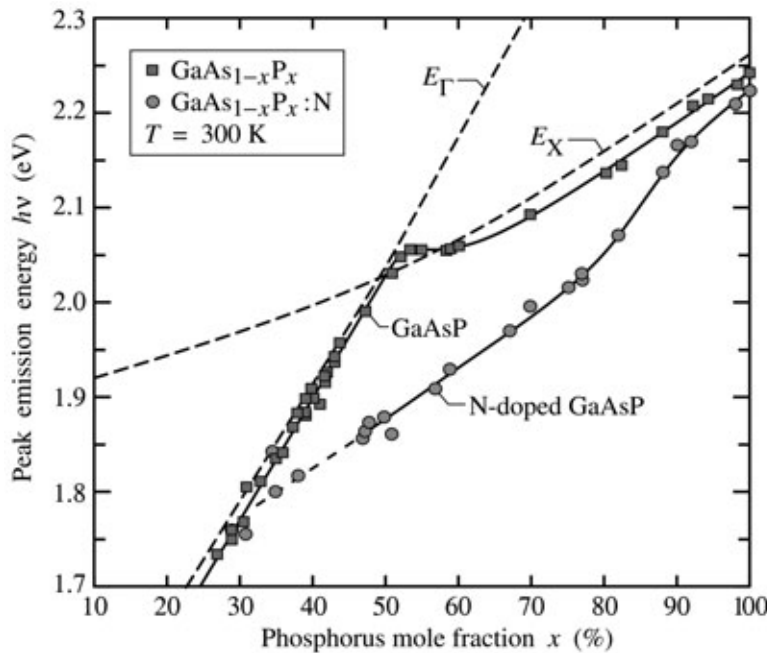


Fig. 8.2. Room-temperature peak emission energy versus alloy composition for undoped and nitrogen-doped GaAsP light-emitting diodes injected with a current density of 5 A/cm². Also shown is the energy gap of the direct- to-indirect (E_{Γ} -to- E_X) transition. The direct-indirect crossover occurs at $x \approx 50\%$ (after Craford *et al.*, 1972).

Groves *et al.* (1978a, 1978b) showed that this advantage is particularly pronounced, if only the active region is doped with nitrogen. In this case, the region of the p-n junction plane and the regions located within the carrier diffusion lengths from the junction plane are doped with

nitrogen. Other regions, such as the confinement and window layer, are not doped with the isoelectronic impurity, so that reabsorption of light by the isoelectronic impurities is limited to the narrow active region. Quantum efficiencies of several percent can be attained with GaP:N LEDs in which the nitrogen doping is limited to the active region.

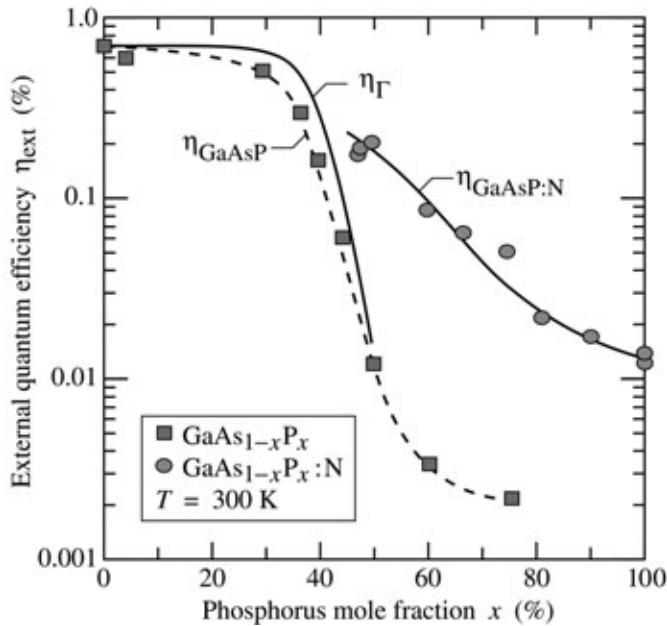


Fig. 8.3. Experimental external quantum efficiency of undoped and N-doped GaAsP versus the P mole fraction. Also shown is the calculated direct-gap (Γ) transition efficiency, η_{Γ} , and the calculated nitrogen (N) related transition efficiency, η_N (solid lines). Note that the nitrogen-related efficiency is higher than the direct-gap efficiency in the indirect bandgap ($x > 50\%$) regime (after Campbell *et al.*, 1974).

The external quantum efficiency of undoped and nitrogen-doped GaAsP is shown in Fig. 8.3. Only the vicinity of the active region is doped with nitrogen. The efficiency of the nitrogen-doped LEDs is strongly enhanced over the entire composition range compared with the GaAsP LEDs without nitrogen doping.

Note that the GaAsP LED efficiency decreases by more than two orders of magnitude in the composition range $x = 40 - 60\%$. This decrease is due to the direct-indirect crossover occurring in GaAsP and due to the increasing dislocation density occurring at higher phosphorus mole fractions. At a phosphorus mole fractions of 75 %, the GaAsP external quantum efficiency is only 0.002 %.

The external quantum efficiency of undoped and nitrogen-doped GaAsP versus emission wavelength is shown in Fig. 8.4. Again, only the vicinity of the active region is doped with nitrogen. The efficiency of nitrogen-doped GaAsP is higher than that of undoped GaAsP, in particular in the orange, yellow, and green wavelength range where the improvement is a factor of 2 – 5. In the red wavelength range, the undoped and nitrogen-doped GaAsP LEDs have similar efficiencies.

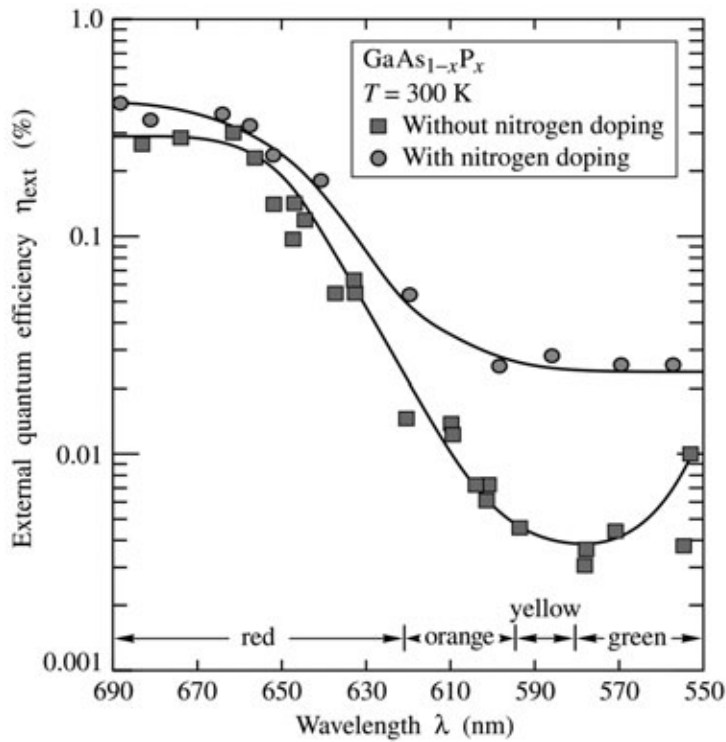


Fig. 8.4. External quantum efficiency versus emission wavelength in undoped and nitrogen-doped $\text{GaAs}_{1-x}\text{P}_x$ (after Groves *et al.*, 1978a, 1978b).

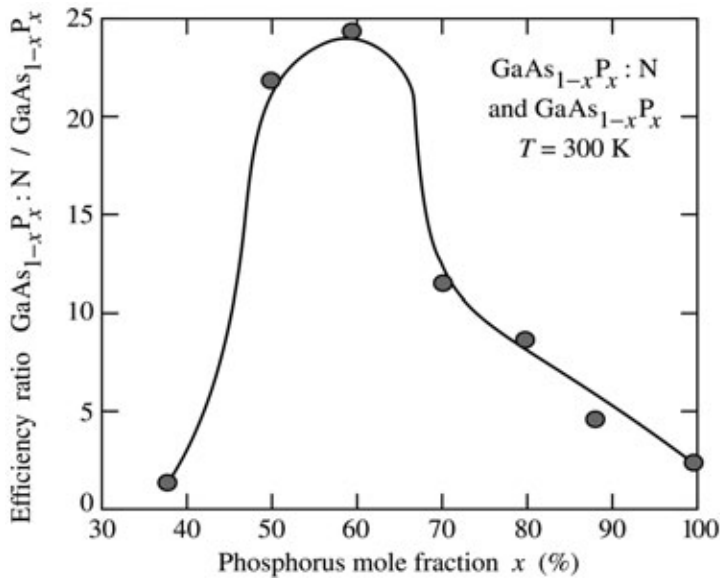


Fig. 8.5. Efficiency ratio between nitrogen-doped and undoped $\text{GaAs}_{1-x}\text{P}_x$ at 300 K (after Groves *et al.*, 1978a, 1978b).

The ratio of external quantum efficiencies of undoped and nitrogen-doped GaAsP LEDs is shown in Fig. 8.5. It is inferred from the figure that nitrogen-doped devices have a higher efficiency over the entire composition range.

The brightness of LEDs based on isoelectronic impurity transitions is limited by the *finite solubility* of nitrogen. For example, nitrogen is soluble in GaP up to nitrogen concentrations of

about 10^{20} cm^{-3} . Since an optical transition via a nitrogen level has a certain lifetime, the maximum nitrogen concentration limits the LED operation to a maximum current beyond which the LED efficiency decreases.

Commercial low-brightness green LEDs are based on nitrogen-doped GaP. The main application of GaP:N LEDs is indicator lamps. However, nitrogen-doped GaP LEDs are not suitable for high-brightness applications, i.e. for applications under bright ambient light conditions such as sunlight. High-brightness green LEDs are based on GaInN.

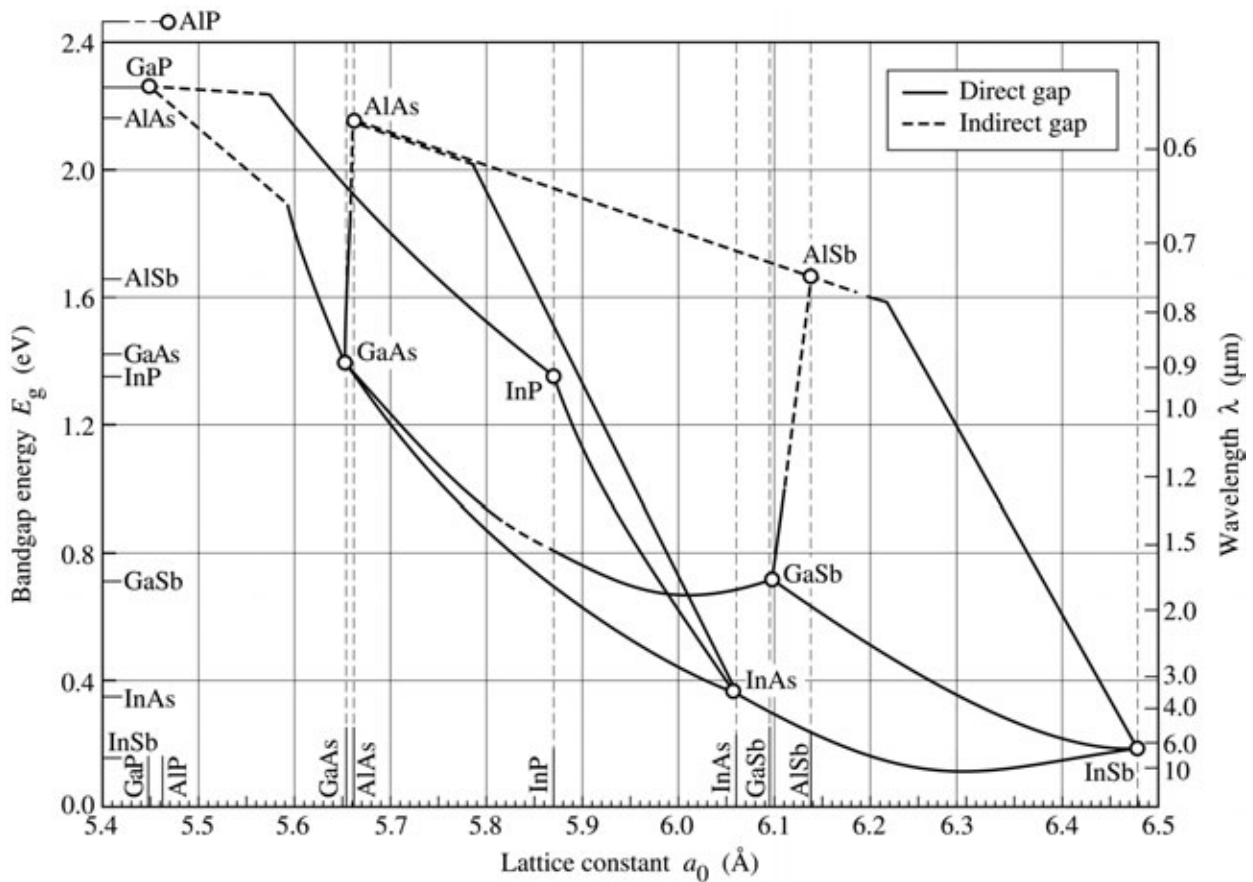


Fig. 8.6. Bandgap energy and lattice constant of various III–V semiconductors at room temperature (adopted from Tien, 1988).

8.2 The AlGaAs/GaAs material system

The $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ material system was developed in the 1970s and early 1980s and it was the first material system suitable for high-brightness LED applications (for a review, see Steranka, 1997). Owing to the very similar atomic radii of Al (1.82 Å) and Ga (1.81 Å), the material system $\text{Al}_x\text{Ga}_{1-x}\text{As}$ (or briefly AlGaAs) is lattice matched to GaAs for all Al mole fractions. The lack of dependence of the lattice constant on the Al mole fraction can be inferred from Fig. 8.6,

which shows the energy gap and lattice constant of several III–V semiconductors and of its ternary and quaternary alloys as a function of the lattice constant (adopted from Tien, 1988).

GaAs and $\text{Al}_x\text{Ga}_{1-x}\text{As}$ for Al mole fractions $x < 0.45$ are direct-gap semiconductors. The energy gap of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ versus the Al mole fraction is shown in Fig. 8.7 (Casey and Panish, 1978). For Al mole fractions $x < 45\%$, the Γ conduction-band valley is the *lowest* minimum and the semiconductor has a *direct* gap. For $x > 45\%$, the X valleys are the *lowest* conduction band minimum and the semiconductor becomes *indirect*.

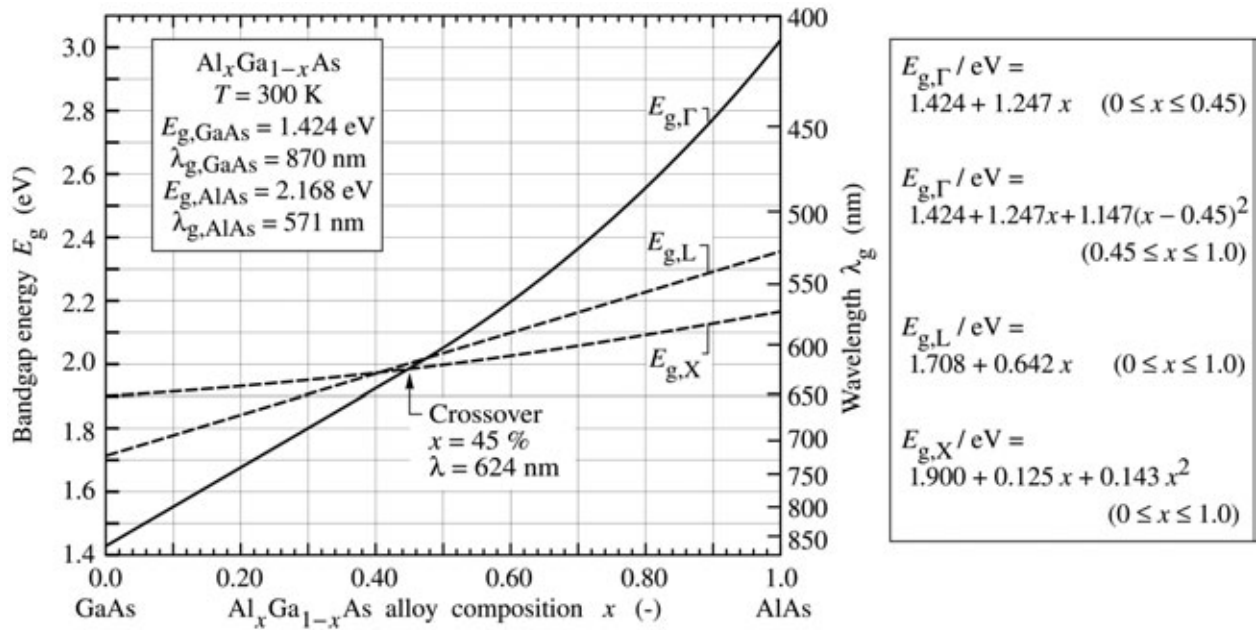


Fig. 8.7. Bandgap energy and emission wavelength of AlGaAs at room temperature. E_{Γ} denotes the direct gap at the Γ point and E_L and E_X denote the indirect gap at the L and X point of the Brillouin zone, respectively (adopted from Casey and Panish, 1978).

The AlGaAs material system is suited for high-brightness visible-spectrum LEDs emitting in the red wavelength range. The direct–indirect crossover occurs at a wavelength of 621 nm. At that wavelength, the radiative efficiency of the AlGaAs system becomes quite low due to the direct–indirect transition. To maintain high efficiency, the emission energy must be several kT lower than the bandgap energy at the direct–indirect crossover point.

There are several possible strategies for AlGaAs-based red LEDs, including $\text{Al}_x\text{Ga}_{1-x}\text{As}$ bulk active regions, $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ quantum well active regions, and $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{Al}_y\text{Ga}_{1-y}\text{As}$ ($x > y$) double heterostructure active regions. The first possibility, $\text{Al}_x\text{Ga}_{1-x}\text{As}$ bulk active regions, lacks the advantages of a heterostructure and this approach is therefore not used in high-brightness LEDs. The two other possibilities are more attractive due to the employment of heterostructures.

The quantum well and double heterostructure active region is used in high-efficiency red LEDs and the two structures are shown schematically in Fig. 8.8. In the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ quantum well case shown in Fig. 8.8 (a), size quantization is used to increase the emission energy. In the case of the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{Al}_y\text{Ga}_{1-y}\text{As}$ double heterostructures shown in Fig. 8.8 (b), AlGaAs is used for both the barrier region and the well region. A drawback of the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ quantum well active regions is the requirement of very thin GaAs quantum wells clad by $\text{Al}_x\text{Ga}_{1-x}\text{As}$ barriers. Vertical transport in multi-quantum well (MQW) structures can lead to non-uniform carrier distribution in the MQW active region unless the barriers are very thin. Consequently, the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{Al}_y\text{Ga}_{1-y}\text{As}$ double heterostructure approach is usually preferred.

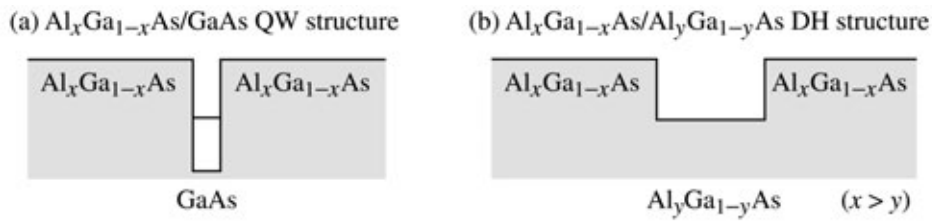


Fig. 8.8. Band diagrams of AlGaAs/GaAs structures suited for emission in the red part of the visible spectrum. (a) AlGaAs/GaAs quantum well (QW) structure with a thin GaAs well. (b) AlGaAs/AlGaAs double heterostructure (DH) with an AlGaAs active region.

AlGaAs/GaAs LEDs have been fabricated as homostructures, single heterostructures, and double heterostructures (Nishizawa *et al.*, 1983). The most efficient AlGaAs red LEDs are double-heterostructure transparent-substrate (DH-TS) devices (Ishiguro *et al.*, 1983; Steranka *et al.* 1988; Ishimatsu and Okuno, 1989). **AlGaAs DH-TS LEDs** are grown on temporary GaAs substrates and consist of a thick (e.g. 125 μm) $\text{Al}_x\text{Ga}_{1-x}\text{As}$ lower confinement layer with an Al mole fraction $x > 60\%$, an $\text{Al}_x\text{Ga}_{1-x}\text{As}$ active layer ($x = 35\%$, for red devices), and a thick (e.g. 125 μm) $\text{Al}_x\text{Ga}_{1-x}\text{As}$ upper confinement layer, also with an Al mole fraction $x > 60\%$. For devices emitting in the IR, the Al mole fractions of the active and confinement layers can be lower. After epitaxial growth, the absorbing GaAs substrate is removed by polishing and selective wet chemical etching. AlGaAs DH-TS LEDs are more than a factor of two brighter than double-heterostructure absorbing-substrate (DH-AS) devices (Steranka *et al.*, 1988).

In the 1980s the growth method of choice for AlGaAs DH-TS LEDs was liquid-phase epitaxy (LPE). This growth method is capable of growing, at a high growth rate, very thick high-quality AlGaAs layers with high Al content. LPE can be scaled up for high-volume production (Ishiguro *et al.*, 1983; Steranka *et al.*, 1988; Ishimatsu and Okuno, 1989). AlGaAs/GaAs DH-AS

LEDs have also been grown by OMVPE (Bradley *et al.*, 1986). However, the OMVPE growth rate is lower than the LPE growth rate. OMVPE growth of thick layers, as required for DH-TS devices, is therefore difficult. Historically AlGaAs DH-TS LEDs were the first high-brightness LEDs suitable for demanding applications such as automotive brake lights and traffic lights, which must be clearly visible under bright ambient conditions.

The reliability of AlGaAs devices is known to be lower than that of AlGaInP devices that do not contain any AlGaAs. High-Al-content AlGaAs layers are subject to oxidation and corrosion, thereby lowering the device lifetime. Dallesasse *et al.* (1990) reported the deterioration of AlGaAs/GaAs heterostructures by hydrolysis. Cracks, fissures, and pinholes were found after long-term exposure to room environmental conditions, especially for thick AlGaAs layers ($> 0.1 \mu\text{m}$) with a high Al content such as 85 %. The authors found very thin AlGaAs layers (e.g. 20 nm) to be stable, even for Al contents of 100 %. Hermetic packaging is required to avoid oxidation and hydrolysis of AlGaAs layers. Steranka *et al.* (1988) stated that some AlGaAs devices on the market have exhibited severe degradation. However, accelerated aging data taken at 55 °C with an injection current of 30 mA showed no degradation at all after 1000 h of stress. Such a performance requires excellent understanding and control of the device fabrication and packaging process.

8.3 The AlGaInP/GaAs material system

The AlGaInP material system was developed in the late 1980s and early 1990s and today is the primary material system for high-brightness LEDs emitting in the long-wavelength part of the visible spectrum, i.e. in the red, orange, amber, and yellow wavelength range. The AlGaInP material system and AlGaInP LEDs have been reviewed by Stringfellow and Craford (1997), Chen *et al.* (1997), and Kish and Fletcher (1997). Further reviews and recent developments were published by Mueller (1999, 2000) and Krames *et al.* (2002).

Figure 8.9 shows the energy gap and the corresponding wavelength versus the lattice constant of AlGaInP (Chen *et al.*, 1997). AlGaInP can be lattice matched to GaAs. Replacing all As atoms in the GaAs lattice by *smaller* P atoms and some of the Ga atoms in the GaAs lattice by *larger* In atoms, forms GaInP, which at the particular composition $\text{Ga}_{0.5}\text{In}_{0.5}\text{P}$, is lattice matched to GaAs. Since Al and Ga have very similar atomic radii, the material $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ is also lattice matched to GaAs.

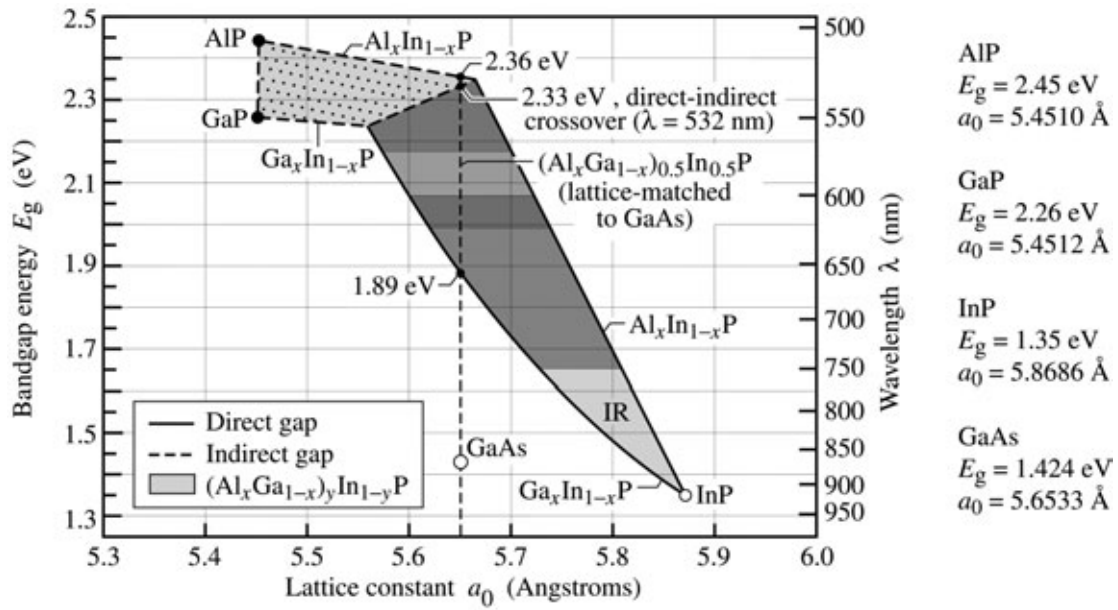


Fig. 8.9. Bandgap energy and corresponding wavelength versus lattice constant of $(Al_xGa_{1-x})_yIn_{1-y}P$ at 300 K. The dashed vertical line shows $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ lattice matched to GaAs (adopted from Chen *et al.*, 1997).

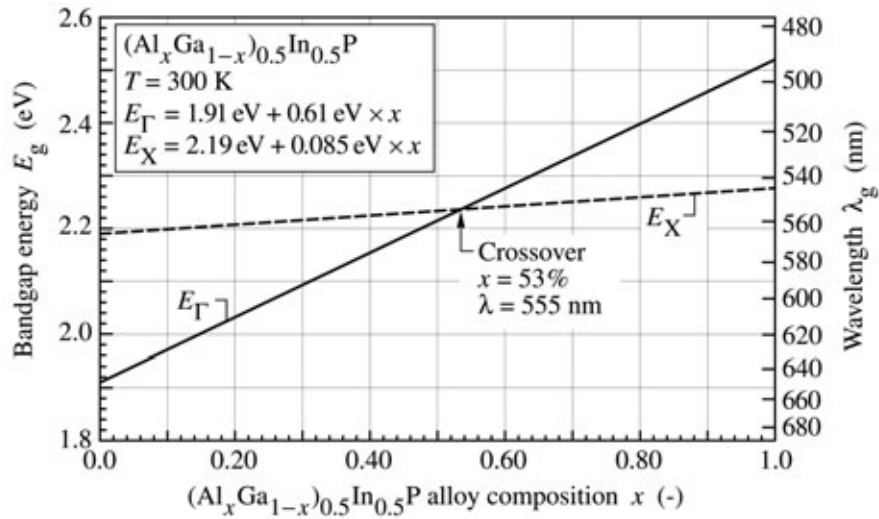


Fig. 8.10. Bandgap energy and emission wavelength of unordered AlGaInP lattice-matched to GaAs at room temperature. E_{Γ} denotes the direct gap at the Γ point and E_X denotes the indirect gap at the X point of the Brillouin zone (adopted from Prins *et al.*, 1995 and Kish and Fletcher, 1997).

According to Chen *et al.* (1997), $(Al_xGa_{1-x})_{0.5}In_{0.5}P$ has a direct bandgap for $x < 0.5$ and an indirect bandgap for $x > 0.5$. At the crossover point ($x = 0.5$), the bandgap energy is 2.33 eV, corresponding to a wavelength of 532 nm. Kish and Fletcher (1997) compiled data from Prins *et*

al. (1995) and concluded that $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}$ is a direct-gap semiconductor for Al mole fractions $x < 0.53$. The energy gap versus Al mole fraction is shown in Fig. 8.10 (Prins *et al.*, 1995; Kish and Fletcher, 1997). At Al mole fractions $x < 53\%$, the Γ conduction-band valley is the lowest minimum and the semiconductor has a direct gap. For $x > 53\%$, the X valleys are the lowest conduction band minimum and the semiconductor becomes indirect. The emission wavelength at the direct–indirect crossover point is approximately 555 nm. The exact wavelength of the crossover point may depend on the degree of atomic ordering present in a particular material (Kish and Fletcher, 1997).

A contour plot of the lattice constant and the energy gap of the AlGaInP materials system is shown in Fig. 8.11 (Chen *et al.*, 1997). The bandgap energy values and the composition of the direct–indirect crossover shown in Fig. 8.11 are slightly different from the data shown in Fig. 8.10, which can be attributed to atomic ordering in AlGaInP. Atomic ordering lowers the bandgap energy by values up to 190 meV (Kish and Fletcher, 1997).

The AlGaInP material system is suited for high-brightness visible-spectrum LEDs emitting in the red, orange, amber, and yellow wavelength range. At the direct–indirect crossover the radiative efficiency of the AlGaInP system is quite low due to the direct–indirect transition. To maintain high efficiency, the emission energy should be several kT lower than the bandgap energy at the direct–indirect crossover point.

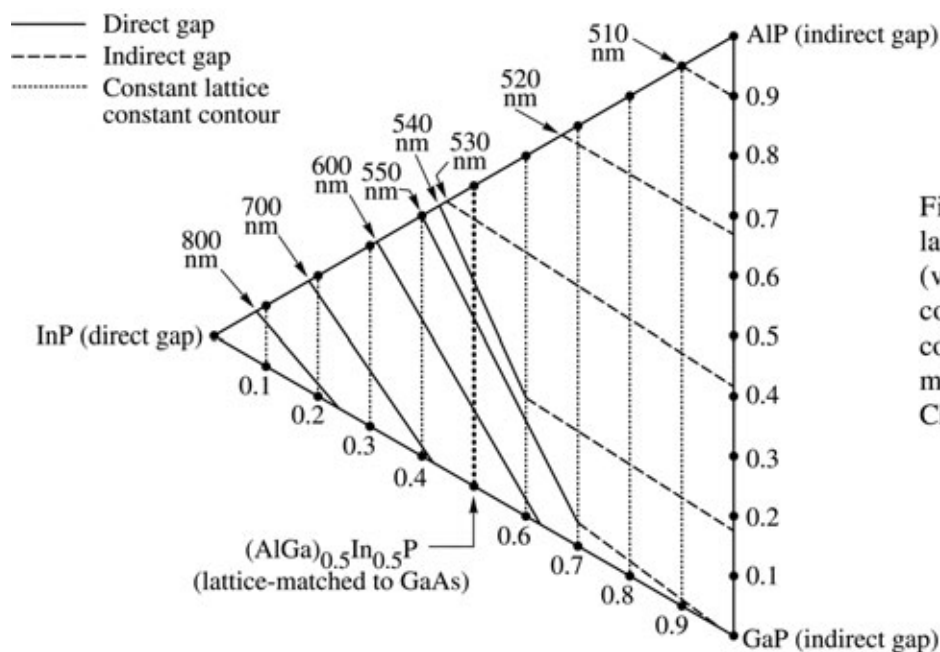


Fig. 8.11. Constant lattice constant contours (vertical lines) and constant emission line contours of the AlGaInP materials system (after Chen *et al.*, 1997).

8.4 The GaInN material system

The GaInN material system was developed in the early 1990s. GaInN LEDs emitting in the blue and green wavelength range have become commercially available in the late 1990s. To date GaInN is the primary material system for high-brightness blue and green LEDs. The GaInN material system and GaInN LEDs have been reviewed by Nakamura and Fasol (1997) and by Strite and Morkoc (1992).

One of the greatest surprises of the GaInN materials system is its high radiative efficiency despite the presence of a very high concentration of threading dislocations in GaInN/GaN epitaxial films. These threading dislocations are due to the lattice mismatch between the commonly used sapphire and SiC substrates and the GaN and GaInN epitaxial films. Typical densities of the threading dislocations are in the $10^7 - 10^9 \text{ cm}^{-2}$ range.

In the III–V arsenide and III–V phosphide material systems, misfit dislocations have disastrous consequences for the radiative efficiency. The lack of such strongly detrimental consequences in the GaInN material system is not fully understood. However, it is believed that the small diffusion length of holes and an apparently low electrical activity of dislocations in GaN and GaInN allows for high radiative efficiencies.

It has also been postulated that fluctuations of the In content in GaInN causes carriers to be localized in potential minima, thus preventing carriers from reaching dislocations. The carriers localized in potential minima will eventually recombine radiatively. Although the physical mechanisms of the high radiative efficiency of GaInN are not yet understood in detail, it is clear that the optical properties of III–V nitrides are much less affected by dislocation than those of III–V arsenides and III–V phosphides.

The bandgap energy versus the lattice constant in the nitride material family is shown in Fig. 8.12. Inspection of the figure indicates that GaInN is, in theory, suitable for covering the entire visible spectrum. However, the growth of high-quality GaInN becomes increasingly more difficult as the In composition is increased, in part due to re-evaporation of In from the growth surface. As a result, the GaInN material system is exclusively used for ultraviolet (UV), blue, and green LEDs at the present time and rarely for longer wavelengths.

Prior to the year 2002, the generally accepted value for the InN bandgap energy was 1.9 eV. However, Wu *et al.* (2002a, 2002b) showed by luminescence measurements that the bandgap of InN is lower, namely between 0.7 and 0.8 eV. Luminescence measurements also indicated that the InN bandgap exhibits an unusual blue shift with increasing temperature.

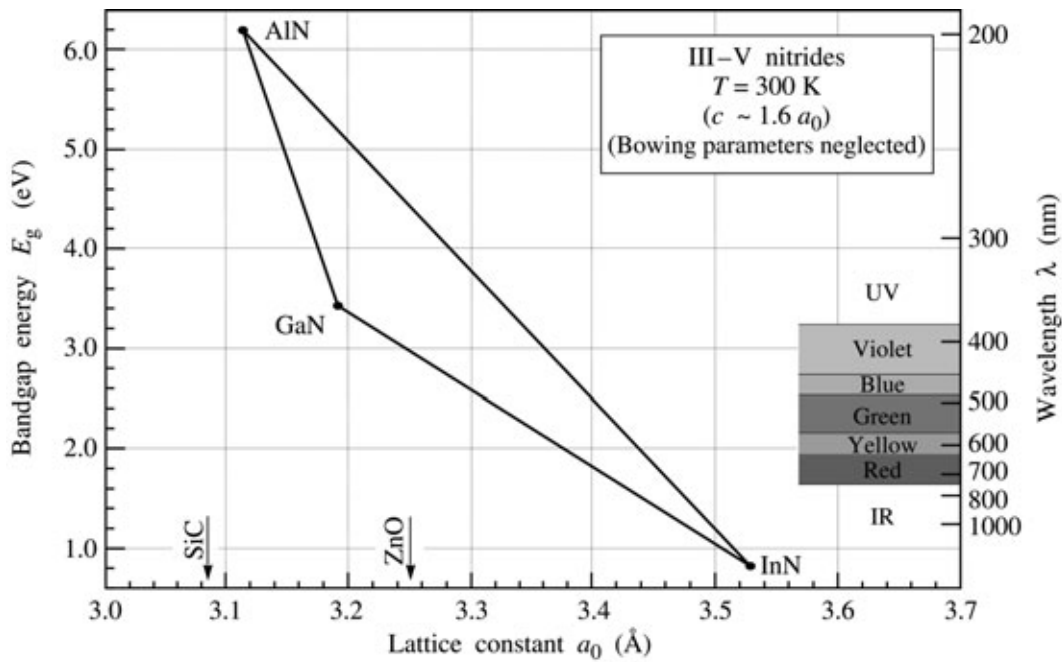


Fig. 8.12. Bandgap energy versus lattice constant of III-V nitride semiconductors at room temperature.

8.5 General characteristics of high-brightness LEDs

The improvement in luminous efficiency of visible-spectrum LEDs has been truly breathtaking. The advancement of LED efficiency can be compared to the advancement made in Si integrated circuits where the performance increase versus time has been characterized by “Moore’s law”. This “law” states that the performance of Si integrated circuits doubles approximately every 18 months.

The historical development of the luminous efficiency of visible-spectrum LEDs is shown in Fig. 8.13 (Craford, 1997, 1999). The chart illustrates the modest beginnings of visible-spectrum LED technology which started in the 1960s. If the progress from 1960 to 2000 is assumed to be continuous, then the LED luminous efficiency has doubled every 4 years. The following types of LEDs are shown in the figure.

- GaAsP LEDs grown on GaAs substrates. The GaAsP/GaAs material system is lattice mismatched so that an abundance of misfit dislocations occurs in GaAsP epitaxial films. As a result, these LEDs have a low luminous efficiency (of the order of only 0.1 lm/W). Red GaAsP LEDs are still being manufactured due to the simple epitaxial growth and low fabrication cost.

- GaP LEDs doped with radiative recombination centers. Pure GaP is an indirect semiconductor and therefore has a low radiative efficiency. However, when doped with isoelectronic impurities such as N or co-doped with Zn and O, radiative transitions in the red and green spectral range occur via these centers.
- GaAsP/GaAs LEDs doped with N emitting in the red. Again a mismatched materials system with low efficiency due to the abundance of misfit dislocations.
- AlGaAs/GaAs LEDs emitting in the red. These LEDs employ GaAs quantum well active regions.
- AlGaAs/AlGaAs double heterostructure LEDs emitting in the red using AlGaAs active regions and AlGaAs barriers.
- AlGaInP/GaAs LEDs with absorbing GaAs substrates
- AlGaInP/GaP LEDs and with wafer-bonded transparent GaP substrates.
- Also included in the chart is a result by Krames *et al.* (1999) who reported LEDs with efficiencies exceeding 100 lm/W using AlGaInP/GaP LEDs with truncated inverted-pyramid-shaped dies.
- GaInN LEDs emitting in the blue and green wavelength range.

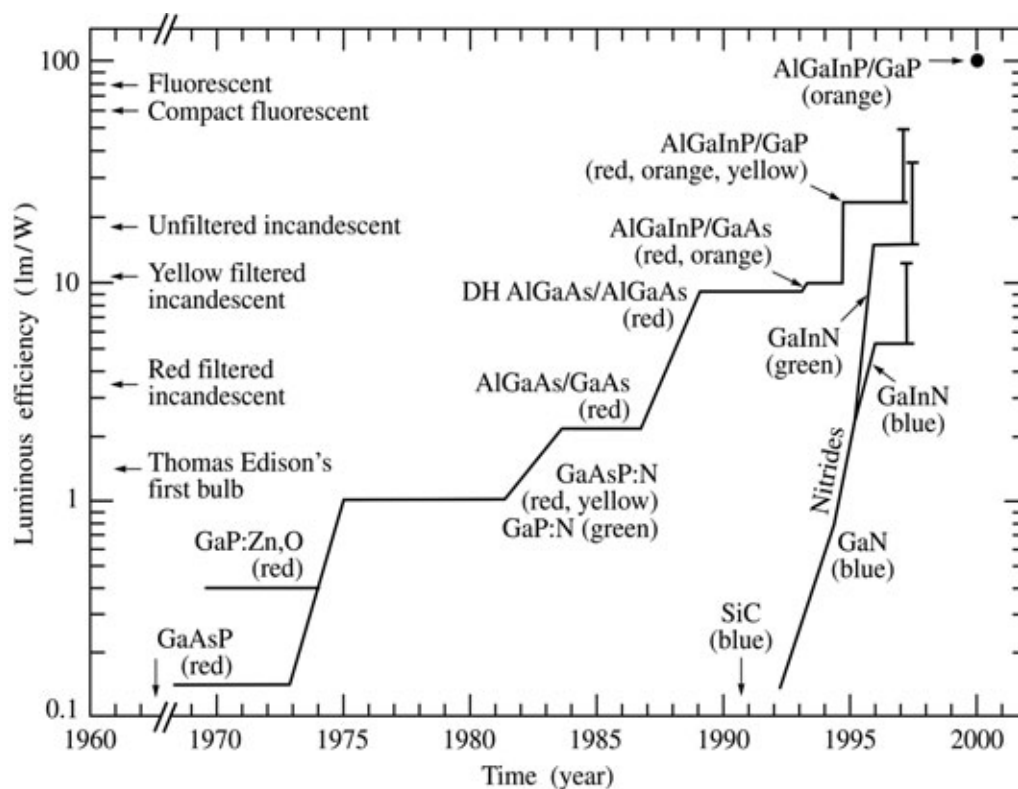


Fig. 8.13. Luminous efficiency of visible-spectrum LEDs and other light sources versus time (adopted from Craford, 1997, 1999, updated 2000).

Figure 8.13 also shows the luminous efficiency of conventional light sources including Edison's first light bulb (1.4 lm/W) and red and yellow filtered incandescent lamps. Inspection of the figure reveals that LEDs outperform filtered red and yellow incandescent lights by a large margin.

The luminous efficiency of high-brightness LEDs and of some low-cost LEDs is shown versus wavelength in Fig. 8.14 (United Epitaxy Corp., 1999). The figure indicates that yellow (590 nm) and orange (605 nm) AlGaInP and green (525 nm) GaInN LEDs are excellent choices for high luminous efficiency devices.

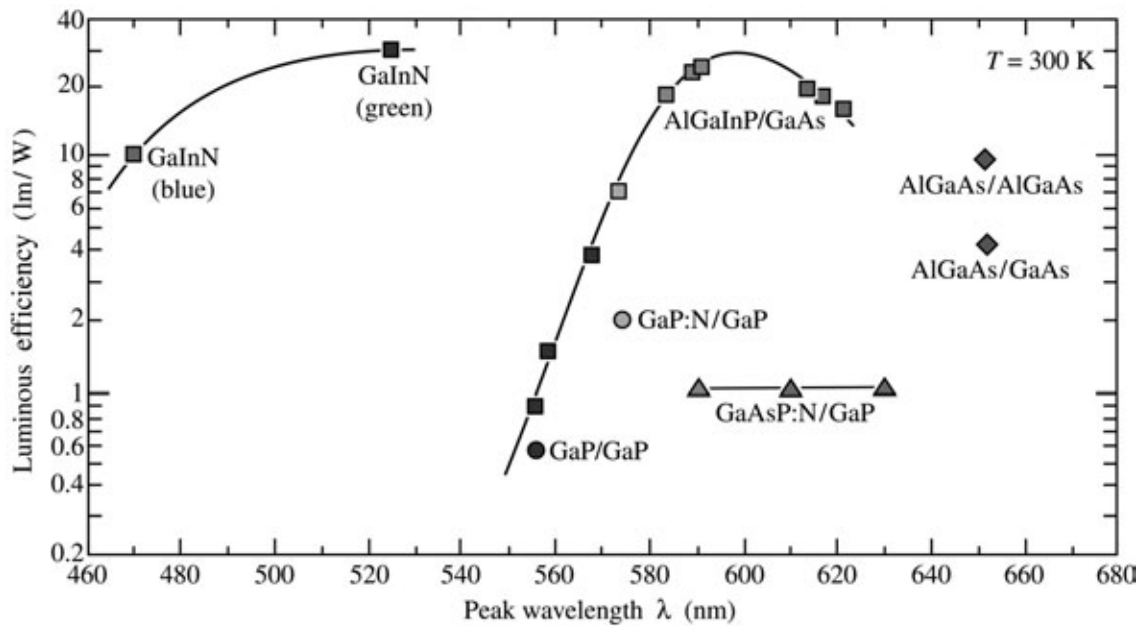


Fig. 8.14. Overview of luminous efficiency of visible LEDs made from the phosphide, arsenide, and nitride material system (adopted from United Epitaxy Corporation, 1999; updated 2000).

The fact that amber (dark yellow or orange–yellow) and orange AlGaInP LEDs provide an excellent luminous efficiency is also due to the high eye sensitivity in this wavelength range. Note that the maximum eye sensitivity occurs in the green at 555 nm, so that a green LED emitting at that wavelength appears brighter than an LED with the same optical power emitting at any other wavelength.

Owing to the high luminous efficiency and the lower manufacturing cost compared with green GaInN LEDs, amber AlGaInP LEDs are used in applications where high brightness and low power consumption are desirable, e.g. in highway signage applications. In the 1980s, such highway signs were made with power-hungry incandescent lamps powered by an electrical

generator which, in turn, was powered by a gasoline engine. Today such signs use energy-saving amber LEDs powered by solar cells (during daytime) and batteries (at night).

Figures 8.13 and 8.14 also show low-power and low-cost LEDs such as GaAsP and GaP:N LEDs with much lower luminous efficiency. These LEDs are not suitable for high-brightness applications due to their inherently lower quantum efficiency. The GaAsP LEDs are mismatched to the GaAs substrate and therefore have a low internal efficiency. The GaP:N LEDs also have a low efficiency due to the nitrogen-impurity-assisted nature of the radiative transition.

Not only the *luminous efficiency* but also the *total power* emitted by an LED is of importance for many applications, in particular for applications where a high luminous flux is required. This is the case, for example, in signage, traffic light, and illumination applications. Whereas conventional light sources such as incandescent light bulbs can be easily scaled up to provide high light power, individual LEDs are low-power devices. The historical development of the luminous flux per LED package, measured in lumens, is shown in Fig. 8.15 (Krames *et al.*, 2000). The figure shows that the luminous flux per LED package has increased by about four orders of magnitude over a period of 30 years.

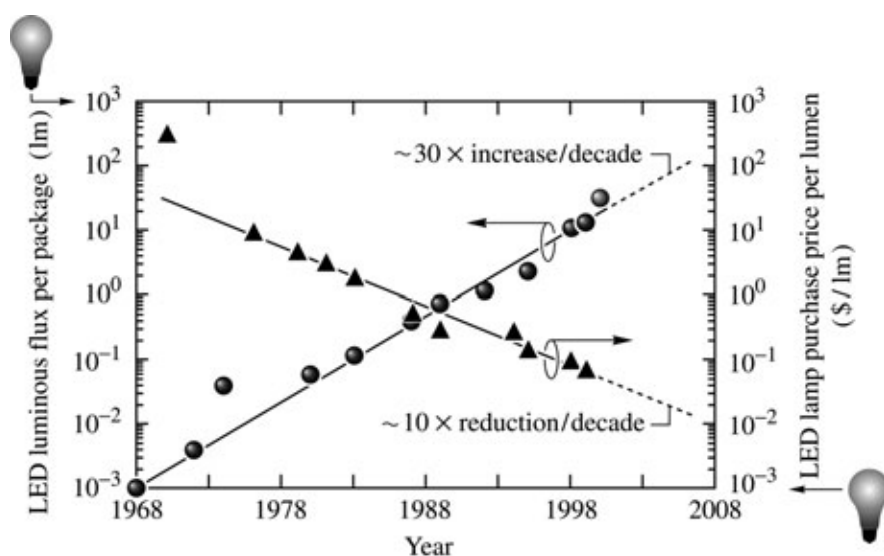


Fig. 8.15. LED luminous flux per package and LED lamp purchase price per lumen versus year. Also shown are the values for a 60 W incandescent tungsten-filament light bulb with a luminous performance of about 17 lm/W and a luminous flux of 1000 lm with an approximate price of 1.00 US\$ (after Krames *et al.*, 2000).

For comparison, Fig. 8.15 shows the luminous flux and an approximate purchase price of a 60 W incandescent light bulb with a luminous flux of 1000 lm. The figure illustrates that continued progress in the performance and manufacturing cost of LEDs is required to enable LEDs to enter into the general illumination market. Note that the cost shown in the figure is just the purchase price of the lamp and does not include the cost for the electrical power consumed over the lifetime of the lamp. The cost of the electrical power required to run an incandescent

light bulb is much higher than the purchase price of the lamp so that efficient light sources can have a cost advantage over incandescent lights even if the initial purchase price is much higher.

8.6 Optical characteristics of high-brightness LEDs

Optical emission spectra of red AlGaInP and green and blue GaInN LEDs are shown in Fig. 8.16 (Toyoda Gosei, 2000). Comparison of the emission spectra reveals that the green LED has a wider emission spectrum than either the blue or the red LED. This can be attributed to the well-known difficulties of growing GaInN with a high In content. It has been found that ***In-rich clusters*** or ***quantum dots*** form during the growth of GaInN, especially in GaInN with a high In content. It is known that the formation of such In clusters depends strongly on the growth conditions.

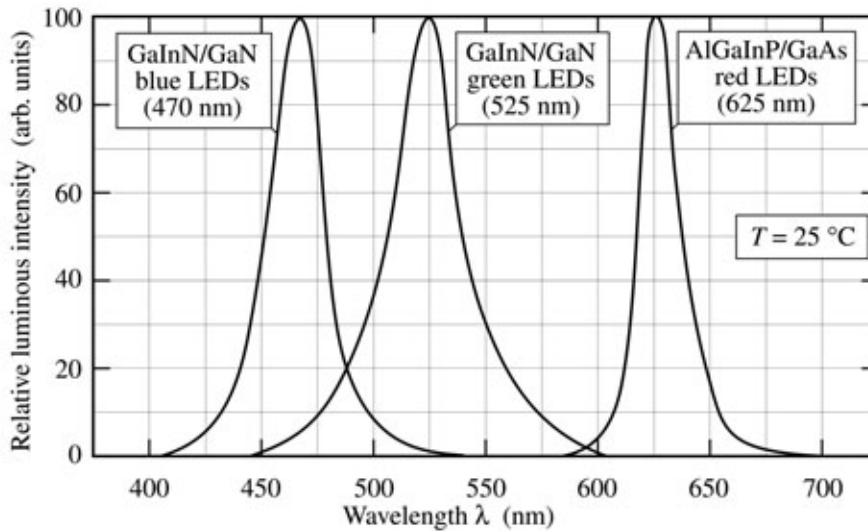


Fig. 8.16. Typical emission spectrum of GaInN/GaN blue, GaInN/GaN green, and AlGaInP/GaAs red LEDs at room temperature (after Toyoda Gosei Corp., 2000).

All LEDs shown in Fig. 8.16 have an active region comprised of a ***semiconductor alloy***. Alloy broadening, i.e. the broadening of the emission band due to random fluctuations of the chemical composition of the active material, will lead to spectral broadening that goes beyond the $1.8 kT$ linewidth expected for thermally broadened emission bands.

The light power versus injection current is shown in Fig. 8.17. A linear dependence with unit slope, i.e. unit differential quantum efficiency, is expected for light-versus-current curves in ideal LEDs. The unit-slope line is represented by the dashed line in Fig. 8.17. The mature AlGaInP LED closely follows the unit-slope line. However, the green LED has a large deviation from the

unit differential quantum efficiency slope due to the lower maturity of the GaInN material system, especially with high concentrations of In.

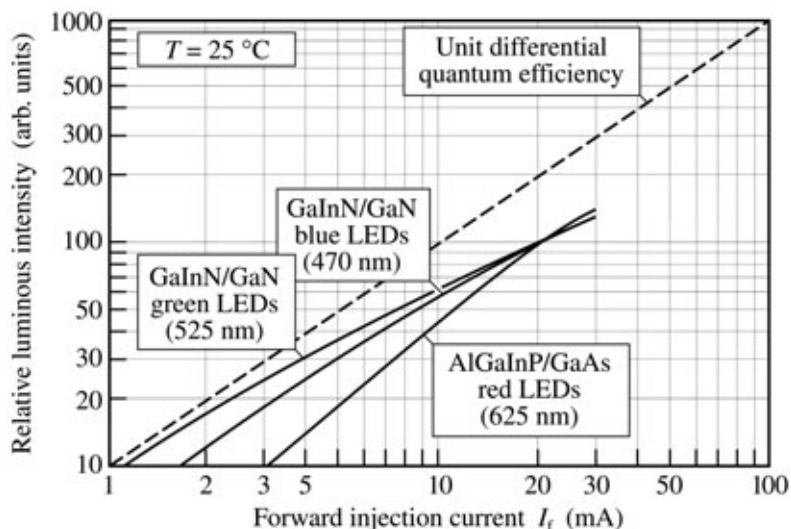


Fig. 8.17. Typical light output power vs. injection current of GaInN/GaN blue, GaInN/GaN green, and AlGaInP/GaAs red LEDs at room temperature (adopted from Toyoda Gosei Corp., 2000).

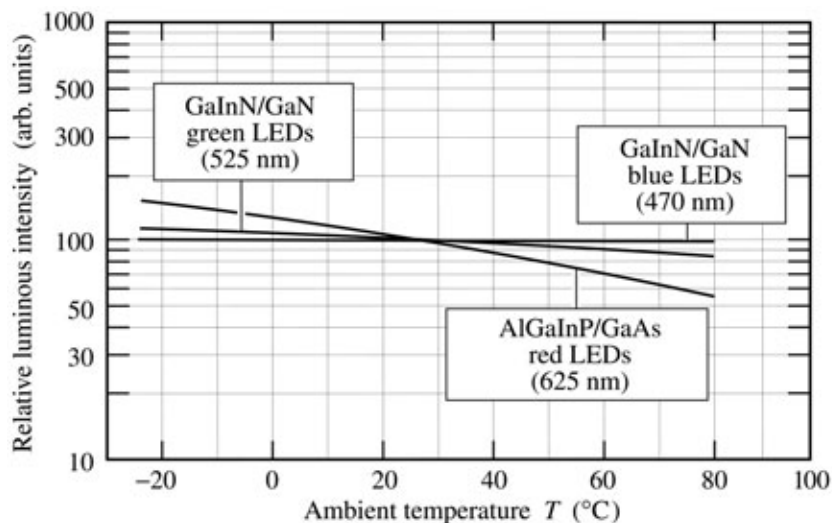


Fig. 8.18. Typical output intensity of GaInN/GaN blue, GaInN/GaN green, and AlGaInP/GaAs red LEDs versus ambient temperature (after Toyoda Gosei Corp., 2000).

The temperature dependence of the optical emission intensity is shown in Fig. 8.18. The figure reveals that III–V nitride diodes have a much weaker temperature dependence than the AlGaInP LED. Two factors contribute to the weaker temperature dependence. *First*, the active-to-confinement barriers are higher in the wide-gap III–V nitride material system than in other III–V material systems. Consequently, carriers in the GaInN active region are very well confined. Thus carrier leakage out of the well and carrier overflow are of little relevance in GaInN LEDs. *Secondly*, AlGaInP has a direct–indirect transition of the bandgap at about

555 nm. At elevated temperatures, the indirect valleys become increasingly populated so that the radiative efficiency decreases.

8.7 Electrical characteristics of high-brightness LEDs

The forward current–voltage (I – V) characteristics of a blue GaInN, a green GaInN, and a red AlGaInP LED are shown in Fig. 8.19. The forward turn-on voltage scales with the emission energy, indicating a well-behaved characteristic. Closer inspection of the forward voltage (at 1 mA) of the green LED ($V_{f,\text{green}} = 2.65$ V) indicates that it is very similar to the blue LED ($V_{f,\text{blue}} = 2.75$ V) even though the emission energy of the blue and green LED are quite different ($\lambda_{\text{blue}} = 470$ nm, $h\nu_{\text{blue}} = 2.64$ eV; $\lambda_{\text{green}} = 525$ nm, $h\nu_{\text{green}} = 2.36$ eV). The small difference in forward voltage indicates that carriers probably lose energy by phonon emission when injected from the GaN barrier into the GaInN active region. They lose more energy when being injected from the GaN barrier into the In-rich active region of the green LED. The energy, dissipated by emission of phonons, is supplied by the external voltage applied to the LED.

The diode series resistance can be inferred from the slope of the I – V characteristics. The blue and green diodes have a higher series resistance than the red AlGaInP diode. The larger resistance in GaInN LEDs can be attributed to several factors including the “lateral” resistance in the n-type buffer layer for devices grown on sapphire substrates, strong polarization effects occurring in the nitride material family, lower p-type conductivity in the cladding layer, and higher p-type contact resistance. The lower p-type conductivity is due to the high acceptor activation energy (approximately 200 meV) in GaN and GaInN so that only a small fraction of acceptors is activated.

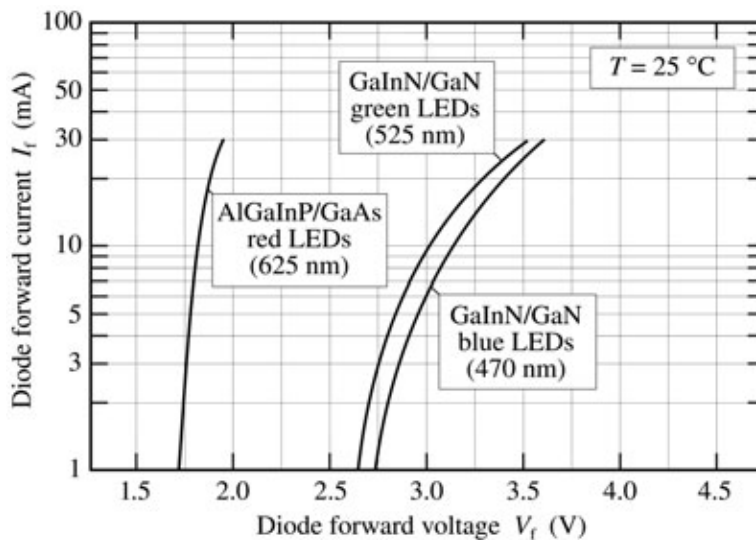


Fig. 8.19. Typical forward current–voltage (I – V) characteristic of GaInN/GaN blue, GaInN/GaN green, and AlGaInP/GaAs red LEDs at room temperature (after Toyoda Gosei Corporation, 2000).

The temperature dependence of the forward voltage at a diode current of 30 mA is shown in Fig. 8.20. For all diodes shown, the forward voltage decreases as the temperature is increased. The decrease in forward voltage is due to the decrease of the bandgap energy. In the blue and green GaInN diodes, the lower forward voltage is also due to the decrease in series resistance occurring at high temperatures. This resistance decrease is due to the higher acceptor activation occurring at elevated temperatures and the resulting higher conductivity of the p-type GaN and GaInN layers.

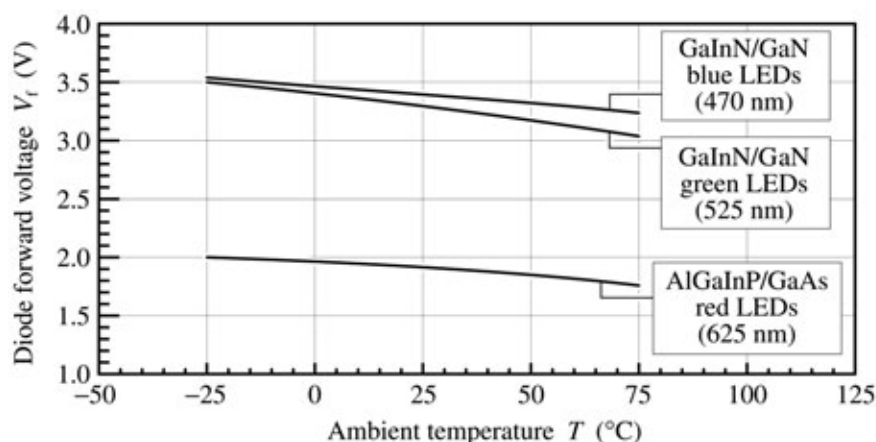


Fig. 8.20. Typical diode forward voltage at a current of 30 mA of GaInN/GaN blue, GaInN/GaN green, and AlGaInP/GaAs red LEDs versus temperature (after Toyoda Gosei Corp., 2000).

References

- Bradley R. R., Ash R. M., Forbes N. W., Griffiths R. J. M., Jebb D. P., and Shepard H. E. "Metalorganic chemical vapor deposition of junction isolated GaAlAs/GaAs LED structures" *J. Crystal Growth* **77**, 629 (1986)
- Campbell J. C., Holonyak Jr. N., Craford M. G., and Keune D. L. "Band structure enhancement and optimization of radiative recombination in GaAsP (and InGaP:N)" *J. Appl. Phys.* **45**, 4543 (1974)
- Casey H. C. Jr. and Panish M. B. "Heterostructure lasers, Part A" and "Heterostructure lasers, Part B" (Academic Press, San Diego, 1978)
- Chen C. H., Stockman S. A., Peanasky M. J., and Kuo C. P. "OMVPE growth of AlGaInP for high-efficiency visible light-emitting diodes" in "High brightness light emitting diodes" edited by G. B. Stringfellow and M. G. Craford *Semiconductors and Semimetals* Vol. **48**, (Academic Press, San Diego, 1997)
- Craford M. G. "Overview of device issues in high-brightness light-emitting diodes" in "High brightness light emitting diodes" edited by G. B. Stringfellow and M. G. Craford *Semiconductors and Semimetals* Vol. **48** (Academic Press, San Diego, 1997)

- Craford M. G. "The bright future of light-emitting diodes" Plenary talk on light emitting diodes at the MRS Fall Meeting, Boston Massachusetts December (1999)
- Craford M. G., Shaw R. W., Herzog A. H., and Groves W. O. "Radiative recombination mechanisms in GaAsP diodes with and without nitrogen doping" *J. Appl. Phys.* **43**, 4075 (1972)
- Dallesasse J. M., El-Zein N., Holonyak Jr. N., Hsieh K. C., Burnham R. D., and Dupuis R. D. "Environmental degradation of AlGaAs–GaAs quantum-well heterostructures" *J. Appl. Phys.* **68**, 2235 (1990)
- Grimmeiss H. G. and Scholz H. "Efficiency of recombination radiation in GaP" *Phys. Lett.* **8**, 233 (1964)
- Groves W. O. and Epstein A. S. "Epitaxial deposition of III–V compounds containing isoelectronic impurities" US Patent 4,001,056 (1977)
- Groves W. O., Herzog A. H., and Craford M. G. "Process for the preparation of electroluminescent III–V materials containing isoelectronic impurities" US Patent Re. 29,648 (1978a)
- Groves W. O., Herzog A. H., and Craford M. G. "GaAsP electroluminescent device doped with isoelectronic impurities" US Patent Re. 29,845 (1978b)
- Holonyak Jr. N. and Bevacqua S. F. "Coherent (visible) light emission from Ga(AsP) junctions" *Appl. Phys. Lett.* **1**, 82 (1962)
- Holonyak Jr. N., Bevacqua S. F., Bielan C. V., and Lubowski S. J. "The 'direct–indirect' transition in Ga(AsP) p–n junctions," *Appl. Phys. Lett.* **3**, 47 (1963)
- Holonyak, Jr. N., Nuese C. J., Sirkis M. D., and Stillman G. E., "Effect of donor impurities on the direct–indirect transition in Ga(AsP)" *Appl. Phys. Lett.* **8**, 83 (1966)
- Ishiguro H., Sawa K., Nagao S., Yamanaka H., and Koike S. "High efficient GaAlAs light emitting diodes of 660 nm with double heterostructure on a GaAlAs substrate" *Appl. Phys. Lett.* **43**, 1034 (1983)
- Ishimatsu S. and Okuno Y. "High efficiency GaAlAs LED" *Optoelectronics – Devices and Technologies* **4**, 21 (1989)
- Kish F. A. and Fletcher R. M. "AlGaInP light-emitting diodes" in "High brightness light emitting diodes" edited by G. B. Stringfellow and M. G. Craford *Semiconductors and Semimetals* Vol. **48** (Academic Press, San Diego, 1997)
- Krames M. R. *et al.* "High-brightness AlGaInN light emitting diodes" *Proceedings of the SPIE* **3938**, 2 (2000)

- Krames M. R. *et al.* “High-power truncated-inverted-pyramid $(\text{Al}_x\text{Ga}_{1-x})_{0.5}\text{In}_{0.5}\text{P}/\text{GaP}$ light emitting diodes exhibiting $> 50\%$ external quantum efficiency” *Appl. Phys. Lett.* **75**, 2365 (1999)
- Krames M. R., Amano H., Brown J. J., and Heremans P. L. “High-efficiency light-emitting diodes” Special Issue of *IEEE J. Select. Top. Quant. Electronics* **8**, 185 (2002)
- Logan R. A., White H. G., and Trumbore F. A. “P-n junctions in compensated solution grown GaP” *J. Appl. Phys.* **38**, 2500 (1967a)
- Logan R. A., White H. G., and Trumbore F. A. “P-n junctions in GaP with external electroluminescence efficiencies $\sim 2\%$ at 25°C ” *Appl. Phys. Lett.* **10**, 206 (1967b)
- Logan R. A., White H. G., and Wiegmann W. “Efficient green electroluminescent junctions in GaP” *Solid State Electronics* **14**, 55 (1971)
- Mueller G. “Electroluminescence I” *Semiconductors and Semimetals* Vol. **64** (Academic Press, San Diego, 1999)
- Mueller G. “Electroluminescence II” *Semiconductors and Semimetals* Vol. **65** (Academic Press, San Diego, 2000)
- Nakamura S. and Fasol G. “The blue laser diode” (Springer, Berlin, 1997)
- Nishizawa J., Koike M., and Jin C. C. “Efficiency of GaAlAs heterostructure red light emitting diodes” *J. Appl. Phys.* **54**, 2807 (1983)
- Nuese C. J., Stillman G. E., Sirkis M. D., and Holonyak, Jr. N., “Gallium arsenide-phosphide: crystal, diffusion, and laser properties” *Solid State Electron.* **9**, 735 (1966)
- Nuese C. J., Tietjen J. J., Gannon J. J., and Gossenberger H. F. “Optimization of electroluminescent efficiencies for vapor-grown GaAsP diodes” *J. Electrochem. Soc.: Solid State Science* **116**, 248 (1969)
- Pilkun M. and Rupprecht H. “Electroluminescence and lasing action in GaAsP” *J. Appl. Phys.* **36**, 684 (1965)
- Prins A. D., Sly J. L., Meney A. T., Dunstan D. J., O'Reilly E. P., Adams A. R., and Valster A. J. *Phys. Chem. Solids* **56**, 349 (1995)
- Steranka F. M. “AlGaAs red light-emitting diodes” in “High brightness light emitting diodes” edited by G. B. Stringfellow and M. G. Craford *Semiconductors and Semimetals* Vol. **48** (Academic Press, San Diego, 1997)
- Steranka F. M., DeFevre D. C., Camras M. D., Tu C.-W., McElfresh D. K., Rudaz S. L., Cook L. W., and Snyder W. L. “Red AlGaAs light emitting diodes” *Hewlett-Packard Journal* p. 84 August (1988)

- Stringfellow G. B. and Craford M. G. (Editors) “High brightness light emitting diodes” *Semiconductors and Semimetals* Vol. **48** (Academic Press, San Diego, 1997)
- Strite S. and Morkoc H., “GaN, AlN, and InN: A review” *J. Vac. Sci. Technol.* **B 10**, 1237 (1992)
- Tien P. K. Original version of the graph is courtesy of P. K. Tien of AT&T Bell Laboratories. (1988)
- Toyoda Gosei Corporation, Japan, LED product catalogue (2000)
- United Epitaxy Corporation, Taiwan, General LED and wafer product catalog (1999)
- Wolfe C. M., Nuese C. J. and Holonyak Jr. N. “Growth and dislocation structure of single-crystal Ga(AsP)” *J. Appl. Phys.* **36**, 3790 (1965)
- Wu J., Walukiewicz W., Yu K. M., Ager III J. W., Haller E. E., Lu H., Schaff W. J., Saito Y. and Nanishi Y. “Unusual properties of the fundamental bandgap of InN” *Appl. Phys. Lett.* **80**, 3967 (2002a)
- Wu J., Walukiewicz W., Yu K. M., Ager III J. W., Haller E. E., Lu H., and Schaff W. J. “Small bandgap bowing in $\text{In}_{1-x}\text{Ga}_x\text{N}$ alloys” *Appl. Phys. Lett.* **80**, 4741 (2002b)

Spontaneous emission from resonant cavities

9.1 Modification of spontaneous emission

Radiative transitions, i.e. transitions of electrons from an initial quantum state to a final state and the simultaneous emission of a light quantum, are one of the most fundamental processes in optoelectronic devices. There are two distinct ways by which the emission of a photon can occur, namely by *spontaneous* and *stimulated* emission. The two distinct emission processes were first postulated by Einstein (1917).

Stimulated emission is employed in semiconductor lasers and superluminescent LEDs. It was realized in the 1960s that the stimulated emission mode can be used in semiconductors to drastically change the radiative emission characteristics. The efforts to harness stimulated emission resulted in the first room-temperature operation of semiconductor lasers (Hayashi *et al.*, 1970) and the first demonstration of a superluminescent LED (Hall *et al.*, 1962).

Spontaneous emission implies the notion that the recombination process occurs *spontaneously*, that is without a means to influence this process. In fact, spontaneous emission has long been believed to be uncontrollable. However, research in microscopic optical resonators, where spatial dimensions are of the order of the wavelength of light, showed the possibility of controlling the spontaneous emission properties of a light-emitting medium. The changes of the emission properties include the spontaneous emission rate, spectral purity, and emission pattern. These changes can be employed to make more efficient, faster, and brighter semiconductor devices. The changes in spontaneous emission characteristics in resonant cavity (RC) and photonic bandgap (PBG) structures were reviewed by Joannopolous (1995).

Microcavity structures have been demonstrated with different active media and different microcavity structures. The first microcavity structure was proposed by Purcell (1946) for emission frequencies in the radio frequency (rf) regime. Small metallic spheres were proposed as the resonator medium. However, no experimental reports followed Purcell's theoretical publication. In the 1980s and 1990s, several microcavity structures have been realized with different types of optically active media. The emission media included organic dyes (De Martini

et al., 1987; Suzuki *et al.*, 1991), semiconductors (Yablonovitch *et al.*, 1988; Yokoyama *et al.*, 1990), rare-earth-doped silica (Schubert *et al.*, 1992b; Hunt *et al.*, 1995b), and organic polymers (Nakayama *et al.*, 1993; Dodabalapur *et al.*, 1994). In these publications, clear changes in spontaneous emission were demonstrated including changes in spectral, spatial, and temporal emission characteristics.

At the beginning of the 1990s, current-injection **resonant-cavity light-emitting diodes** (RCLEDs) were first demonstrated in the GaAs material system (Schubert *et al.*, 1992a) and subsequently in organic light-emitting materials (Nakayama *et al.*, 1993). Both publications reported an emission line narrowing due to the resonant cavities. RCLEDs have many advantageous properties when compared with conventional LEDs, including higher brightness, increased spectral purity, and higher efficiency. For example, the spectral power density in RCLEDs was shown to be enhanced by more than one order of magnitude (Hunt *et al.*, 1992, 1995a).

The changes in optical gain in VCSELs due to the enhancement in spontaneous emission was analyzed by Deppe and Lei (1992). The comparison of a *macro*cavity, in which the cavity is much *longer* than the emission wavelength ($\lambda \ll L_{\text{cav}}$), with a *micro*cavity ($\lambda \approx L_{\text{cav}}$) revealed that the gain can be enhanced by factors of 2 – 4 for typical GaAs emission linewidths at room temperature (50 nm). Thus laser threshold currents can be lower in microcavity structures due to the higher gain.

It is important to distinguish between emission *inside* the cavity and emission *out of* the cavity. The enhancement of the spontaneous emission *inside* the cavity and emission through one of the mirrors *out of* the cavity can be very different. At moderate values of the cavity finesse, the spontaneous emission inside and out of the cavity is enhanced. However, for very high finesse cavities (see, for example, Jewell *et al.*, 1988), the overall emission out of the cavity *decreases* (Schubert *et al.*, 1996). In the limit of very high reflectivity reflectors ($R_1 = R_2 \rightarrow 100\%$), the emission out of the cavity becomes zero. This effect will be discussed in detail below.

A device in which all the spontaneous emission occurs into a single optical mode has been proposed by Kobayashi *et al.* (1982, 1985). This device has been termed a **zero-threshold laser** (Yokoyama, 1992) and **single-mode LED** (Yablonovitch, 1994). In a conventional laser, only a small portion of the spontaneous emission couples into a single state of the electromagnetic field controlled by the laser cavity. The rest is lost to free space modes that radiate out the side of the laser. The idea of a thresholdless laser is simple. It assumes a wavelength size cavity in which

only one optical mode exists. Thus spontaneous as well as stimulated emission couples to this optical mode. The thresholdless laser should lack a threshold, i.e. the clear distinction between the spontaneous and the lasing regime which is observed in the light-output versus current characteristic of conventional lasers. Clearly, the prospects of such a device are intriguing. Even though several attempts to demonstrate a thresholdless laser have been reported (Yokoyama *et al.*, 1990; Yokoyama 1992; Numai *et al.*, 1993), a thresholdless laser has not yet been demonstrated.

9.2 Fabry–Perot resonators

The simplest form of optical cavity consists of two coplanar mirrors separated by a distance L_{cav} . About one century ago, Fabry and Perot (1899) were the first to build and analyze optical cavities with coplanar reflectors. These cavities had a large separation between the two reflectors, i.e. $L_{\text{cav}} \gg \lambda$. However, if the distance between the two reflectors is of the order of the wavelength of light, new physical phenomena occur, including the enhancement of the optical emission from an active material inside the cavity. Very small cavities, with typical dimensions of $L_{\text{cav}} \approx \lambda$, will be denoted as *microcavities*.

Coplanar microcavities are the simplest form of optical microcavities and their properties are summarized below. For a detailed discussion of the optical properties of Fabry–Perot cavities, the reader is referred to the literature (Coldren and Corzine, 1995; Saleh and Teich, 1991). Fabry–Perot cavities with two reflectors of reflectivity R_1 and R_2 are shown in Figs 9.1 (a) and (b). Plane waves propagating inside the cavity can interfere constructively and destructively resulting in stable (allowed) optical modes and attenuated (disallowed) optical modes, respectively. (Note that the photon length is much longer than the microcavity length.) For lossless (non-absorbing) reflectors, the transmittance through the two reflectors is given by $T_1 = 1 - R_1$ and $T_2 = 1 - R_2$. Taking into account multiple reflections inside the cavity, the transmittance through a Fabry–Perot cavity can be expressed in terms of a geometric series. The transmitted light intensity (transmittance) is then given by

$$T = \frac{T_1 T_2}{1 + R_1 R_2 - 2 \sqrt{R_1 R_2} \cos 2\phi} \quad (9.1)$$

where ϕ is the phase change of the optical wave for a single pass between the two reflectors. Phase changes at the reflectors are neglected. The maxima of the transmittance occur if the

condition of constructive interference is fulfilled, i.e. if $2\phi = 0, 2\pi, \dots$. Insertion of these values into Eq. (9.1) yields the transmittance maxima as

$$T_{\max} = \frac{T_1 T_2}{(1 - \sqrt{R_1 R_2})^2}. \quad (9.2)$$

For asymmetric cavities ($R_1 \neq R_2$), it is $T_{\max} < 1$. For symmetric cavities ($R_1 = R_2$), the transmittance maxima are unity, $T_{\max} = 1$.

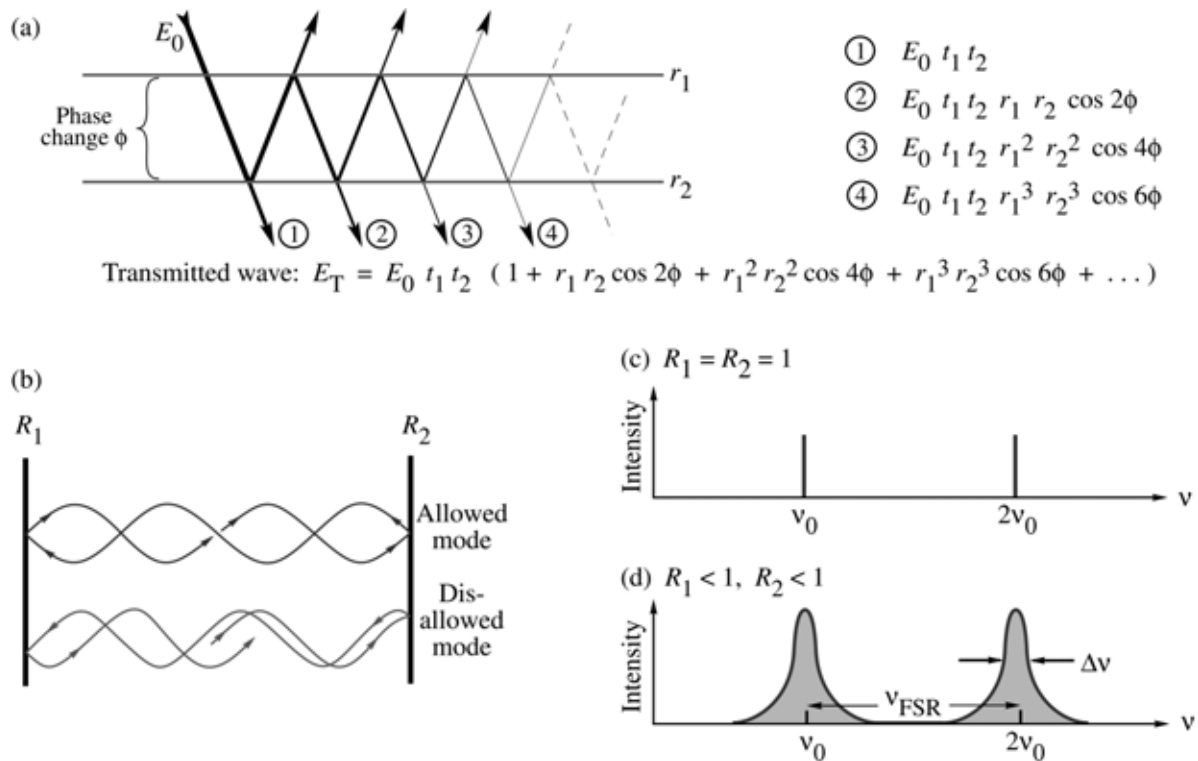


Fig. 9.1. (a) Transmission of a light wave with electric field amplitude E_0 through a Fabry–Perot resonator. (b) Schematic illustration of allowed and disallowed optical modes in a Fabry–Perot cavity consisting of two coplanar reflectors. Optical mode density for a resonator with (c) no mirror losses ($R_1 = R_2 = 100\%$) and (d) mirror losses.

Exercise. Transmission through a Fabry–Perot cavity. Derive Eq. (9.1) by calculating the transmitted wave intensity in terms of a geometric series as illustrated in Fig. 9.1 (a).

Near $\phi = 0, 2\pi, \dots$, the cosine term in Eq. (9.1) can be expanded into a power series ($\cos 2\phi \approx 1 - 2\phi^2$). One obtains

$$T = \frac{T_1 T_2}{\left(1 \pm \sqrt{R_1 R_2}\right)^2 + \sqrt{R_1 R_2} 4\phi^2} \quad (9.3)$$

Equation (9.3) indicates that near the maxima, the transmittance can be approximated by a lorentzian function. The transmittance T in Eq. (9.3) has a maximum at $\phi = 0$. The transmittance decreases to half of the maximum value at $\phi_{1/2} = [1 - (R_1 R_2)^{1/2}] / [4 (R_1 R_2)^{1/2}]^{1/2}$. For high values of R_1 and R_2 , i.e. $R_1 \approx 1$ and $R_2 \approx 1$, it is $\phi_{1/2} = (1/2) [1 - (R_1 R_2)^{1/2}]$.

The cavity finesse, F , is defined as the ratio of the transmittance peak separation to the transmittance full-width at half-maximum, i.e.

$$F = \frac{\text{Peak separation}}{\text{Peak width}} = \frac{\pi}{2 \phi_{1/2}} = \frac{\pi \sqrt[4]{R_1 R_2}}{1 - \sqrt{R_1 R_2}} \approx \frac{\pi}{1 - \sqrt{R_1 R_2}} \quad (9.4)$$

Inspection of Eq. (9.4) shows that the finesse becomes very large for high values of R_1 and R_2 .

The wavelength and frequency of light are practically more accessible than the phase. Equations (9.1) – (9.4) can be converted to wavelength and frequency using

$$\phi = 2\pi \frac{\bar{n} L_{\text{cav}}}{\lambda} = 2\pi \frac{\bar{n} L_{\text{cav}} \nu}{c} \quad (9.5)$$

where L_{cav} is the length of the cavity, λ is the wavelength of light in vacuum, ν is the frequency of light, and \bar{n} is the refraction index inside the cavity. Figures 9.1 (c) and (d) show the transmittance through a cavity with $R = 1$ and $R < 1$, respectively. In the frequency domain, the transmittance peak separation is called the *free spectral range* ν_{FSR} , as shown in Fig. 9.1 (d). The finesse of the cavity in the frequency domain is then given by $F = \nu_{\text{FSR}} / \Delta\nu$.

Frequently the *cavity quality factor* Q rather than the finesse is used. The cavity Q is defined as the ratio of the transmittance peak frequency to the peak width. Using this definition and Eq. (9.4), one obtains

$$Q = \frac{\text{peak frequency}}{\text{peak width}} = \frac{2\bar{n} L_{\text{cav}}}{\lambda} \frac{\pi \sqrt[4]{R_1 R_2}}{1 - \sqrt{R_1 R_2}} \approx \frac{2\bar{n} L_{\text{cav}}}{\lambda} \frac{\pi}{1 - \sqrt{R_1 R_2}} \quad (9.6)$$

where the peak width is measured in units of frequency.

Figure 9.2 shows an example of a reflectance spectrum of a microcavity consisting of a four-pair Si/SiO₂ distributed Bragg reflector (DBR) deposited on a Si substrate, a SiO₂ center region, a 2.5-pair Si/SiO₂ DBR top reflector. The resonance wavelength of the cavity is approximately 1.0 μm . The reflectance of the cavity does not approach zero at the resonance wavelength due to the unequalness of the reflectivities of the two reflectors.

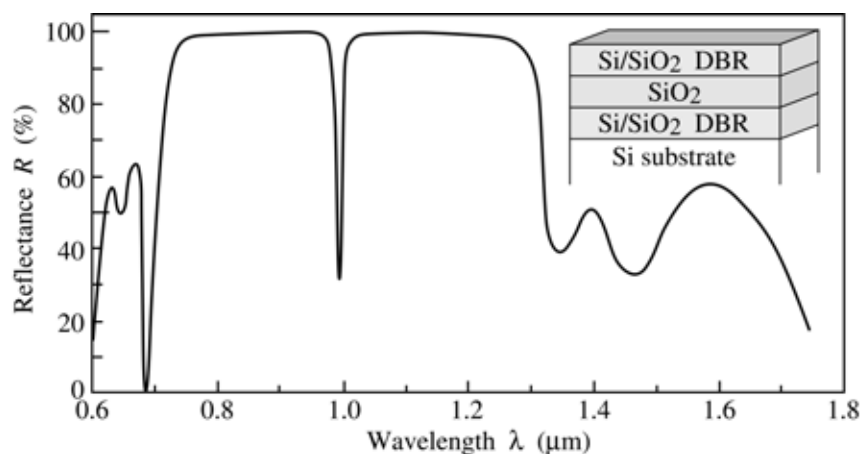


Fig. 9.2. Reflectance of a Fabry-Perot cavity consisting of two Si/SiO₂ reflectors and a SiO₂ center region. At the resonance wavelength ($\lambda \approx 1000$ nm), the reflectivity has a sharp dip.

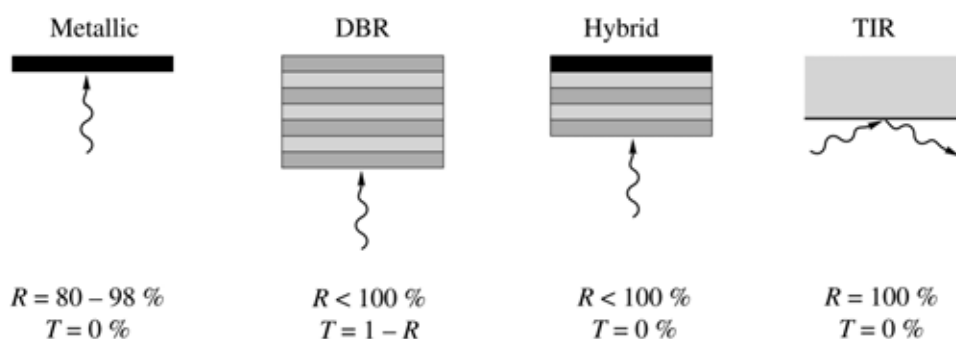


Fig. 9.3. Different types of reflectors including metallic reflector, distributed Bragg reflector (DBR), hybrid reflector, and total internal reflector (TIR). Also given are typical reflectances and transmittances.

9.3 Reflectors

Different types of reflectors are shown in Fig. 9.3 including metallic reflectors, distributed Bragg reflectors (DBRs), hybrid metal-DBR reflectors, and reflectors based on total internal reflection (TIR). Metallic reflectors and hybrid reflectors are absorbing reflectors and should not be used as light-exit reflectors. That is, the transmittance of hybrid reflectors is near zero, unless the

thickness of the metal is very thin (Tu *et al.*, 1990). Total internal reflectors require that the angle of incidence be shallow in order to achieve high reflectivity. TIR have been used in edge-emitting lasers (Smith *et al.*, 1993) and in whispering gallery lasers (McCall *et al.*, 1992).

The reflectance spectra of a metal mirror and a DBR are compared in Fig. 9.4. Whereas metal reflectors exhibit a broad band with high reflectivity, DBRs display only a narrow band of high reflectivity denoted as the *stop band*. The reflectance spectrum of the DBR displays interference fringes adjacent to the stop band. The properties of DBRs have been analyzed in detail by Coldren and Corzine (1995), Yariv (1989), and Björk *et al.* (1995). Here, only a brief summary will be given.

Consider a distributed Bragg reflector consisting of m pairs of two dielectric, lossless materials with refractive indices \bar{n}_1 and \bar{n}_2 . The thickness of the two layers is assumed to be a quarter wave, i.e. $L_1 = \lambda_{\text{Bragg}} / (4\bar{n}_1)$, and $L_2 = \lambda_{\text{Bragg}} / (4\bar{n}_2)$. The period of the DBR is $L_1 + L_2$. The reflectivity of a single interface is given by Fresnel's equation for normal incidence

$$r = \frac{\bar{n}_1 - \bar{n}_2}{\bar{n}_1 + \bar{n}_2}. \quad (9.7)$$

Multiple reflections at the interfaces of the DBR and constructive interference of the multiple reflected waves increase the reflectivity with increasing numbers of pairs. The reflectivity has a maximum at the Bragg wavelength λ_{Bragg} . The reflectivity at the Bragg wavelength of a DBR with m quarter-wave pairs is given by (Coldren and Corzine, 1995)

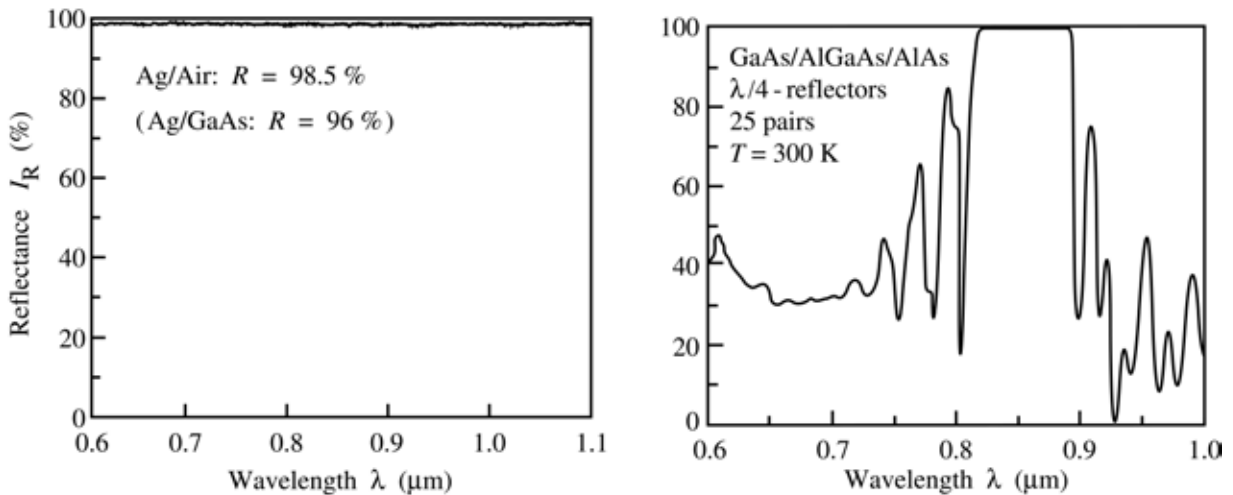


Fig. 9.4. Reflectance of a silver/air reflector and a 25-pair AlAs/GaAs distributed Bragg reflector (DBR).

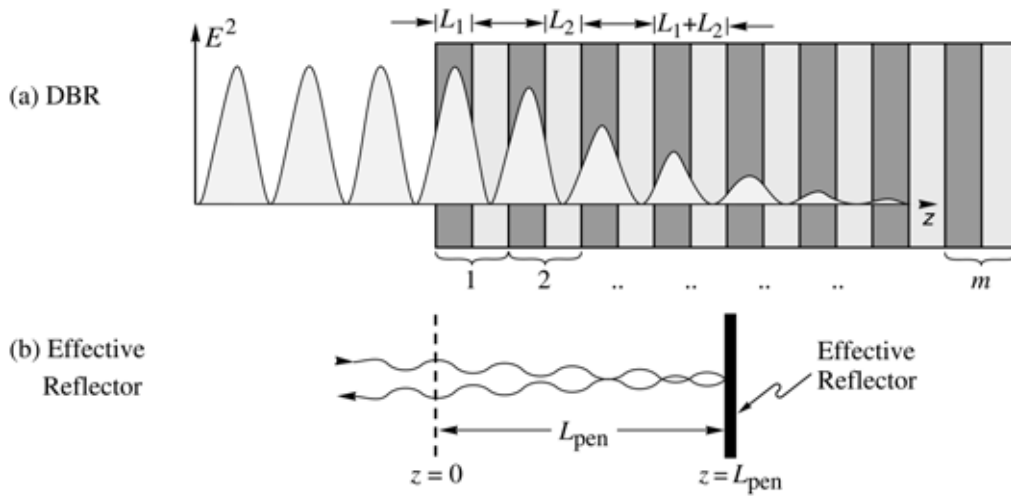


Fig. 9.5. Illustration of the DBR penetration depth. (a) DBR consisting of two materials with thickness L_1 and L_2 . (b) Ideal (metallic) reflector displaced from the DBR surface by the penetration depth.

$$R_{\text{DBR}} = |r_{\text{DBR}}|^2 = \left[\frac{1 - (\bar{n}_1/\bar{n}_2)^{2m}}{1 + (\bar{n}_1/\bar{n}_2)^{2m}} \right]^2. \quad (9.8)$$

The *high-reflectivity* or **stop band** of a DBR depends on the difference in refractive index of the two constituent materials, $\Delta\bar{n}$. The spectral width of the stop band is given by (Yariv, 1989)

$$\Delta\lambda_{\text{stop band}} = \frac{2 \lambda_{\text{Bragg}} \Delta\bar{n}}{\pi \bar{n}_{\text{eff}}} \quad (9.9)$$

where \bar{n}_{eff} is the effective refractive index of the mirror. It can be calculated by requiring the same optical path length normal to the layers for the DBR and the effective medium. The effective refractive index is then given by

$$\bar{n}_{\text{eff}} = 2 \left(\frac{1}{\bar{n}_1} + \frac{1}{\bar{n}_2} \right)^{-1}. \quad (9.10)$$

The length of a cavity consisting of two metal mirrors is the physical distance between the two mirrors. For DBRs, the optical wave penetrates into the reflector by one or several quarter-wave pairs. The penetration of the optical field is shown schematically in Fig. 9.5 (a). Only a

finite number out of the total number of quarter-wave pairs are effective in reflecting the optical wave. The effective number of pairs seen by the wave electric field is given by (Coldren and Corzine, 1995)

$$m_{\text{eff}} \approx \frac{1}{2} \frac{\bar{n}_1 + \bar{n}_2}{\bar{n}_1 - \bar{n}_2} \tanh \left(2m \frac{\bar{n}_1 - \bar{n}_2}{\bar{n}_1 + \bar{n}_2} \right). \quad (9.11)$$

For very thick DBRs ($m \rightarrow \infty$) the tanh function approaches unity and one obtains

$$m_{\text{eff}} \approx \frac{1}{2} \frac{\bar{n}_1 + \bar{n}_2}{\bar{n}_1 - \bar{n}_2}. \quad (9.12)$$

At the Bragg wavelength ($\lambda = \lambda_{\text{Bragg}}$), the phase change of the reflected wave is zero. In the vicinity of the Bragg wavelength ($\lambda \approx \lambda_{\text{Bragg}}$), the phase of the reflected wave changes *linearly* with wavelength. It is therefore possible to approximate a DBR with a metal-like mirror located a distance L_{pen} behind the first dielectric interface, as shown in Fig. 9.5. The reflection of the DBR can thus be expressed as

$$r_{\text{DBR}} \approx |r_{\text{DBR}}| e^{-2i(\beta - \beta_{\text{Bragg}})L_{\text{pen}}} \quad (9.13)$$

where $\beta = 2\pi / \lambda$ is the average phase constant of the wave. The phase change at $z = 0$ (see Fig. 9.5) of the wave reflected by the metal mirror is given by

$$r_{\text{metal}}|_{z=0} = |r_{\text{metal}}| e^{2i(2\pi/\lambda)L_{\text{pen}}}. \quad (9.14)$$

Equating the phase changes given by Eqs. (9.13) and (9.14) and using the phase changes of a DBR (Coldren and Corzine, 1995), the penetration depth is given by

$$L_{\text{pen}} = \frac{L_1 + L_2}{4r} \tanh(2mr). \quad (9.15)$$

For a large number of pairs ($m \rightarrow \infty$), the penetration depth is given by

$$L_{\text{pen}} \approx \frac{L_1 + L_2}{4r} = \frac{(L_1 + L_2)}{4} \frac{\bar{n}_1 + \bar{n}_2}{\bar{n}_1 - \bar{n}_2}. \quad (9.16)$$

Comparison of Eqs. (9.16) and (9.12) yields that

$$L_{\text{pen}} = \frac{1}{2} m_{\text{eff}} (L_1 + L_2). \quad (9.17)$$

The factor of (1/2) in Eq. (9.17) is due to the fact that m_{eff} applies to the effective number of periods seen by the *electric field*, whereas L_{pen} applies to the optical power. The optical power is equal to the square of the electric field and hence it penetrates half as far into the mirror. The effective length of a cavity consisting of two DBRs is thus given by the sum of the thickness of the center region plus the two penetration depths into the DBRs. The effective length of a cavity with DBRs is thus longer than the effective length of a cavity with metal mirrors.

9.4 Optical mode density in a one-dimensional resonator

In this section, the enhancement of spontaneous emission will be calculated based on the changes of the *optical mode density* in a one-dimensional (1D) resonator, i.e. a coplanar Fabry–Perot microcavity. We first discuss the basic physics causing the changes of the spontaneous emission from an optically active medium located inside a microcavity and then derive analytical formulas for the spectral and integrated emission enhancement. The spontaneous radiative transition rate in an optically active, homogeneous medium is given by (see, for example, Yariv, 1982)

$$W_{\text{spont}} = \tau_{\text{spont}}^{-1} = \int_0^\infty W_{\text{spont}}^{(\ell)} \rho(\mathbf{v}_\ell) d\mathbf{v}_\ell \quad (9.18)$$

where $W_{\text{spont}}^{(\ell)}$ is the spontaneous transition rate into the optical mode ℓ and $\rho(\mathbf{v}_\ell)$ is the optical mode density. Assuming that the optical medium is homogeneous, the spontaneous emission lifetime, τ_{spont} , is the inverse of the spontaneous emission rate. However, if the optical mode density in the device depends on the spatial direction, as in the case of a cavity structure, then the emission rate given in Eq. (9.18) depends on the direction. Equation (9.18) can be applied to some small range of solid angle along a certain direction, for example the direction perpendicular to the reflectors of a Fabry–Perot cavity. Thus, Eq. (9.18) can be used to calculate the emission rate along a specific direction, in particular the optical axis of a cavity.

The spontaneous emission rate into the optical mode ℓ , $W_{\text{spont}}^{(\ell)}$, contains the dipole matrix element of the two electronic states involved in the transition (Yariv, 1982). Thus $W_{\text{spont}}^{(\ell)}$ will *not* be changed by placing the optically active medium inside an optical cavity. However, the optical mode density, $\rho(\nu_\ell)$, is strongly modified by the cavity. Next, the changes in optical mode density will be used to calculate the changes in spontaneous emission rate.

We first compare the optical mode density in free space with the optical mode density in a microcavity. For simplicity, we restrict our considerations to the one-dimensional case, i.e. to the case of a coplanar Fabry–Perot microcavity. Furthermore, we restrict our considerations to the emission along the optical axis of the cavity.

In a one-dimensional homogeneous medium, the density of optical modes per unit length per unit frequency is given by

$$\boxed{\rho^{1D}(\nu) = \frac{2\bar{n}}{c}} \quad (9.19)$$

where \bar{n} is the refractive index of the medium. Equation (9.19) can be derived using a similar formalism commonly used for the derivation of the mode density in free space. The constant optical mode density given by Eq. (9.19) is shown in Fig. 9.6 (a).

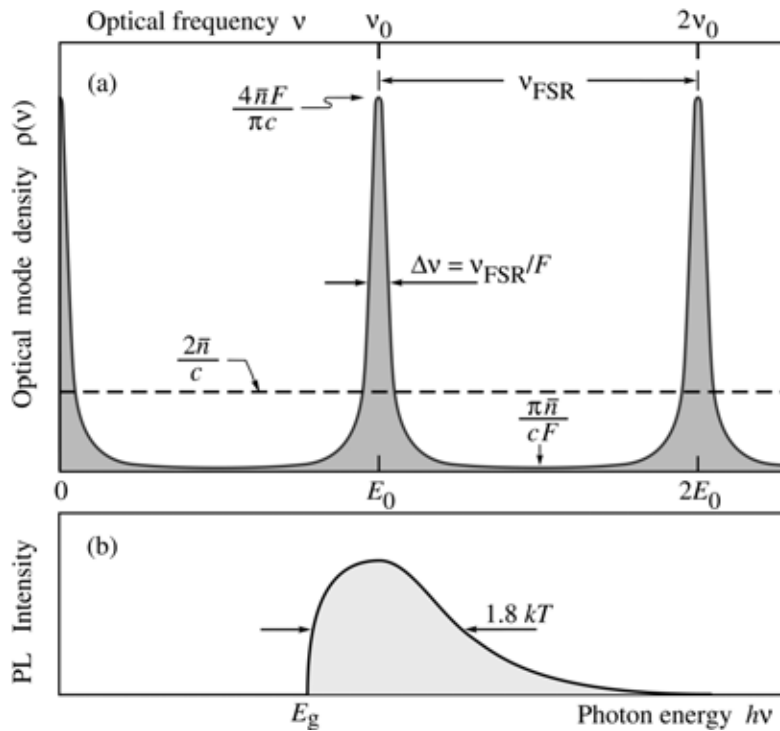


Fig. 9.6. (a) Optical mode density of a one-dimensional planar microcavity (solid line) and of homogeneous one-dimensional space. (b) Theoretical shape of the luminescence spectrum of bulk semiconductors.

Exercise. Optical mode density. Derive Eq. (9.19), i.e. optical mode density in a 1D space.

In planar microcavities, the optical modes are discrete and the frequencies of these modes are integer multiples of the fundamental mode frequency. The optical mode density of a planar microcavity is shown schematically in Fig. 9.6 (a). The fundamental and first excited mode occur at frequencies of ν_0 and $2\nu_0$, respectively. For a cavity with two metallic reflectors (no distributed Bragg reflectors) and a π phase shift of the optical wave upon reflection, the fundamental frequency is given by $\nu_0 = c / 2\bar{n}L_{\text{cav}}$, where c is the velocity of light in vacuum and L_{cav} is the length of the cavity. In a *resonant microcavity*, the emission frequency of an optically active medium located inside the cavity equals the frequency of one of the cavity modes.

The optical mode density along the cavity axis can be derived using the relation between the mode density in the cavity and the optical transmittance through the cavity, $T(\nu)$,

$$\rho(\nu) = K T(\nu) \quad (9.20)$$

where K is a constant. The value of this constant will be determined by a normalization condition, i.e. by considering a single optical mode. Using Eq. (9.1), the transmission through a Fabry–Perot cavity can be written as

$$T(\nu) = \frac{T_1 T_2}{1 + R_1 R_2 - 2 \sqrt{R_1 R_2} \cos(4\pi \bar{n} L_{\text{cav}} \nu / c)} . \quad (9.21)$$

The transmittance has maxima at $\nu = 0, \nu_0, 2\nu_0, \dots$, and minima at $\nu = \nu_0/2, 3\nu_0/2, 5\nu_0/2, \dots$. The lorentzian approximation of a transmittance maximum at $\nu = 0$ is obtained by expanding the cosine term in Eq. (9.21) into a power series using $\cos x \approx 1 - x^2/2$, so that

$$T(\nu) = \frac{T_1 T_2 \left(\sqrt{R_1 R_2}\right)^{-1} (4\pi \bar{n} L_{\text{cav}} / c)^{-2}}{\frac{(1 - \sqrt{R_1 R_2})^2}{\left(\sqrt{R_1 R_2}\right)^{-1} (4\pi \bar{n} L_{\text{cav}} / c)^2} + \nu^2} . \quad (9.22)$$

Integrating $\rho(\nu)$ over all frequencies and the cavity length yields a single optical mode, i.e.

$$K \int_0^{L_c} \int_{-\infty}^{\infty} \rho(\nu) d\nu dL = 1. \quad (9.23)$$

The lower and upper limit of the frequency integration can be chosen to be $\pm \infty$ since the lorentzian approximation of Eq. (9.22) has only one maximum at $\nu = 0$. Using Eqs. (9.20), (9.22), and the integration formula $\int_{-\infty}^{\infty} (a^2 + x^2)^{-1} dx = \pi/a$ yields

$$K = \frac{(R_1 R_2)^{3/4}}{T_1 T_2} \frac{4\bar{n}}{c} \left(1 - \sqrt{R_1 R_2}\right). \quad (9.24)$$

Using Eq. (9.20), the optical mode density of a one-dimensional cavity for emission along the cavity axis is then given by

$$\rho(\nu) = \frac{(R_1 R_2)^{3/4}}{T_1 T_2} \frac{4\bar{n}}{c} \left(1 - \sqrt{R_1 R_2}\right) T(\nu) \quad (9.25)$$

Equation (9.25) allows one to calculate the density of optical modes at the maxima and minima. At the *maxima*, the mode density is given by

$$\rho_{\max} = \frac{(R_1 R_2)^{3/4}}{1 - \sqrt{R_1 R_2}} \frac{4\bar{n}}{c}. \quad (9.26)$$

Using $(R_1 R_2)^{3/4} \approx 1$ and the expression derived for the finesse F (see Eq. 9.4), one obtains an approximate expression for the mode density at the maxima

$$\rho_{\max} \approx \frac{4}{\pi} \frac{\bar{n} F}{c}. \quad (9.27)$$

That is, the mode density at the maxima is proportional to the finesse of the cavity. At the *minima*, the mode density is given by

$$\rho_{\min} = \frac{(R_1 R_2)^{3/4} (1 - \sqrt{R_1 R_2})}{(1 + \sqrt{R_1 R_2})^2} \frac{4\bar{n}}{c}. \quad (9.28)$$

Using $(R_1 R_2)^{3/4} \approx 1$ and the expression derived for the finesse F (see Eq. 9.4), one obtains an approximate expression for the mode density at the minima

$$\rho_{\min} \approx \pi \frac{\bar{n}}{cF}. \quad (9.29)$$

That is, the mode density at the minima is *inversely* proportional to the finesse of the cavity.

The comparison of the optical mode densities of a one-dimensional (1D) free space and a 1D planar cavity is shown in Fig. 9.6. Note that the *mode density is conserved*, i.e. the area below the 1D mode density and the 1D cavity mode density are the same.

9.5 Spectral emission enhancement

Because the emission rate at a given wavelength is directly proportional to the optical mode density (see Eq. 9.18), the emission rate *enhancement spectrum* is given by the ratio of the 1D cavity mode density to the 1D free space mode density. As calculated earlier, the cavity enhancement spectrum has a lorentzian lineshape. The enhancement factor at the resonance wavelength is thus given by the ratio of the optical mode densities with and without a cavity, i.e.

$$G_e = \frac{\rho_{\max}}{\rho_{\text{1D}}} \approx \frac{2}{\pi} F \approx \left(\frac{2}{\pi}\right) \frac{\pi (R_1 R_2)^{1/4}}{1 - \sqrt{R_1 R_2}} \quad (9.30)$$

The equation shows that a strong enhancement of the spontaneous emission rate along the cavity axis can be achieved with microcavities.

Equation (9.30) represents the *average* emission rate enhancement out of *both* reflectors of the cavity. To find the enhancement out a *single* direction, we multiply the enhancement given by Eq. (9.30) by the fraction of the light exiting the mirror with reflectivity R_1 (i.e. $1 - R_1$) divided by the average loss of the two mirrors for one round trip in the cavity (i.e. $(1/2) [(1 - R_1) + (1 - R_2)]$). For large R_1 and R_2 , this gives for the enhancement of the emission exiting R_1

$$G_e \approx \frac{2(1-R_1)}{2-R_1-R_2} \frac{2F}{\pi} \approx \frac{1-R_1}{1-\sqrt{R_1 R_2}} \frac{2F}{\pi} \approx \frac{2}{\pi} \frac{\pi (R_1 R_2)^{1/4} (1-R_1)}{(1-\sqrt{R_1 R_2})^2} \quad (9.31)$$

where we used the approximation $1 - (R_1 R_2)^{1/2} \approx (1/2)(1 - R_1 R_2) \approx (1/2)(2 - R_1 - R_2)$. Equation (9.31) represents the emission rate enhancement from a *single* reflector with reflectivity R_1 .

Next we take into account the standing wave effect, that is, the distribution of the optically active material relative to the nodes and antinodes of the optical wave. The antinode enhancement factor ξ has a value of 2, if the active region is located exactly at an antinode of the standing wave inside the cavity. The value of ξ is unity, if the active region is smeared out over many periods of the standing optical wave. Finally, $\xi = 0$ if the active material is located at a node.

The emission rate enhancement is then given by

$$G_e = \frac{\xi}{2} \frac{2}{\pi} \frac{\pi (R_1 R_2)^{1/4} (1 - R_1)}{(1 - \sqrt{R_1 R_2})^2} \frac{\tau_{\text{cav}}}{\tau} \quad (9.32)$$

where R_1 is the reflectivity of the light-exit mirror and therefore $R_1 < R_2$. Equation (9.32) also takes into account changes in the spontaneous emission lifetime in terms of τ , the lifetime without cavity, and τ_{cav} , the lifetime with cavity. The factor of τ_{cav} / τ ensures that the enhancement decreases if the cavity lifetime is reduced as a result of the cavity. For planar microcavities, the ratio of the spontaneous lifetime with a cavity, τ_{cav} , and the lifetime without a cavity, τ , is $\tau_{\text{cav}} / \tau \geq 0.9$ (Vredenberg *et al.*, 1993). Thus, the emission lifetime is changed by only a minor amount in a planar microcavity.

9.6 Integrated emission enhancement

The total enhancement *integrated over wavelength*, rather than the enhancement at the resonance wavelength, is relevant for many practical devices. *On resonance*, the emission is enhanced along the axis of the cavity. However, sufficiently far *off resonance*, the emission is suppressed. Because the natural emission spectrum of the active medium (without a cavity) can be much broader than the cavity resonance, it is, *a priori*, not clear whether the integrated emission is enhanced at all. To calculate the wavelength-integrated enhancement, the spectral width of the

cavity resonance and the spectral width of the natural emission spectrum must be determined. The resonance spectral width can be calculated from the finesse of the cavity or the cavity quality factor.

The theoretical width of the emission spectrum of bulk semiconductors is $1.8 kT$ (see, for example, Schubert, 1993), where k is Boltzmann's constant and T is the absolute temperature. At room temperature, $1.8 kT$ corresponds to an emission linewidth of $\Delta\lambda_n = 31$ nm for an emission wavelength of 900 nm. For a cavity resonance width of 5 – 10 nm, one part of the spectrum is strongly enhanced, whereas the rest of the spectrum is suppressed. The integrated enhancement ratio (or suppression ratio) can be calculated analytically by assuming a Gaussian natural emission spectrum. For semiconductors at 300 K, the linewidth of the natural emission is, in the case of high-finesse cavities, *larger* than the width of the cavity resonance. The gaussian emission spectrum has a width of $\Delta\lambda_n = 2\sigma (2 \ln 2)^{1/2}$ and a peak value of $(\sigma (2\pi)^{1/2})^{-1}$, where σ is the standard deviation of the gaussian function. The integrated enhancement ratio (or suppression ratio) is then given by (Hunt *et al.*, 1993)

$$G_{\text{int}} = \frac{\pi}{2} G_e \Delta\lambda \frac{1}{\sigma \sqrt{2\pi}} = G_e \sqrt{\pi \ln 2} \frac{\Delta\lambda}{\Delta\lambda_n} \quad (9.33)$$

where the factor of $\pi / 2$ is due to the lorentzian lineshape of the enhancement spectrum. Hence, the integrated emission enhancement depends on the natural emission linewidth of the active material. The value of G_{int} can be quite different for different types of optically active materials. Narrow atomic emission spectra can be enhanced by several orders of magnitude (Schubert *et al.*, 1992b). On the other hand, materials having broad emission spectra such as dyes or polymers (de Martini *et al.*, 1987; Suzuki *et al.*, 1991) may not exhibit any integrated enhancement at all. Equation (9.33) also shows that the width of the resonance has a profound influence on the integrated enhancement. Narrow resonance spectral widths, i.e. high finesse values or long cavities (Hunt *et al.*, 1992), reduce the integrated enhancement.

Example. Spectral enhancement and integrated enhancement of a resonant-cavity structure.

As an example, we calculate the spectral and wavelength-integrated enhancement of a semiconductor resonant-cavity structure using Eqs. (9.32) and (9.33). With the reflectivities $R_1 = 90$ % and $R_2 = 97$ %, an antinode enhancement factor of $\xi = 1.5$ and $\tau_{\text{cav}} / \tau \approx 1$, one obtains a finesse of $F = 46$, and a peak enhancement factor of $G_e = 68$ using Eq. (9.32). Insertion of this

value into Eq. (9.33), using a cavity resonance bandwidth of $\Delta\lambda = 6.5$ nm (Schubert *et al.*, 1994), and the theoretical 300 K natural emission linewidth of $\Delta\lambda = 31$ nm, one obtains a theoretical integrated enhancement factor of $G_{\text{int}} = 13$. Experimental enhancement factors of 5 have been demonstrated (Schubert *et al.*, 1994) for the reflectivity values assumed above. The lower experimental enhancement is in part due to a broader natural emission linewidth, which exceeds the theoretical value of $1.8 kT$.

The spontaneous emission spectrum of a bulk semiconductor is shown schematically in Fig. 9.6 (b). For maximum enhancement along the cavity axis, the cavity must be in resonance with the natural emission spectrum. Note that additional broadening mechanisms, such as alloy broadening will broaden the natural emission spectrum over its theoretical value of $1.8 kT$. Quantum-well structures have inherently narrower spectra ($0.7 kT$), due to the step-function-like density of states. Low temperatures and excitonic effects can further narrow the natural emission linewidth. Thus, higher enhancements are expected for low temperatures and quantum well active regions.

References

- Björk G., Yamamoto Y., and Heitmann H. “Spontaneous emission control in semiconductor microcavities” in “Confined electrons and photons” edited by E. Burstein and C. Weisbuch (Plenum Press, New York, 1995)
- Coldren L. A. and Corzime S. W. “Diode lasers and photonic integrated circuits” (John Wiley and Sons, New York, 1995)
- De Martini F., Innocenti G., Jacobovitz G. R., and Mataloni P. “Anomalous spontaneous emission time in a microscopic optical cavity” *Phys. Rev. Lett.* **59**, 2955 (1987)
- Deppe D. G. and Lei C. “Spontaneous emission and optical gain in a Fabry–Perot microcavity” *Appl. Phys. Lett.* **60**, 527 (1992)
- Dodabalapur A., Rothberg L. J., and Miller T. M. “Color variation with electroluminescent organic semiconductors in multimode resonant cavities” *Appl. Phys. Lett.* **65**, 2308 (1994)
- Einstein A. “On the quantum theory of radiation (translated from German)” *Z. Phys.* **18**, 121 (1917)
- Fabry G. and Perot A. “Theory and applications of a new interference method for spectroscopy” (translated from French)” *Ann. Chim. Phys.* **16**, 115 (1899)
- Hall R. N., Fenner G. E., Kingsley J. D., Soltys T. J., and Carlson R. O. “Coherent light emission

- from GaAs junctions” *Phys. Rev. Lett.* **9**, 366 (1962)
- Hayashi I., Panish M. B., Foy P. W., and Sumski S. “Junction lasers which operate continuously at room temperature” *Appl. Phys. Lett.* **17**, 109 (1970)
- Hunt N. E. J., Schubert E. F., Logan R. A., and Zydzik G. J. “Enhanced spectral power density and reduced linewidth at 1.3 μm in an InGaAsP quantum well resonant cavity light emitting diode” *Appl. Phys. Lett.* **61**, 2287 (1992)
- Hunt N. E. J., Schubert E. F., Kopf R. F., Sivco D. L., Cho A. Y., and Zydzik G. J. “Increased fiber communications bandwidth from a resonant cavity light emitting diode emitting at $\lambda = 940 \text{ nm}$ ” *Appl. Phys. Lett.* **63**, 2600 (1993)
- Hunt N. E. J., Schubert E. F., Sivco D. L., Cho A. Y., Kopf R. F., Logan R. A., and Zydzik G. J. “High efficiency, narrow spectrum resonant cavity light-emitting diodes” in “Confined electrons and photons” edited by E. Burstein and C. Weisbuch (Plenum Press, New York, 1995a)
- Hunt N. E. J., Vredenberg A. M., Schubert E. F., Becker P. C., Jacobson D. C., Poate J. M., and Zydzik G. J. “Spontaneous emission control of Er^{3+} in Si/SiO₂ microcavities” in “Confined electrons and photons” edited by E. Burstein and C. Weisbuch (Plenum Press, New York, 1995b)
- Jewell J. L., Lee Y. H., McCall S. L., Harbison J. P., and Florez L. T. “High-finesse AlGaAs interference filters grown by molecular beam epitaxy” *Appl. Phys. Lett.* **53**, 640 (1988)
- Joannopolous J. D., Meade R. D., and Winn J. N. “Photonic crystals” (Princeton University Press, Princeton, 1995)
- Kobayashi T., Segawa T., Morimoto A., and Sueta T. *Meeting of the Jpn. Soc. of Appl. Phys.* Tokyo (1982); see also Yokoyama (1992)
- Kobayashi T., Morimoto A., and Sueta T. *Melting of the Jpn. Soc. of Appl. Phys.* Tokyo (1985); see also Yokoyama (1992)
- Nakayama T., Itoh Y., and Kakuta A. “Organic photo- and electroluminescent devices with double mirrors” *Appl. Phys. Lett.* **63**, 594 (1993)
- Numai T., Kosaka H., Ogura I., Kurihara K., Sugimoto M., and Kasahara K. “Indistinct threshold laser operation in a pnpn vertical to surface transmission electrophotonic device with a vertical cavity” *IEEE J. Quantum Electronics* **29**, 403 (1993)
- McCall S. L., Levi A. F. J., Slusher R. E., Pearton S. J., and Logan R. A. “Whispering-gallery mode microdisk lasers” *Appl. Phys. Lett.* **60**, 289 (1992)
- Purcell E. M. “Spontaneous emission probabilities at radio frequencies” *Phys. Rev.* **69**, 681

(1946)

- Saleh B. E. A. and Teich M. C. “Fundamentals of photonics” (John Wiley and Sons, New York, 1991)
- Schubert E. F. “Doping in III–V semiconductors” page 512 (Cambridge University Press, Cambridge U.K., 1993)
- Schubert E. F., Wang Y. H., Cho A. Y., Tu L. W., and Zydzik G. J. “Resonant cavity light emitting diode” *Appl. Phys. Lett.* **60**, 921 (1992a)
- Schubert E. F., Vredenberg A. M., Hunt N. E. J., Wong Y. H., Becker P. C., Poate J. M., Jacobson D. C., Feldman L. C., and Zydzik G. J. “Giant enhancement in luminescence intensity in Er-doped Si/SiO₂ resonant cavities” *Appl. Phys. Lett.* **61**, 1381 (1992b)
- Schubert E. F., Hunt N. E. J., Micovic M., Malik R. J., Sivco D. L., Cho A. Y., and Zydzik G. J. *Science* **265**, 943 (1994)
- Schubert E. F., Hunt N. E. J., Malik R. J., Micovic M., and Miller D. L. “Temperature and modulation characteristics of resonant cavity light-emitting diodes” *IEEE J. Lightwave Technol.* **14**, 1721 (1996)
- Smith G. M., Forbes D. V., Coleman J. J., and Verdeyen J. T. “Optical properties of reactive ion etched corner reflector strained-layer InGaAs–GaAs–AlGaAs quantum-well lasers” *IEEE Photonics Technology Lett.* **5**, 873 (1993)
- Suzuki M., Yokoyama H., Brorson S. D., and Ippen E. P. “Observation of spontaneous emission lifetime change of dye-containing Langmuir–Blodgett films in optical microcavities” *Appl. Phys. Lett.* **58**, 998 (1991)
- Tu L. W., Schubert E. F., Kopf R. F., Zydzik G. J., Hong M., Chu S. N. G., and Mannaerts J. P. “Vertical cavity surface emitting lasers with semitransparent metallic mirrors and high quantum efficiencies” *Appl. Phys. Lett.* **57**, 2045 (1990)
- Vredenberg A. M., Hunt N. E. J., Schubert E. F., Jacobson D. C., Poate J. M., and Zydzik G. J. *Phys. Rev. Lett.* **71**, 517 (1993)
- Yablonovitch E., Gmitter T. J., and Bhat R. “Inhibited and enhanced spontaneous emission from optically thin AlGaAs/GaAs double heterostructures” *Phys. Rev. Lett.* **61**, 2546 (1988)
- Yablonovitch E. personal communication (1994)
- Yariv A. “Theory and applications of quantum mechanics” (John Wiley and Sons, New York, 1982) p. 143
- Yariv A. “Quantum electronics” 3rd edition (John Wiley and Sons, New York, 1989)
- Yokoyama H. “Physics and device applications of optical microcavities” *Science* **256**, 66 (1992)

Yokoyama H., Nishi K., Anan T., Yamada H., Boorson S.D., and Ippen E. P. “Enhanced spontaneous emission from GaAs quantum wells in monolithic microcavities” *Appl. Phys. Lett.* **57**, 2814 (1990)

10

Resonant cavity light-emitting diodes

10.1 Introduction and history

The resonant-cavity light-emitting diode (RCLED) is a light-emitting diode that has a light-emitting region inside an optical cavity. The optical cavity has a thickness of typically one-half or one times the wavelength of the light emitted by the LED, i.e. a fraction of a micrometer for devices emitting in the visible or in the infrared. The resonance wavelength of the cavity coincides or is in resonance with the emission wavelength of the light-emitting active region of the LED. Thus the cavity is a *resonant cavity*. The spontaneous emission properties from a light-emitting region located inside the resonant cavity are enhanced by the resonant-cavity effect. The RCLED is the first practical device making use of spontaneous emission enhancement occurring in microcavities.

The placement of an active region inside a resonant cavity results in multiple improvements of the device characteristics. *First*, the light intensity emitted from the RCLED along the axis of the cavity, i.e. normal to the semiconductor surface, is higher compared with conventional LEDs. The enhancement factor is typically a factor of 2 – 10. *Secondly*, the emission spectrum of the RCLED has a *higher spectral purity* compared with conventional LEDs. In conventional LEDs, the spectral emission linewidth is determined by the thermal energy kT . However, in RCLEDs, the emission linewidth is determined by the quality factor (Q factor) of the optical cavity. As a result, the spectral emission width of the RCLED is a factor of 2 – 5 narrower compared with conventional LEDs. For the same reason, the wavelength shift with temperature is determined by the temperature coefficient of the optical cavity and not by the energy gap of the active material. This results in a significantly higher temperature stability of the RCLED emission wavelength compared with conventional LEDs. *Thirdly*, the emission far-field pattern of the RCLED is more *directed* compared with conventional LEDs. In conventional LEDs, the emission pattern is lambertian (i.e. cosine-function-like). In an RCLED, the emission pattern is directed mostly along the optical axis of the cavity.

These characteristics of RCLEDs are desirable for local-area, medium bit rate optical communication systems. LEDs play an important role in local-area (< 5 km) medium bit rate (< 1 Gbit/s) optical communication networks. In particular, plastic optical fibers are increasingly used for optical communication over short distances. The higher emission intensity and the more directed emission pattern afforded by the RCLED increase the power coupled into the optical fiber. As a result, the RCLED can transmit data over longer distances. Furthermore, the higher spectral purity of RCLEDs results in less chromatic dispersion allowing for higher bit rates.

Light-emitting diodes are the transmitter device of choice for medium bit rate optical communication over distances less than 5 km. Compared with lasers, LEDs are less expensive, more reliable, and less temperature sensitive. The RCLED has improved characteristics compared with conventional LEDs while maintaining the inherent advantages of LEDs. The reflectivity of the RCLED reflectors is lower compared with vertical-cavity surface-emitting lasers (VCSELs), thereby allowing for a lower RCLED manufacturing cost compared with VCSELs. At 650 nm, the preferred communication wavelength for plastic optical fibers, VCSELs are difficult to manufacture due to the lack of high-reflectivity reflectors.

RCLEDs are also used for high-brightness applications (Streubel *et al.*, 1998; Wirth *et al.*, 2001, 2002). In these devices, the resonance wavelength is designed to be at the long-wavelength end of the spontaneous emission spectrum of the semiconductor. This ensures that the emission intensity, integrated over all spatial directions, is maximized.

The enhanced spontaneous emission occurring in resonant-cavity structures can be beneficially employed in semiconductor and polymer LEDs. Resonant-cavity light-emitting diodes were first realized in 1992 (Schubert *et al.*, 1992a) in the GaAs material system. About a year later, RCLEDs were demonstrated in organic materials (Nakayama *et al.*, 1993).

Resonant-cavity structures with enhanced spontaneous emission also include Er-doped microcavities (Schubert *et al.*, 1992b). Owing to the inherently narrow luminescence line of intra-atomic Er radiative transitions, there is a very good overlap between the cavity optical mode and the Er luminescence line. At the present time, no Er-doped current-injection devices exist. However, the great potential of Er-doped resonant cavities makes the realization of Er-doped RCLEDs likely in the future.

10.2 RCLED design rules

The basic structure of an RCLED is shown in Fig. 10.1 and comprises of two mirrors with reflectivity R_1 and R_2 . The reflectivity of the two mirrors is chosen to be unequal so that the light

exits the cavity predominantly through one of the mirrors. This mirror is called the **light-exit mirror**. Here we designate the mirror with reflectivity R_1 as the light-exit mirror. An active region is located between the mirrors, preferably at the antinode location of the standing optical wave of the cavity, as shown in Fig. 10.1. Metal mirrors are assumed in Fig. 10.1 so that the wave amplitude is zero at the location of the mirrors.

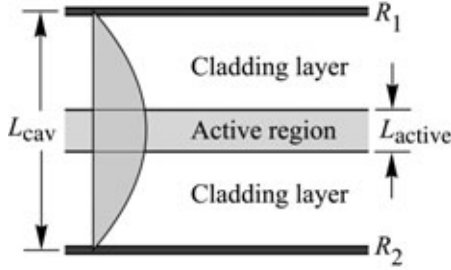


Fig. 10.1. Schematic illustration of a resonant cavity consisting of two metal mirrors with reflectivity R_1 and R_2 . The active region has a thickness L_{active} and an absorption coefficient α . Also shown is the standing optical wave. The cavity length is L_{cav} is equal to $\lambda / 2$.

Next, we summarize several design rules intended to maximize the spontaneous emission enhancement in resonant-cavity structures (Schubert *et al.*, 1994, 1996; Hunt *et al.*, 1995a, 1995b). These rules will provide further insight into the fundamental operating principles of RCLEDs and the differences of these devices with respect to VCSELs.

The **first design criterion** for RCLEDs is that the reflectivity of the light-exit reflector, R_1 , should be much lower than the reflectivity of the back reflector, i.e.

$$R_1 \ll R_2 . \quad (10.1)$$

This condition ensures that light exits the device mainly through the reflector with reflectivity R_1 . Equation (10.1) applies to the design of communication RCLEDs, where light is emitted into the small core of a multimode fiber, and display RCLEDs, where light should be emitted towards the observer.

The **second design criterion** calls for the shortest possible cavity length L_{cav} . In order to derive this criterion, the integrated enhancement, discussed in a preceding section, can be rewritten by using the expressions for the cavity finesse F and cavity quality factor Q . One obtains

$$G_{\text{int}} = \frac{\xi}{2} \frac{2}{\pi} \frac{1 - R_1}{1 - \sqrt{R_1 R_2}} \sqrt{\pi \ln 2} \frac{\lambda}{\Delta \lambda_n} \frac{\lambda_{\text{cav}}}{L_{\text{cav}}} \frac{\tau_{\text{cav}}}{\tau} \quad (10.2)$$

where λ and λ_{cav} are the active region emission wavelengths in vacuum and inside the cavity, respectively. Since the emission wavelength λ and the natural linewidth of the active medium, $\Delta\lambda_n$, are given quantities, Eq. (10.2) shows that minimization of the cavity length L_{cav} maximizes the integrated intensity.

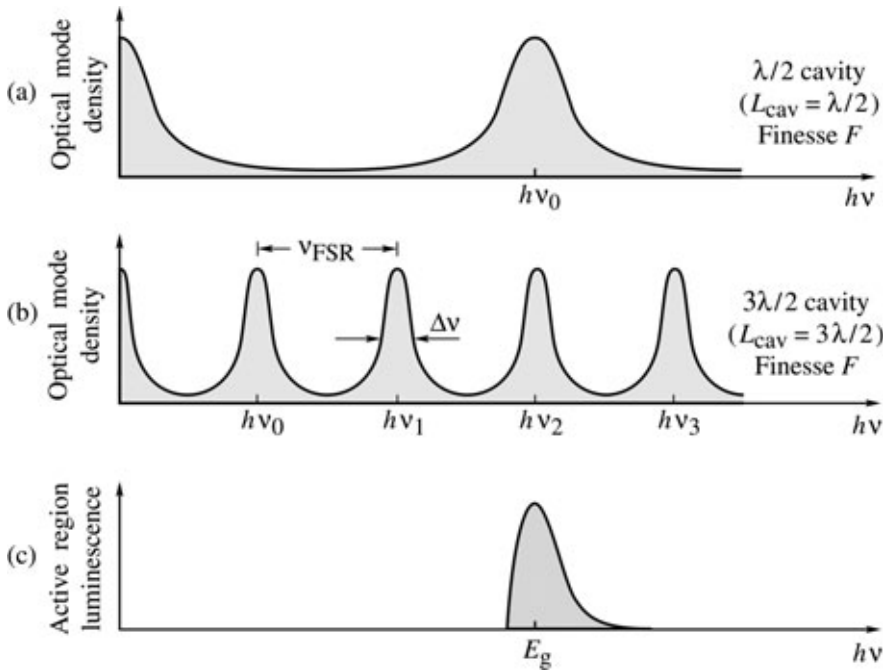


Fig. 10.2. Optical mode density for (a) a short and (b) a long cavity with the same finesse F . (c) Spontaneous free space emission spectrum of an LED active region. The spontaneous emission spectrum has a better overlap with the short-cavity mode spectrum compared with the long-cavity mode spectrum.

The importance of a short cavity length is elucidated by Fig. 10.2. The optical mode density of two different cavities, namely a short and a long cavity are shown in Figs. 10.2 (a) and (b), respectively. Both cavities have the same mirror reflectivities and finesse. The natural emission spectrum of the active region is shown in Fig. 10.2 (c). The best overlap between the resonant optical mode and the active region emission spectrum is obtained for the shortest cavity.

The largest enhancements are achieved with the shortest cavities, which in turn are obtained if the *fundamental* cavity mode is in resonance with the emission from the active medium. The cavity length is also reduced by using a DBR with a short penetration depth, i.e. a DBR consisting of two materials with a large difference in the refractive index.

The **third design criterion** is the minimization of self-absorption in the active region. This criterion can be stated as follows: the reabsorption probability of photons emitted from the active region into the cavity mode should be much smaller than the escape probability of photons through one of the reflectors. Assuming $R_2 \approx 1$, this criterion can be written as

$$2 \xi \alpha L_{\text{active}} < (1 - R_1) \quad (10.3)$$

where α and L_{active} are the absorption coefficient and the thickness of the active region, respectively. If the criterion of Eq. (10.3) were not fulfilled, photons would be most likely to be reabsorbed by the active region. Subsequently, re-emission will, with a certain probability, occur along the lateral direction (waveguided modes), i.e. not into the cavity mode. Another possibility is that the electron-hole pairs generated by re-absorption recombine non-radiatively. In either case, re-absorption processes occurring in high-finesse cavities *reduce the cavity mode emission out of the cavity*. Thus, if the condition of Eq. (10.3) is not fulfilled, the emission intensity of resonant cavities is lowered rather than enhanced.

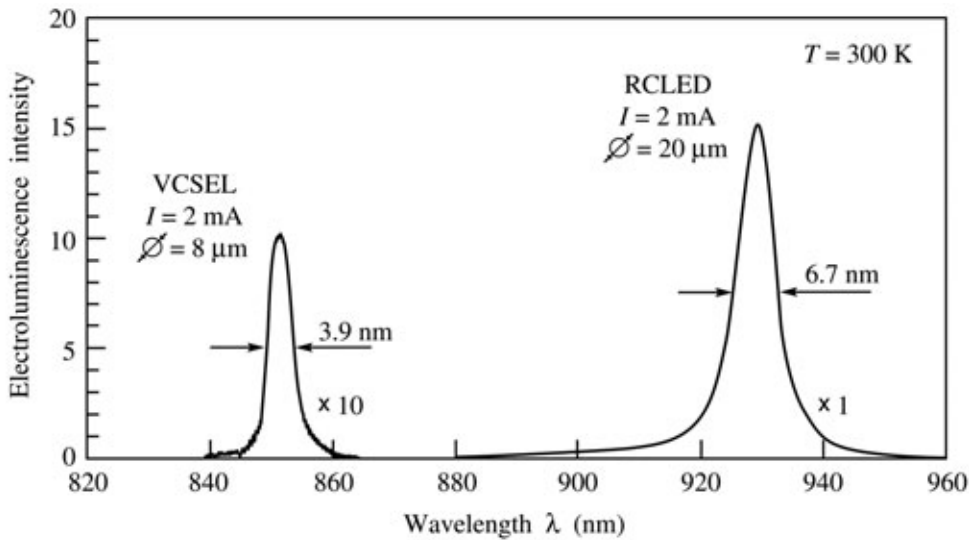


Fig. 10.3. Spontaneous electroluminescence spectrum of a vertical-cavity surface-emitting laser (VCSEL) emitting at 850 nm and of a resonant-cavity light-emitting diode (RCLED) emitting at 930 nm. The drive current for both devices is 2 mA. The VCSEL spectrum is multiplied by a factor of 10. The threshold current of the VCSEL is 7 mA (after Schubert *et al.*, 1996).

Whereas the condition of Eq. (10.3) is fulfilled in RCLEDs, it is clearly not fulfilled in VCSELs. The spontaneous emission intensities of RCLEDs and VCSELs were compared by Schubert *et al.* (1996). In this comparison, the VCSEL and the RCLED were driven by an injection current of 2 mA, which is below the threshold current of the VCSEL of $I_{\text{th}} = 7$ mA. The spontaneous emission spectra of an RCLED and a VCSEL are shown in Fig. 10.3. The VCSEL has an AlGaAs/GaAs quantum-well active region emitting at 850 nm. Both reflectors of the VCSEL are AlGaAs/AlAs DBRs. Figure 10.3 reveals that the emission intensity of the VCSEL

in the *spontaneous* regime is more than a factor of 10 lower than the emission intensity from the RCLED.

Because the magnitude of the maximum gain in semiconductors is always lower than the magnitude of the absorption coefficient in an unpumped semiconductor ($|g| < |\alpha|$), VCSELs could not lase if the condition of Eq. (10.3) were met. Thus, the spontaneous emission intensity in VCSELs is low and *must be low* in order to enable the device to lase. Figure 10.3 also reveals that the emission spectral linewidth of VCSELs is narrower than that of RCLEDs. The higher spectral purity is due to the higher values of R_1 and R_2 as required for VCSELs.

The fulfillment of Eq. (10.3) by RCLEDs also implies that these devices *cannot lase*. As stated above, it is always $|g| < |\alpha|$. Consequently, the mirror loss $(1 - R_1)$ is always larger than the maximum achievable round-trip gain ($2 \xi g L_{\text{active}}$). The fundamental inability of RCLEDs to lase has been verified experimentally by pulsed injecting currents of large magnitude without finding any evidence for lasing. These considerations show that the device physics of RCLEDs and VCSELs is fundamentally different.

The arguments used above imply that the *spontaneous* emission into the fundamental cavity mode in VCSEL structures is very low due to reabsorption of photons by the active region. A reduction of the threshold current by increasing the reflectivity will be accompanied by a further decrease of the *spontaneous* emission below threshold. We therefore conclude that the so-called zero-threshold laser (Kobayashi *et al.* 1982; Yokoyama, 1992) cannot be realized by a planar resonant-cavity structure.

10.3 GaInAs/GaAs RCLEDs emitting at 930 nm

The structure of an RCLED with a GaInAs active region is shown in Fig. 10.4 (a). The cavity is defined by one distributed Bragg reflector (DBR) and one metallic reflector. Also included are two confinement regions and a four-quantum-well active region. The heavily doped n-type substrate is coated with a ZrO_2 anti-reflection layer (Schubert *et al.*, 1994). A picture of the first RCLED is shown in Fig. 10.4 (b).

The motivation for the metal reflector is twofold. *First*, the metallic Ag reflector serves as a non-alloyed ohmic contact to the heavily doped p-type ($N_A \approx 5 \times 10^{19} \text{ cm}^{-3}$) GaAs top layer, thus effectively confining the pumped region to the area below the contact. *Secondly*, it was shown in the preceding section that the cavity length must be kept as short as possible for maximizing the emission enhancement. Owing to the lack of a penetration depth, metal reflectors allow for a short cavity length. Cavities with two metallic reflectors have been reported (Wilkinson *et al.*,

1995). However, optical absorption losses in the light-exit mirror can be large in a double metal mirror structure, unless very thin metallic reflectors are used (Tu *et al.*, 1990). The lack of a p-type DBR also avoids the well-known problem of high resistance in p-type DBRs (Schubert *et al.*, 1992c; Lear and Schneider, 1996). It has been shown that parabolic grading yields the lowest ohmic resistance in DBRs. Such parabolic grading is suited to eliminating heterojunction band discontinuities (Schubert *et al.*, 1992c).

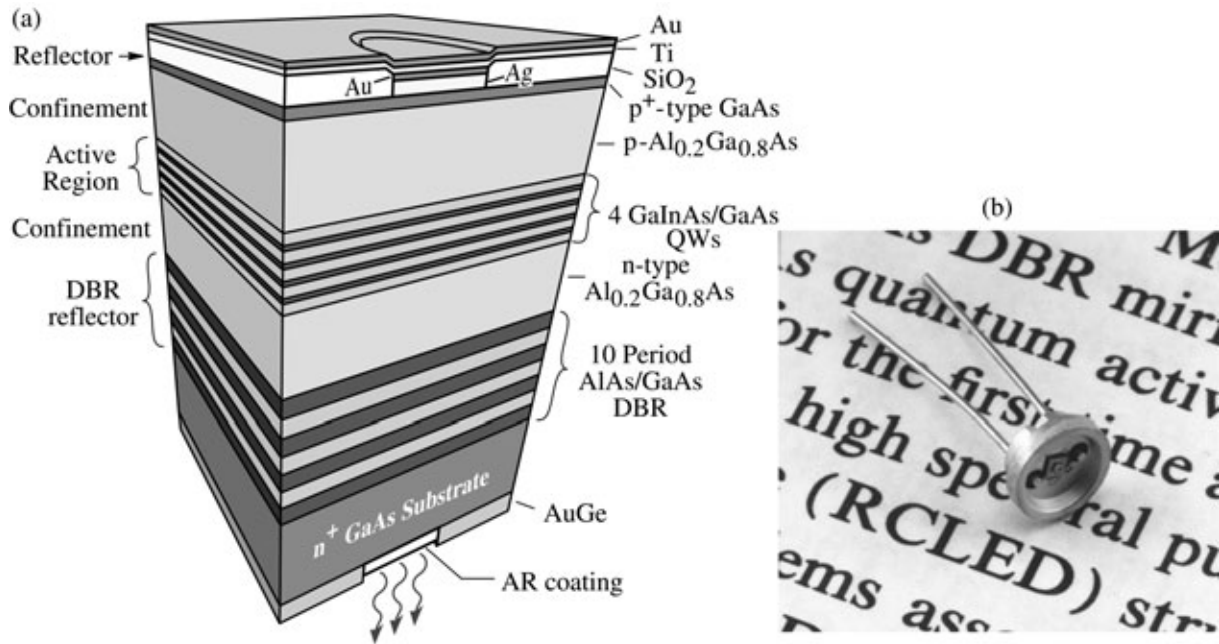


Fig. 10.4. (a) Schematic structure of a substrate-emitting GaInAs/GaAs RCLED consisting of a metal top reflector and a bottom distributed Bragg reflector (DBR). The RCLED emits at 930 nm. The reflectors are an AlAs/GaAs DBR and a Ag top reflector. (b) Picture of the first RCLED (after Schubert *et al.*, 1994).

The magnitude of the reflectivity of the DBR needs to be consistent with Eqs. (10.1) and (10.3). The Ag back mirror has a reflectivity of approximately 96 %. According to Eq. (10.1), the DBR reflectivity must be $< 96 \%$. The second criterion of Eq. (10.3) requires that $2 \xi \alpha L_{\text{active}} < 1 - R_1$. Assuming $\xi = 1.3$, $\alpha = 10^4 \text{ cm}^{-1}$, $L_{\text{active}} = 400 \text{ \AA}$, one obtains the condition $R_1 < 90 \%$. Thus the mirror reflectivity of RCLEDs must be much lower than that of VCSELs. A high reflectivity would increase self-absorption and decrease the light output of the device as discussed earlier. De Neve *et al.* (1995) used an extensive theoretical model to calculate the mirror reflectivity. The maximum efficiency was calculated at a reflectivity of $R_1 = 50 - 60 \%$.

The reflection and emission properties of the RCLED are shown in Figs. 10.5 (a) and (b). The reflection spectrum of the RCLED (Fig. 10.5 (a)) exhibits a highly reflective band for

wavelengths > 900 nm and a dip in the reflectivity at the cavity resonance. The spectral width of the cavity resonance is 6.3 nm. The emission spectrum of an electrically pumped device, shown in Fig. 10.5 (b), has nearly the same shape and width as the cavity resonance.

In conventional LEDs, the spectral characteristics of the devices reflect the thermal distribution of electrons and holes in the conduction and valence band. The spectral characteristics of light emission from microcavities are as intriguing as they are complex. However, restricting our considerations to the optical axis of the cavity simplifies the cavity physics considerably. If we assume that the cavity resonance is much narrower than the natural emission spectrum of the semiconductor, then the on-resonance luminescence is enhanced whereas the off-resonance luminescence is suppressed. The on-axis emission spectrum should therefore reflect the enhancement, that is, the resonance spectrum of the cavity. The experimental results shown in Fig. 10.5 confirm this conjecture.

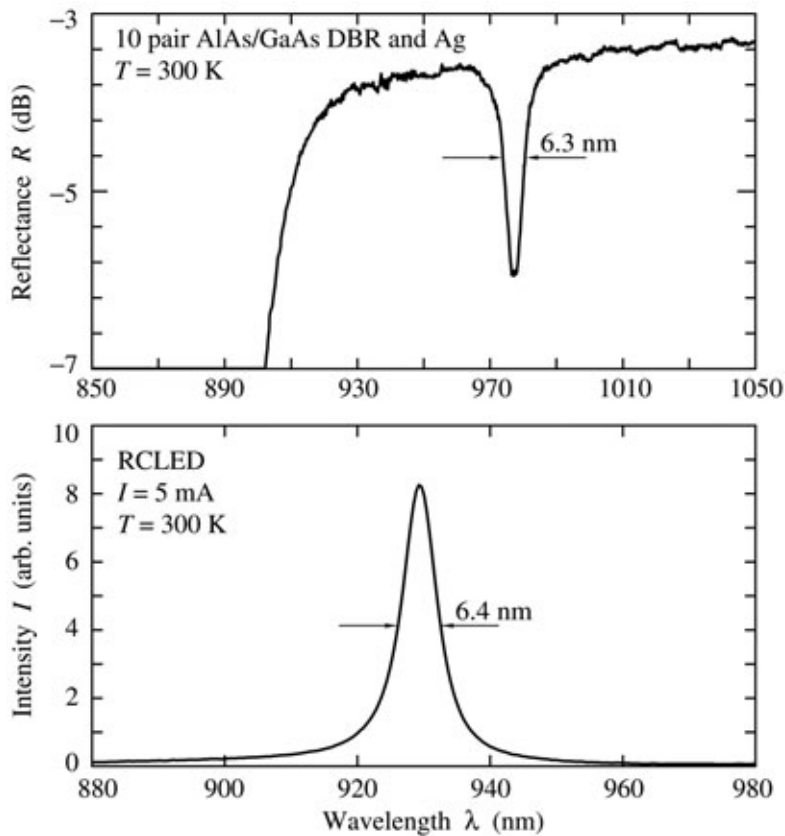


Fig. 10.5. (a) Reflectance of a resonant cavity consisting of a 10-pair AlAs/GaAs distributed Bragg reflector and an Ag reflector. (b) Emission spectrum of a RCLED consisting of a 10-pair AlAs/GaAs distributed Bragg reflector and an Ag reflector (after Schubert *et al.*, 1994).

Owing to the cavity, the emission spectrum of an RCLED is much narrower than the emission spectrum of regular LEDs (Schubert *et al.* 1992a; Hunt *et al.*, 1992, 1993). The spectral width of the RCLED emitted into a certain direction is given by the optical characteristics of the

cavity. In contrast, the spectral width of a regular LED is about $1.8 kT$, a value that is much wider than the RCLED emission spectrum. A comparison of a regular GaAs LED and a GaInAs RCLED emission spectrum is shown in Fig. 10.6. Comparison of the spectra shows that the RCLED emission spectrum is about a factor of about 10 narrower than the spectrum of the GaAs LED.

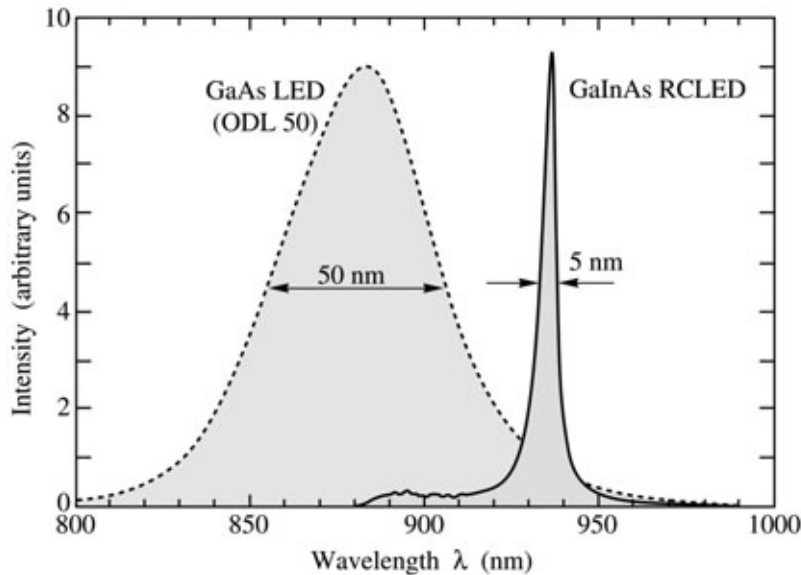


Fig. 10.6. Comparison of the emission spectra of a GaAs LED emitting at 870 nm (AT&T ODL 50 product) and a GaInAs RCLED emitting at 930 nm (after Hunt *et al.*, 1993).

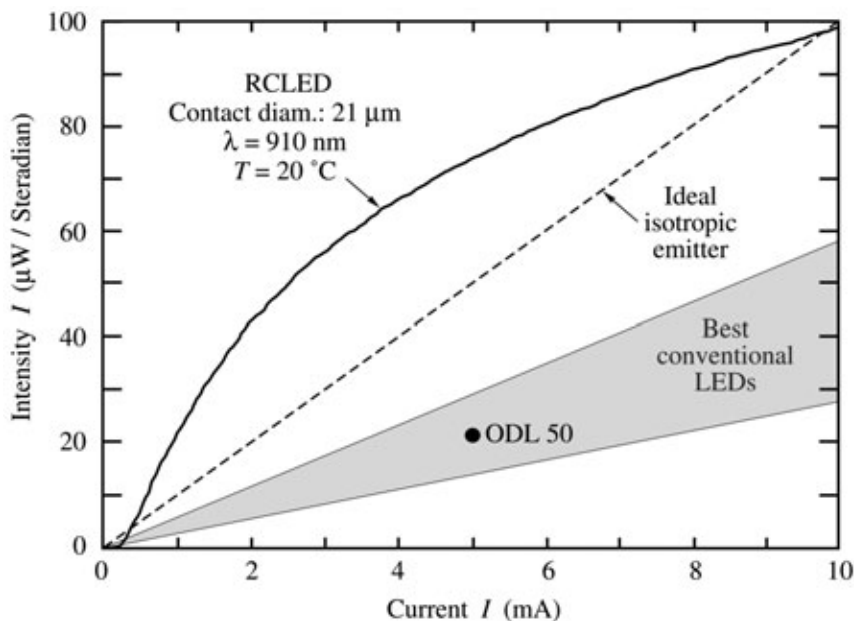


Fig. 10.7. Light-versus-current curves of a GaInAs/GaAs RCLED and of the *ideal isotropic emitter*. The ideal isotropic emitter is a hypothetical device emitting light isotropically with a quantum efficiency of 100%. The shaded region shows the intensity of the best conventional LEDs. The ODL 50 is a commercial LED product (after Schubert *et al.*, 1994).

A regular LED has little or no angle dependence of the emission spectrum. However, the reflective properties of DBRs and of cavities consisting of DBRs are angle-dependent. Consequently, the emission from an RCLED along a certain direction is narrower than that of a regular LED. Integrated over all directions, the RCLED has a broad emission spectrum.

A figure of merit for LEDs used in optical fiber communication systems is the photon flux density emitted from the diode at a given current, which, for a given wavelength, can be characterized in terms of the units microwatts per steradian. The optical power coupled into a fiber is directly proportional to the photon flux density.

The intensity of an RCLED as a function of the injection current is shown in Fig. 10.7. For comparison, the calculated intensity of the *ideal isotropic emitter*, which is a hypothetical device, is also shown. The ideal isotropic emitter is assumed to have an internal quantum efficiency of 100 % and the device is assumed to be clad by an anti-reflection coating providing zero reflectivity ($R = 0$) for all wavelengths emitted from the active region. If the photon emission inside the semiconductor is isotropic, then the optical power per unit current per unit solid angle normal to the planar semiconductor surface is given by

$$\frac{P_{\text{optical}}}{\Omega} = \frac{1}{4\pi \bar{n}^2} \frac{h c}{e \lambda} \quad (10.4)$$

where Ω represents the unit solid angle, \bar{n} is the refractive index of the semiconductor, c is the velocity of light, e is the electronic charge, and λ is the emission wavelength in vacuum. Equation (10.4) is represented by the dashed line in Fig. 10.7. Neither the 100 % internal quantum efficiency nor the hypothetical anti-reflection coating can be reproduced in practice for fundamental reasons. Therefore, the ideal isotropic emitter represents an upper limit for the intensity attainable with any conventional LED. Of course, even the best conventional LEDs have intensities lower than that of the ideal isotropic emitter. Also included in Fig. 10.7 is the state-of-the-art ODL 50 GaAs LED used for optical fiber communication. All devices shown in Fig. 10.7 have planar light-emitting surfaces, and no lensing is used.

Figure 10.7 reveals that the RCLED provides unprecedented intensities in terms of both absolute values and slope efficiencies. The slope efficiency is 7.3 times the efficiency of the best conventional LEDs and 3.1 times the calculated efficiency of the ideal isotropic emitter. At a current of 5 mA, the intensity of the RCLED is 3.3 times that of the best conventional LEDs including the ODL 50. The high efficiencies make the RCLED well suited for optical

interconnect and communication systems.

The higher spectral purity of RCLEDs reduces chromatic dispersion in optical fiber communications (Hunt *et al.*, 1993). The chromatic dispersion is directly proportional to the linewidth of the source. Since RCLEDs have linewidths 5 – 10 times narrower than conventional LEDs, chromatic dispersion effects, which dominate at wavelengths of 800 – 900 nm, are reduced as well. Hunt *et al.* (1993) showed that the bandwidth of RCLEDs is a factor of 5 – 10 higher than that of conventional LEDs. An RCLED-versus-LED comparison in a transmission experiment is shown in Fig. 10.8. The results show the received signal after a transmission length of 5 m and 3.4 km for the two devices. The fiber used is a graded-index multimode fiber with a core diameter of 62.5 μm . After a transmission length of 5 m, no marked difference is found for the two devices. However, a substantial difference is found after a transmission length of 3.4 km. Inspection of Fig. 10.8 reveals that the RCLED exhibits much less pulse broadening as compared to the conventional LED. This difference is due to reduced material dispersion for the RCLED.

Schubert *et al.* (1996) demonstrated the high-speed modulation capability of RCLEDs. Eye diagram measurements with a random bit pattern generator revealed wide-open eyes at 622 Mbit/s. Due to the small size of the current injected region, the parasitic capacitances of communication RCLEDs are small. It is expected that RCLEDs are suitable for modulation frequencies exceeding 1 Gbit/s.

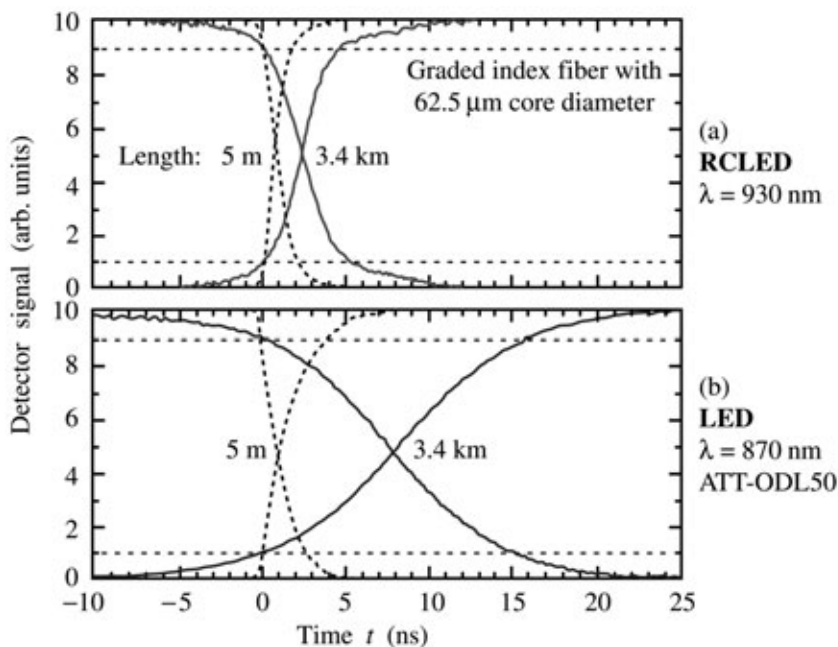


Fig. 10.8. Signal detected at the receiver end of a graded-index multimode fiber with a core diameter of 62.5 μm using an (a) GaInAs RCLED and (b) GaAs LED source. After a short transmission distance of 5 m, no marked difference is found for the two sources. After a transmission distance of 3.4 km, the RCLED exhibits much less pulse broadening than the LED (after Hunt *et al.*, 1993).

10.4 AlGaInP/GaAs RCLEDs emitting at 650 nm

RCLEDs have also been demonstrated in the visible wavelength range using the AlGaInP material system (Streubel *et al.*, 1998; Whitaker, 1999; Wirth *et al.*, 2001, 2002). The AlGaInP material system is commonly used for high-brightness red, orange, and yellow emitters and is can be grown lattice-matched on GaAs substrates. The active region of RCLEDs is an AlGaInP/GaInP multiple quantum well structure emitting at 650 nm. The RCLEDs are suited for use in communication systems using plastic optical fibers. It is difficult to fabricate VCSELs in this wavelength range due to the unavailability of lattice-matched and transparent DBR materials with a large refractive-index contrast.

The basic structure of a top-emitting AlGaInP RCLED emitting at 650 nm is shown in Fig. 10.9. The device consists of an AlGaInP/GaInP MQW active region and AlGaInP cladding layers. The DBRs consist of AlAs/AlGaAs layers. The Al content in the AlGaAs layers of the DBR is chosen to be sufficiently high to make the DBR transparent to the light emitted by the active region. As a result, the index contrast of the AlAs/AlGaAs DBR layers is somewhat low.

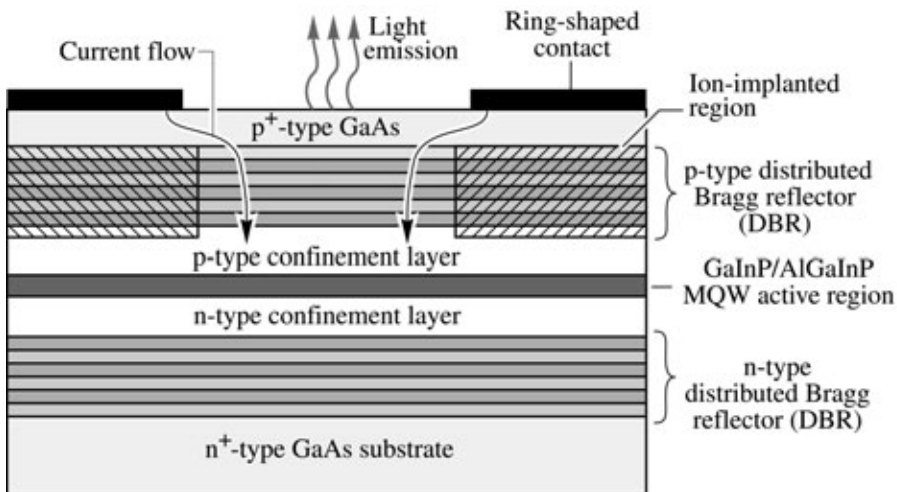


Fig. 10.9. Structure of a GaInP/AlGaInP/GaAs MQW RCLED emitting at 650 nm used for plastic optical fiber applications (after Whitaker, 1999)

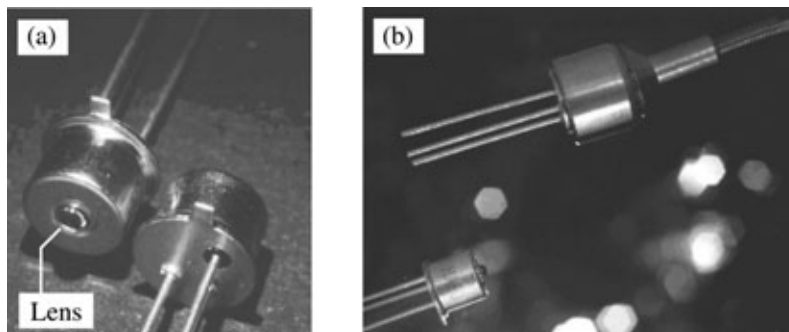


Fig. 10.10. (a) Packaged (TO package) RCLED emitting at 650 nm suited for plastic optical fiber applications. (b) Pig-tailed RCLED (courtesy of Mitel Corporation, Sweden, 1999).

The RCLED has a ring-shaped top contact configuration. The current is funneled into the center region of the ring using ion implantation to create an insulating region under the ring-shaped metalization. Hydrogen and, more frequently, oxygen implantation is used to render the semiconductor highly resistive. Note that the ion-implanted region is located in the p-type region only and does not extend into the active region, thereby avoiding the creation of defects in the active region where they would act as luminescence killers.

Packaged RCLEDs in a lensed TO package and in a pig-tailed package are shown in Figs. 10.10 (a) and (b), respectively. The lens is used for beam collimation, thereby enhancing the coupling efficiency to fibers.

Three RCLEDs under current injection conditions are shown in Fig. 10.11. The picture shows that the emission pattern is directed towards the surface normal of the devices. The emission wavelength is 650 nm.

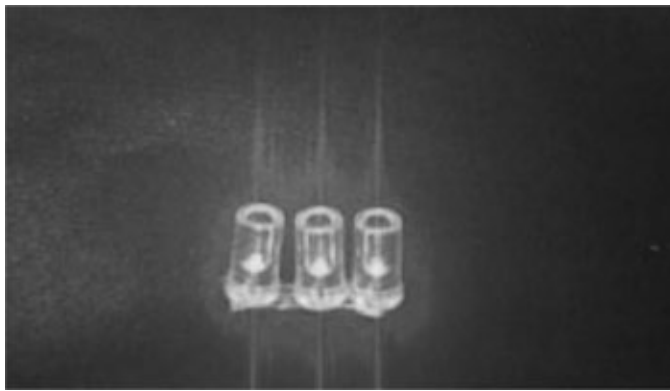


Fig. 10.11. AlGaInP/GaAs RCLEDs emitting at 650 nm. Note the forward-directed emission pattern similar to that of a semiconductor laser (courtesy of Osram Opto Semiconductors Corporation, Germany, 1999).

Optical spectra of a 650 nm RCLED and of a conventional LED injected at different current levels are shown in Fig. 10.12. The spectra are measured after the light is coupled into a plastic optical fiber. Thus, the magnitude of the spectra is a direct measure of the device efficiency *and* of the coupling efficiency. Inspection of the figure reveals several features. *First*, the RCLED has a higher coupled peak power as well as integrated power than the LED. *Secondly*, the RCLED has a higher spectral purity than the LED.

Streubel *et al.* (1998) also reported that the emission spectrum was, at room temperature, intentionally blue-shifted with respect to the cavity resonance in order to improve the temperature stability of the RCLED output power. This cavity tuning results in a heart-shaped (double-lobed) emission pattern at room temperature, because the resonance wavelength of the cavity decreases for off-normal emission directions. As the temperature increases, the natural emission spectrum from the active region redshifts so that the cavity resonance (along the normal

direction) has a better overlap with the natural emission spectrum. As a result, the temperature sensitivity of the RCLED is reduced. The natural emission spectrum redshifts at a rate of about $0.5 \text{ nm}/^\circ\text{C}$, whereas the cavity resonance shifts at only about one-tenth of that rate.

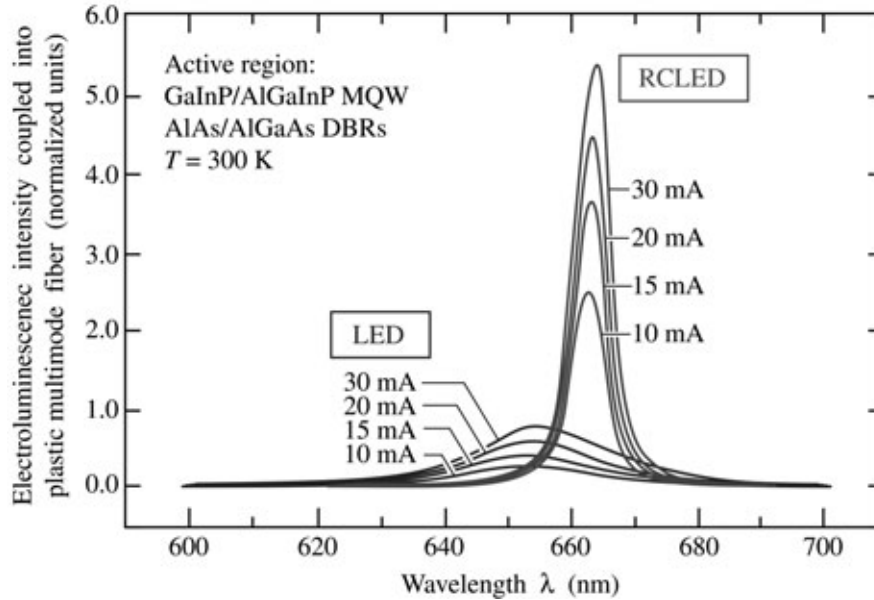


Fig. 10.12. Spectra of light coupled into a plastic optical fiber from an GaInP/AlGaInP MQW RCLED and a conventional GaInP/AlGaInP LED at different drive currents. Note the narrower spectrum and higher coupled power of the RCLED (after Streubel *et al.*, 1998).

10.5 Large-area photon recycling LEDs

Instead of devising ways to *reduce the lateral emission*, one could devise ways to *recycle* lateral emission, thereby redirecting the emission towards the top of the device. By re-absorbing the lateral emission, one can re-capture the energy into the active region, giving it another chance to emit along the desired direction. Below, two examples of such devices are discussed.

The first example of such a device is an optically pumped semiconductor structure reported by Schnitzer *et al.* (1993). Consider the structure shown in Fig. 10.13 (a). The backside of a thin semiconductor layer is coated with gold. If one were to optically pump the sample at low intensities, the active region remains absorbing. Most of the emission (about 95 %) that hits the semiconductor–air interface will totally internally reflect, and stay within the semiconductor. Gold is a good reflector at infrared wavelengths. The semiconductor forming the active region is absorbing at the emission wavelength, so after some average absorption length, L_{abs} , the trapped light has a chance to re-emit. This structure exhibited 72 % external efficiency, which required a 99.7 % internal quantum efficiency. The fabrication of the structure required the active epilayer to be etched and floated off its substrate. Note that the structure had no electrical contacts. The extraction efficiency of contactless structures is always higher compared with current-injection devices with contacts. An electrical device fabricated in this manner would probably have reliability

problems due to the fragile nature of unsupported epitaxial films. A similar but more practical photon-recycling device was made by Blondelle *et al.* (1995).

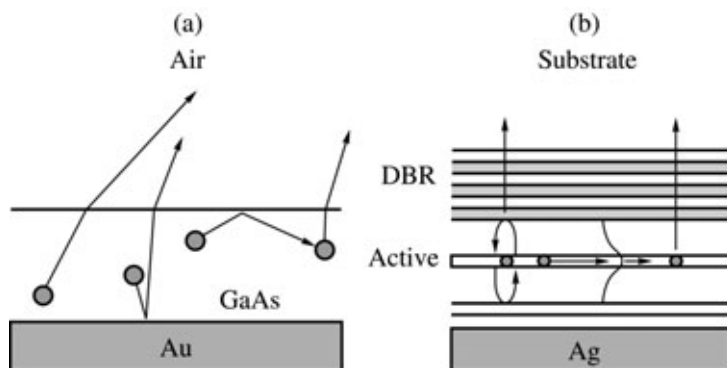


Fig. 10.13. Two approaches to photon recycling LEDs. (a) Bulk epilayer placed on top of gold. Most spontaneous emission that does not escape into air is reabsorbed and has a chance to emit again. (b) Micro-cavity designed with a waveguiding active region. Waveguided light is reabsorbed after some tens of microns, and has a chance to re-emit out of the top of the device.

The concept of photon recycling was used by De Neve *et al.* (1995) in a GaInAs/GaAs/AlGaAs RCLED. A simplified diagram of their structure is shown in Fig. 10.13 (b). If the graded-composition carrier confinement region is thick enough around the quantum wells, it also acts as a graded-index waveguide. De Neve *et al.* calculated that 30 % of the light emitted by the active region goes into this waveguide mode. This will get re-absorbed after some tens of microns, allowing the photons another chance to re-emit out of the top of the structure. In this way, about a 30 % increase in external quantum efficiency was attained. The active region of quantum well RCLEDs is not thick enough to support a strongly guided optical mode. One might think that a strong waveguiding mode would take power away from the normal emission, but this does not appear to be the case. The energy tends to be at the expense of other high-angle modes instead, meaning that just modifying the waveguiding mode does little to change the external efficiency. Making use of this high-angle light by achieving photon recycling of the waveguiding emission is an attractive option.

One drawback to photon recycling is that it requires a large enough diameter device that multiple re-absorption events can occur within the emitting area of the device. This makes them less attractive for fiber-optic communications, where small diameters couple better to fibers, especially when coupling lenses are used. Another drawback is that self-absorption necessarily increases the lifetime of the spontaneous emission, slowing down the maximum modulation rate of the devices.

In fact, the device reported by De Neve *et al.* (1995) was shown to work with highest efficiencies at large diameters anyway, and the authors were not targeting this RCLED for fiber communications. At high current densities, the carrier confinement in the quantum wells is reduced, and the efficiency decreases, so a larger device tends to be more efficient. Also, their device is

resonant to the long-wavelength side of the natural emission peak, rather than at the peak. This shifts the maximum intensity of the main emission lobe to some angle, rather than being on axis, but it maximizes the total amount of emission into the main lobe. The Blondelle *et al.* (1995) and De Neve *et al.* (1995) device is therefore designed for maximization of total emission from the top of the device, rather than for the specific needs of fiber coupling. Display devices or free-space communications devices benefit from this approach. The authors achieved an experimental maximum of 16 % external quantum efficiency, compared with the theoretical 2 % for an ideal bulk emitter of the same index.

10.6 Thresholdless lasers

There have been a number of papers describing a thresholdless laser as a possible communications device (De Martini *et al.*, 1987). This would essentially be a light-emitting device that emitted most or all of its light into the fundamental spatial cavity mode. In this way, one can achieve the lasing condition of there being one photon in the cavity almost without gain. Let us denote the probability that light is emitted into the fundamental spatial cavity mode as β . If the β of the LED were almost 1, then the light versus current curve of the device would be linear, and *indistinguishable* from a laser with no threshold. A comparison of the light versus current for a conventional laser, a high- β laser, and a truly thresholdless laser is given in Fig. 10.14. If one were to envision a thresholdless device as fabricated from a resonant-cavity pillar, there are a number of problems besides those of fabrication. Simply having a good sub-threshold intensity, giving only a small kink in the $L-I$ curve between sub-threshold and lasing, is not good enough for a thresholdless laser. The output modulation of such a device would be extremely slow below threshold compared to the lasing regime. To enable lasing, there must be some gain, because the β will never be 1. For there to be gain, we must pump the active region past transparency first, since it is absorbing under no-pump conditions. This gain condition requires a certain carrier density, which, in fact, determines the threshold.

One can imagine a device such that all optical modes other than the fundamental are highly suppressed, and the internal quantum efficiency is nearly perfect. In this case, with even a small current, the carrier densities within the device would slowly build until the active region became transparent, and some light could get out through the fundamental optical mode. Such a device would have a very small threshold, but one would have to be careful not to ever drive the device below its threshold even momentarily, because of the long time required to build the necessary space charge inside the device for transparency. Also, if the fundamental emission was suppressed

at low carrier densities, it is unlikely that the other modes would be suppressed by an even higher degree, which is what would be required to not lose the injected carriers. In fact, it is the difficulty in suppressing any side emission that makes semiconductor thresholdless lasers difficult if not impractical.

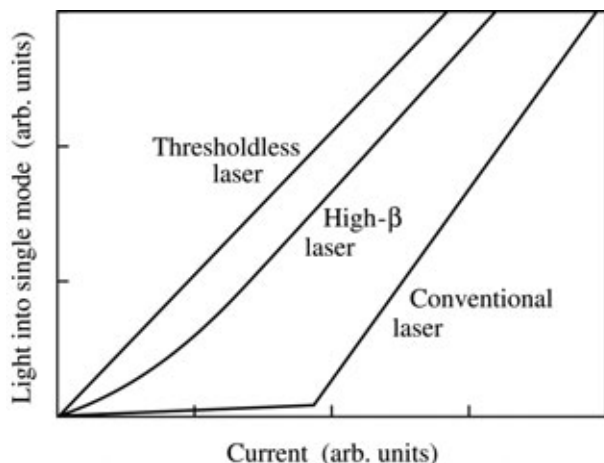


Fig. 10.14. Light-power-versus-current curves for single spatial-mode emission from a (i) conventional laser, (ii) a high β -factor laser, and (iii) a thresholdless laser. The conventional laser has a distinct current threshold. The high β -factor laser has a less distinct threshold. It would be noticeable in the spectrum and device modulation speed, however. A hypothetical thresholdless laser would have a β close to 1, and would somehow suppress all other lossy emission until the carrier density required for gain (or at least transparency) was achieved.

One could, of course, make a laser with a very small threshold by making the laser very small. The problem is that such a device would be capable of only a very small output power. For a hypothetical four quantum-well single mode LED, which does not rely on gain, but achieves a β close to 1, the current densities cannot be more than about $50 \mu\text{A}/\mu\text{m}^2$ or $5 \text{ kA}/\text{cm}^2$. Using the formulas for β for a dielectric pillar ($\beta > 0.5$), and this current density, one can find that less than $2 \mu\text{W}$ could be emitted into a single mode for an RCLED pillar structure. This is clearly too small for high-speed communications in the 100 – 650 Mbit/s regime, where $10 \mu\text{W}$ is considered a minimum for an emitter. An RCLED with a very short carrier lifetime owing to extreme enhancement of the desired emission could have higher pump currents, but no semiconductor design can do this. If the device is a laser, higher powers could be drawn because of the reduced carrier lifetime, but the device sizes are incredibly small for real fabrication, or for any sort of heat dissipation.

In conclusion, semiconductor RCLEDs should remain devices larger than about 5 or 10 microns in diameter, and should remain multi-mode emitters. Semiconductor vertical-cavity lasers are best fabricated at sizes that make sense technologically, without regard to the β factor of the spontaneous emission. A high- β laser may not be desirable anyway, since the spontaneous emission will introduce excess noise into the emission.

10.7 Other RCLED devices

It is likely that the future of resonant-cavity LEDs and confined-photon emitters will involve new materials and applications from those currently used in communications and display systems. Spontaneously emitting confined-photon devices will always have competition from lasers, conventional LEDs, and other forms of devices. It is only a matter of time, however, before the right materials combination at the right wavelengths, for the right application, will make commercial devices a reality. It is therefore worthwhile to mention recent confined photonic LED materials systems.

Wilkinson *et al.* (1995) fabricated an AlGaAs/GaAs thin-film emitter with metal mirrors on silicon substrate, emitting at 880 nm. Pavesi *et al.* (1996) fabricated porous silicon RCLEDs at 750 nm wavelength. Fisher *et al.*, (1995) investigated a conjugate polymer RCLED designed for 650 nm. Hadji *et al.* (1995) have realized CdHgTe/HgTe RCLEDs operating at a wavelength of 3.2 μm .

A structure of particular note is a GaAs/Al_xO_y RCLED (Huffaker *et al.*, 1995) operating at 950 nm, where the aluminum oxide in the output mirror was produced from AlAs by oxidation. This composition of output mirror allows the effective cavity length to remain small, maximizing the output enhancement in the RCLED. The back mirror of the structure was Ag. Another particularly interesting device is the broadly tunable RCLED by Larson and Harris (1995). The top mirror is a deformable membrane which can be moved by electrostatic forces. Tunable emission was shown from 938 to 970 nm.

10.8 Other novel confined-photon emitters

A complete discussion of other light-emitting structures that confine photons is beyond the scope of this chapter. It is useful, however, to discuss the properties of some confined-photon emitters. One such structure is the microdisk laser (McCall *et al.*, 1992) that is fabricated as a thin dielectric disk, which couples light out the edges of the disk. Lasing modes can be described by a mode number M , where $\exp(i M \phi)$ is the form of the electric field around the cylindrical disk. Because waves can propagate both ways, M can be positive or negative. The disk can be fabricated with a thickness such that the emission perpendicular to the disk is suppressed. Small disks will only support a few modes, and therefore can have a high spontaneous emission factor β . The Q of these modes are also high enough to achieve lasing. One attractive aspect of such disks is that the lasing emission occurs in the plane of the sample, from a very small device. This could be useful for integration of many photonic devices on a single wafer. However, the output is difficult to efficiently couple into

waveguides and fibers, as it only couples evanescently. Advances have been made in improving the longevity, operating temperature range, active-region passivation of such devices (Mohideen *et al.*, 1993). Room-temperature CW electrical pumping is still a problem, however.

Photonic bandgap structures may someday yield practical LEDs. These structures involve two- or three-dimensional photon confinement by periodic patterning of the emitting material. Examples of photonic bandgap structures were given by Baba and Matsuzaki (1996) and Fan *et al.* (1997). Such a structure could consist of a series of rods or holes in a regular pattern, such as a hexagonal close-packed array. The periodicity of the array creates an optical bandgap for lateral emission at certain emission energies and one or both polarizations. By suppressing the lateral emission, the rod structure will have a large bandgap for TM emission and a smaller gap for TE emission, but not at the same emission energies. However, if the emitting region had a dipole oriented mainly along the rods (such as in quantum well electron-to-light-hole recombination), the lateral emission could be efficiently suppressed. The structure with the patterned holes will have a smaller bandgap than that of the rod structure, but it has the advantage that it can have a true optical bandgap for both polarizations of light. By themselves, or combined with a planar resonant cavity, longitudinal emission could theoretically be enhanced.

References

- Baba T. and Matsuzaki T., "GaInAsP/InP 2-dimensional photonic crystals" in "Microcavities and photonic bandgaps" edited by J. Rarity and C. Weisbuch, p. 193 (Kluwer Academic Publishers, Netherlands, 1996)
- Blondelle J., De Neve H., Demeester P., Van Daele P., Borghs G., and Baets R. "16 % external quantum efficiency from planar microcavity LED's at 940 nm by precise matching of the cavity wavelength" *Electron. Lett.* **31**, 1286 (1995)
- De Martini F., Innocenti G., Jacobovitz G. R., and Mataloni P. "Anomalous spontaneous emission time in a microscopic optical cavity" *Phys Rev. Lett.* **59**, 2955 (1987)
- De Neve H., Blondelle J., Baets R., Demeester P., Van Daele P., and Borghs G. "High efficiency planar microcavity LEDs: Comparison of design and experiment" *IEEE Photonics Tech. Lett.* **7**, 287 (1995)
- Fan S., Villeneuve P. R., Joannopolous J. D., and Schubert E. F. "High extraction efficiency of spontaneous emission from slabs of photonic crystals" *Phys. Rev. Lett.* **78**, 3294 (1997)
- Fisher T. A., Lidzey D. G., Pate M. A., Weaver M. S., Whittaker D. M., Skolnick M. S., and Bradley D. D. C. "Electroluminescence from a conjugated polymer microcavity structure"

- Appl. Phys. Lett.* **67**, 1355 (1995)
- Hadji E., Bleuse J., Magnea N., and Pautrat J. L. “3.2 μm infrared resonant cavity light emitting diode” *Appl. Phys. Lett.* **67**, 2591 (1995)
- Huffaker D. L., Lin C. C., Shin J., and Deppe D. G. “Resonant cavity light emitting diode with an $\text{Al}_x\text{O}_y/\text{GaAs}$ reflector” *Appl. Phys. Lett.* **66**, 3096 (1995)
- Hunt N. E. J., Schubert E. F., Logan R. A., and Zydzik G. J. “Enhanced spectral power density and reduced linewidth at 1.3 μm in an InGaAsP quantum well resonant-cavity light-emitting diode” *Appl. Phys. Lett.* **61**, 2287 (1992)
- Hunt N. E. J., Schubert E. F., Kopf R. F., Sivco D. L., Cho A. Y., and Zydzik G. J. “Increased fiber communications bandwidth from a resonant cavity light-emitting diode emitting at $\lambda = 940 \text{ nm}$ ” *Appl. Phys. Lett.* **63**, 2600 (1993)
- Hunt N. E. J., Schubert E. F., Sivco D. L., Cho A. Y., Kopf R. F., Logan R. A., and Zydzik G. J. “High efficiency, narrow spectrum resonant cavity light-emitting diodes” in “Confined electrons and photons” edited by E. Burstein and C. Weisbuch (Plenum Press, New York, 1995a)
- Hunt N. E. J., Vredenberg A. M., Schubert E. F., Becker P. C., Jacobson D. C., Poate J. M., and Zydzik G. J. “Spontaneous emission control of Er^{3+} in Si/SiO₂ microcavities” in “Confined electrons and photons” edited by E. Burstein and C. Weisbuch (Plenum Press, New York, 1995b)
- Kobayashi T., Segawa T., Morimoto A., and Sueta T., paper presented at the 43rd fall meeting of the Japanese Society of Applied Physics, Tokyo, Sept. (1982)
- Larson M. C. and Harris J. S. Jr. “Broadly tunable resonant-cavity light emission” *Appl. Phys. Lett.* **67**, 590 (1995)
- Lear K. L. and Schneider Jr. R. P. “Uniparabolic mirror grading for vertical cavity surface emitting lasers” *Appl. Phys. Lett.* **68**, 605 (1996)
- McCall S. L., Levi A. F. J., Slusher R. E., Pearton S. J., and Logan R. A. “Whispering-gallery mode microdisk laser” *Appl. Phys. Lett.* **60**, 289 (1992)
- Mitel Corporation, Sweden. Photograph of RCLED is gratefully acknowledged (1999)
- Mohideen U., Hobson W. S., Pearton J., Ren F., and Slusher R. E. “GaAs/AlGaAs microdisk lasers” *Appl. Phys. Lett.* **64**, 1911 (1993)
- Nakayama T., Itoh Y., and Kakuta A. “Organic photo- and electroluminescent devices with double mirrors” *Appl. Phys. Lett.* **63**, 594 (1993)
- Osram Opto Semiconductors Corporation, Germany. Photograph of RCLED is gratefully

acknowledged (1999)

- Pavesi L., Guardini R., and Mazzoleni C. "Porous silicon resonant cavity light emitting diodes" *Solid State Comm.* **97**, 1051 (1996)
- Schnitzer I., Yablonovitch E., Caneau C., and Gmitter T. J. "Ultra-high spontaneous emission quantum efficiency, 99.7 % internally and 72 % externally, from AlGaAs/GaAs/AlGaAs double heterostructures" *Appl. Phys. Lett.* **62**, 131 (1993)
- Schubert E. F., Wang Y.-H., Cho A. Y., Tu L.-W., and Zydzik G. J. "Resonant cavity light-emitting diode" *Appl. Phys. Lett.* **60**, 921 (1992a)
- Schubert E. F., Vredenberg A. M., Hunt N. E. J., Wong Y. H., Becker P. C., Poate J. M., Jacobson D. C., Feldman L. C., and Zydzik G. J. "Giant enhancement of luminescence intensity in Er-doped Si/SiO₂ resonant cavities" *Appl. Phys. Lett.* **61**, 1381 (1992b)
- Schubert E. F., Tu L. W., Zydzik G. J., Kopf R. F., Benvenuti A., and Pinto M. R. "Elimination of heterojunction band discontinuities by modulation doping" *Appl. Phys. Lett.* **60**, 466 (1992c)
- Schubert E. F., Hunt N. E. J., Micovic M., Malik R. J., Sivco D. L., Cho A. Y., and Zydzik G. J. "Highly efficient light-emitting diodes with microcavities" *Science* **265**, 943 (1994)
- Schubert E. F., Hunt N. E. J., Malik R. J., Micovic M., and Miller D. L. "Temperature and modulation characteristics of resonant cavity light-emitting diodes" *IEEE J. Lightwave Technol.* **14**, 1721 (1996)
- Streubel K., Helin U., Oskarsson V., Backlin E., and Johanson A. "High-brightness visible (660 nm) resonant-cavity light-emitting diode" *IEEE Photon. Technol. Lett.* **10**, 1685 (1998)
- Tu L. W., Schubert E. F., Zydzik G. J., Kopf R. F., Hong M., Chu S. N. G., and Mannaerts J. P. "Vertical cavity surface emitting lasers with semitransparent metallic mirrors and high quantum efficiencies" *Appl. Phys. Lett.* **57**, 2045 (1990)
- Whitaker T. "Resonant Cavity LEDs" *Compound Semiconductors*, **5**, 32 (May 1999)
- Wilkinson S. T., Jokerst N. M., and Leavitt R. P. "Resonant-cavity-enhanced thin-film AlGaAs/GaAs/AlGaAs LED's with metal mirrors" *Appl. Optics* **34**, 8298 (1995)
- Wirth R., Karnutsch C., Kugler S., and Streubel K. "High-efficiency resonant-cavity LEDs emitting at 650 nm" *IEEE Photonics Technol. Lett.* **13**, 421 (2001)
- Wirth R., Huber W., Karnutsch C., and Streubel K. "Resonators provide LEDs with laser-like performance" *Compound Semiconductors* **8**, 49 (2002)
- Yokoyama H. "Physics and device applications of optical microcavities" *Science* **256**, 66 (1992)

Human vision

The recipient of the light emitted by most visible-spectrum LEDs is the human eye. In this chapter, the properties of the human eye and human vision are summarized, in particular the properties relating to the light intensity and color. *Colorimetry*, or the science of colors and human vision, has attracted scientists for many centuries (for a review, see MacAdam, 1993).

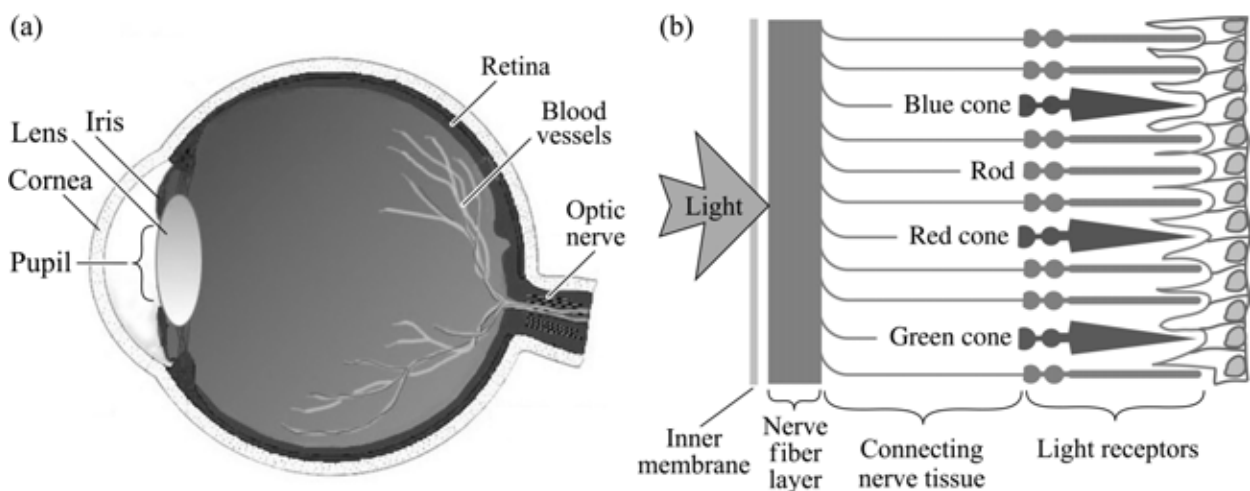


Fig. 11.1. (a) Cross section through a human eye. (b) Schematic view of the retina including rod and cone light receptors (adapted from Encyclopedia Britannica, 1994).

11.1 Light receptors of the human eye

Figure 11.1 (a) shows a schematic illustration of the human eye (Encyclopedia Britannica, 1994). The inside of the eyeball is clad by the retina, which is the light-sensitive part of the eye. Figure 11.1 (b) shows the cell structure of the retina including the light-sensitive *rod cells* and *cone cells* and the nerves transmitting the information to the brain. Rod cells are more abundant and more light sensitive than cone cells. Rods are sensitive over the entire visible spectrum. There are three types of cone cells, namely cone cells sensitive in the red, green, and blue spectral range. The cone cells are therefore denoted as the red, green, and blue cones.

Photopic vision relates to human vision at high ambient light levels (e.g. during daylight conditions) when vision is mediated by the cones. **Scotopic vision** relates to human vision at low ambient light levels (e.g. at night) when vision is mediated by rods. Rods have a much higher sensitivity than the cones. The sense of color is essentially lost in the scotopic vision regime. At night, objects lose their colors and only appear as objects with different gray levels. The following discussion relates to the photopic vision regime.

11.2 Basic radiometric and photometric units

The physical properties of electromagnetic radiation are characterized by **radiometric units**. Using radiometric units, we can characterize light in terms of physical quantities; for example, the number of photons, photon energy, and **optical power** (in the lighting community frequently called the **radiant flux**). However, the radiometric units are irrelevant when it comes to light perception by a human being. For example, infrared radiation causes no luminous sensation in the eye. To characterize the light and color sensation by the human eye, **photometric units** are used.

The **luminous intensity**, which is a photometric quantity, represents the light intensity of an optical source, as perceived by the human eye. The luminous intensity is measured in units of **candela** (cd), which is a unit of the International System of Units (SI unit). The present definition of luminous intensity is as follows: *a monochromatic light source emitting an optical power of (1/683) Watt at 555 nm into the solid angle of 1 steradian (sr) has a luminous intensity of 1 candela (cd).*

The unit *candela* has great historical significance. All light intensity measurements can be traced back to the candela. It evolved from an older unit, the **candlepower**, or simply, the **candle**. One candlepower had been defined as the light intensity emitted from a real candle, similar to a plumber's candle. This candle had a specified construction and dimensions:

one standardized candle emits a luminous intensity of 1.0 cd

The luminous intensity of a light source can thus be characterized by giving the number of standardized candles that together emit the same luminous intensity. Note that *candlepower* and *candle* are non-SI units that are rarely used at the present time.

The **luminous flux**, which is also a photometric quantity, represents the light power of a source as perceived by the human eye. The unit of luminous flux is the **lumen** (lm). It is defined as follows: *a monochromatic light source emitting an optical power of (1/683) Watt at 555 nm*

has a luminous flux of 1 lumen (lm). The lumen is an SI unit.

A comparison of the definitions for the candela and lumen reveals that 1 candela equals 1 lumen per steradian or $\text{cd} = \text{lm}/\text{sr}$. Thus, an isotropically emitting light source with luminous intensity of 1 cd has a luminous flux of $4\pi \text{ lm} = 12.57 \text{ lm}$.

The **illuminance** is the luminous flux incident per unit area. The illuminance measured in **lux** ($\text{lux} = \text{lm}/\text{m}^2$). It is an SI unit used when characterizing the illumination in certain environments. Table 11.1 gives typical values of the illuminance in different environments.

Illumination condition	Illuminance
Full moon	1 lux
Street lighting	10 lux
Home lighting	30 to 300 lux
Office desk lighting	100 to 1000 lux
Surgery lighting	10 000 lux
Direct sunlight	100 000 lux

Table 11.1. Typical illuminances in different environments.

Exercise. Photometric units. A 60 W incandescent light bulb has a luminous flux of 1000 lm. Assume that light is emitted isotropically from the bulb.

- What is the luminous efficiency (i.e. the number of lumens emitted per Watt of electrical input power) of the light bulb?
- What number of standardized candles emit the same luminous intensity?
- What is the illuminance, E_{lum} , in units of lux, on a desk located 1.5 m below the bulb?
- Is the illuminance level obtained under (c) sufficiently high for reading?
- What is the luminous intensity, I_{lum} , in units of candela, of the light bulb?
- Derive the relationship between the illuminance at a distance r from the light bulb, measured in lux, and the luminous intensity, measured in candela?
- Derive the relationship between the illuminance at a distance r from the light bulb, measured in lux, and the luminous flux, measured in lumen?
- The definition of the cd involves the optical power of (1/683) W. What, do you suppose, is the origin of this particular power level?

Solution: (a) 16.7 lm/W. (b) 80 candles. (c) $E_{\text{lum}} = 35.4 \text{ lm}/\text{m}^2 = 35.4 \text{ lux}$. (d) Yes.
 (e) $79.6 \text{ lm}/\text{sr} = 79.6 \text{ cd}$. (f) $E_{\text{lum}} r^2 = I_{\text{lum}}$. (g) $E_{\text{lum}} 4\pi r^2 = \Phi_{\text{lum}}$.

(h) Originally, the unit of luminous intensity had been defined as the intensity emitted by a real candle. Subsequently the unit was defined as the intensity of a light source with specified wavelength and optical power. When the power of that light source is (1/683) W, it has the same intensity as the candle. Thus this particular power level has a historical origin and results from the effort to maintain continuity.

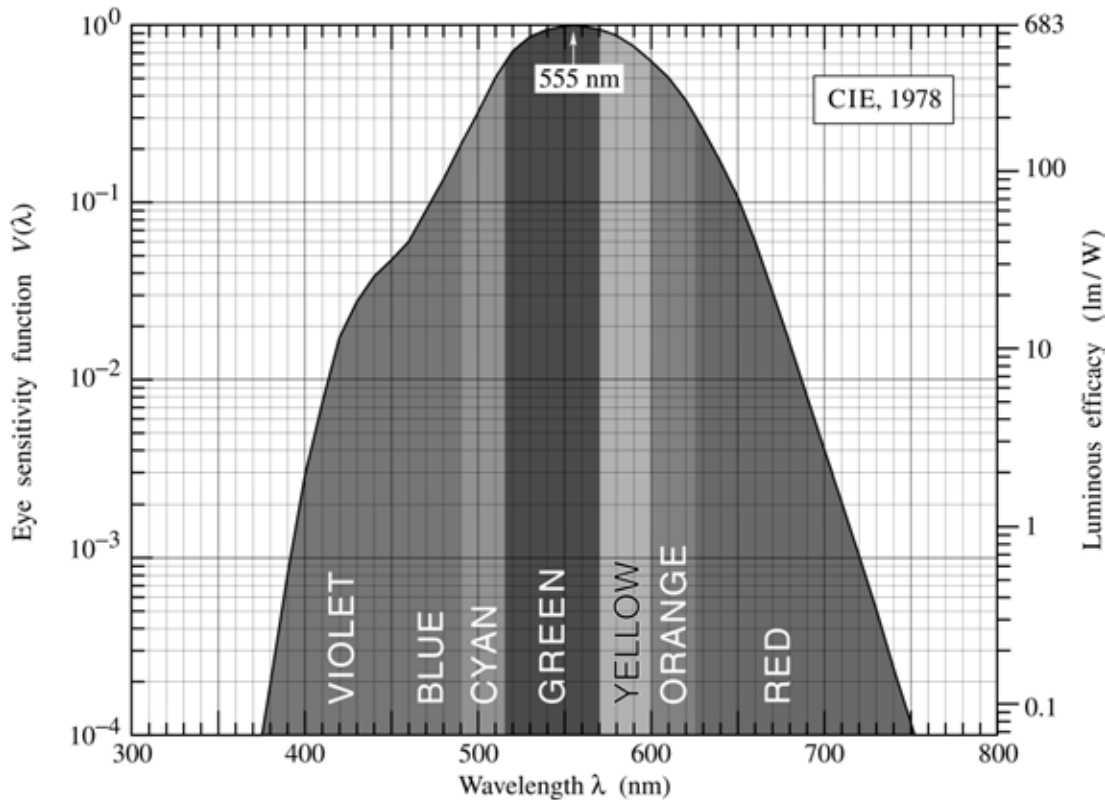


Fig. 11.2. Eye sensitivity function, $V(\lambda)$, (left ordinate) and luminous efficacy, measured in lumens per Watt of optical power (right ordinate). The eye sensitivity is greatest at 555 nm. Also given is a polynomial approximation for $V(\lambda)$ (after 1978 CIE data).

11.3 Eye sensitivity function, luminous efficacy, and luminous efficiency

The conversion between radiometric and photometric units is provided by the *eye sensitivity function*, $V(\lambda)$, shown in Fig. 11.2. The eye has its peak sensitivity in the green at 555 nm, where the eye sensitivity function has a value of unity, i.e. $V(555 \text{ nm}) = 1$. Numerical values of $V(\lambda)$ are tabulated in Appendix 11.1. For wavelengths ranging from 390 to 720 nm, the eye sensitivity function is greater than 10^{-3} . Although the human eye is sensitive to light with wavelengths $< 390 \text{ nm}$ and $> 720 \text{ nm}$, the sensitivity at these wavelengths is extremely low. The relationship between color and wavelength is given in Table 11.2.

Color	Wavelength	Color	Wavelength
Ultraviolet	< 390 nm	Yellow	570 – 600 nm
Violet	390 – 455 nm	Amber	590 – 600 nm
Blue	455 – 490 nm	Orange	600 – 625 nm
Cyan	490 – 515 nm	Red	625 – 720 nm
Green	515 – 570 nm	Infrared	> 720 nm

Table 11.2. Colors and associated typical LED peak wavelength ranges.

The **luminous flux**, Φ_{lum} , is obtained from the radiometric light power using the equation

$$\Phi_{\text{lum}} = 683 \frac{\text{lm}}{\text{W}} \int_{\lambda} V(\lambda) P(\lambda) d\lambda \quad (11.1)$$

where $P(\lambda)$ is the power spectral density, i.e. the light power emitted per unit wavelength, and the prefactor 683 lm/W is a normalization factor. The optical power emitted by a light source is then given by

$$P = \int_{\lambda} P(\lambda) d\lambda . \quad (11.2)$$

High performance visible-spectrum LEDs have a luminous flux of about 1.0 – 5.0 lm at an injection current of 50 – 100 mA.

The **luminous efficacy of optical radiation**, measured in units of lumens per Watt of optical power, is the conversion efficiency from optical power to luminous flux. The luminous efficacy is defined as

$$\text{luminous efficacy} = \frac{\Phi_{\text{lum}}}{P} = \left[683 \frac{\text{lm}}{\text{W}} \int_{\lambda} V(\lambda) P(\lambda) d\lambda \right] / \left[\int_{\lambda} P(\lambda) d\lambda \right] \quad (11.3)$$

Amber light has a higher luminous efficacy than red light. For strictly monochromatic light sources ($\Delta\lambda \rightarrow 0$), the luminous efficacy is equal to the eye sensitivity function $V(\lambda)$ multiplied by 683 lm/W. However, for multicolor light sources and especially for white light sources, the luminous efficacy needs to be calculated by integration over all wavelengths.

The **luminous efficiency of a light source**, also measured in units of lm/W, is the luminous flux of the light source divided by the electrical input power.

$$\text{luminous efficiency} = \Phi_{\text{lum}} / (IV) \quad (11.4)$$

where the product (IV) is the electrical input power of the device (in the lighting community, luminous efficiency is often referred to as *luminous efficacy of the source*).

Inspection of Eqs. (11.3) and (11.4) yields that the luminous efficiency is the product of luminous efficacy and electrical-to-optical power conversion efficiency. The luminous efficiency of common light sources is given in Table 11.3.

Light source		Luminous efficiency
Edison's first light bulb	(a)	1.4 lm/W
Tungsten filament light bulbs	(a)	15 – 20 lm/W
Quartz halogen light bulbs	(a)	20 – 25 lm/W
Fluorescent light tubes and compact bulbs	(b)	50 – 80 lm/W
Mercury vapor light bulbs	(c)	50 – 60 lm/W
Metal halide light bulbs	(c)	80 – 125 lm/W
High-pressure sodium vapor light bulbs	(c)	100 – 140 lm/W

Table 11.3. Luminous efficiencies of different light sources. (a) Incandescent sources. (b) Fluorescent sources. (c) High-intensity discharge (HID) sources.

The luminous efficiency is a highly relevant figure of merit for visible-spectrum LEDs. It is a measure of the perceived light power normalized to the electrical power expended to operate the LED. For light sources with a perfect electrical-power-to-optical-power conversion, the luminous efficiency is equal to the luminous efficacy.

Exercise. Luminous efficacy and luminous efficiency of LEDs. Consider a red and an amber LED emitting at 625 and 590 nm, respectively. For simplicity assume that the emission spectra are monochromatic ($\Delta\lambda \rightarrow 0$). What is the luminous efficacy of the two light sources? Calculate the luminous efficiency of the LEDs, assuming that the red and amber LEDs have an external quantum efficiency of 50 %. Assume that the LED voltage is given by $V = E_g / e = h\nu / e$.

Assume next that the LED spectra are thermally broadened and have a gaussian lineshape with a linewidth of $1.8 kT$. Again calculate the luminous efficacy and luminous efficiency of the two light sources. How accurate are the results obtained with the approximation of monochromaticity?

The **luminance** of an LED is the ratio of the luminous intensity emitted in a certain direction (measured in cd) divided by the projected LED chip area in that direction (measured in cm^2). The luminance is measured in units of cd/cm^2 . In most cases, the direction of interest is normal to the chip surface. In this case, the luminance is the luminous intensity emitted along the chip-normal direction divided by the chip area. It is desirable to maximize luminous intensity and luminous flux while keeping the LED chip area minimal. Thus the luminance is a measure of how efficiently the valuable semiconductor wafer area is used to attain, at a given injection current, a certain luminous intensity.

The *projected LED chip area* mentioned above depends on the angle ϕ according to a cosine law, i.e. the projected area is $A_{\text{projected}} = A_{\text{chip}} \cos \phi$, where ϕ is the angle between the direction considered and the chip surface normal. The luminous intensity of LEDs with lambertian emission pattern also depends on the angle ϕ according to a cosine law. Thus the luminance of lambertian LEDs is a constant, independent of angle.

Some LED structures attain excellent power efficiency by using small light-emitting areas (current injection in a small area of chip) and advanced light-output-coupling structures (see, for example, Schmid *et al.*, 2002). However, such devices have low luminance because only a small fraction of the chip area is injected with current.

Figure of merit	Explanation	Unit
Luminous efficacy	Luminous flux per optical unit power	lm/W
Luminous efficiency	Luminous flux per input electrical unit power	lm/W
Luminous intensity efficiency	Luminous flux per sr per input electrical unit power	cd/W
Luminance	Luminous flux per sr per chip unit area	cd/cm^2
Power efficiency	Optical output power per input electrical unit power	%
Internal quantum efficiency	Photons emitted in active region per electron injected	%
External quantum efficiency	Photons emitted from LED per electron injected	%
Extraction efficiency	Escape probability of photons emitted in active region	%

Table 11.4. Summary of photometric, radiometric, and quantum performance measures for LEDs.

All LEDs have a certain **radiation pattern** or **far-field pattern**. The intensity, measured in W/cm^2 , depends on the longitudinal and azimuth angle and the distance from the LED. The total optical power emitted by the LED is obtained by integration over the area of a sphere.

$$P = \int_A \int_{\lambda} I(\lambda) d\lambda dA \quad (11.5)$$

where $I(\lambda)$ is the **spectral light intensity** (measured in W per nm per cm²) and A is the surface area of the sphere. Integration is carried out over the entire surface area.

Table 11.4 summarizes frequently used figures of merit for light-emitting diodes.

11.4 Color matching functions and chromaticity diagram

Light causes different levels of excitation of the red, green, and blue cones. However, the sensation of color and luminous flux of a particular color varies slightly among different individuals. Furthermore, the sensation of color is a subjective quantity that cannot be measured objectively. For these reasons, *The International Commission for Illumination* (*Commission Internationale de l'Eclairage*, CIE) has *standardized the measurement of color* using **color matching functions** and the **chromaticity diagram**. Note that neither the color matching functions nor the chromaticity diagram are unique (see, for example, Judd, 1951 or Vos, 1978). In fact there have been several different versions of color matching functions and chromaticity diagram.

The CIE color matching functions are shown in Fig. 11.3. The values of the three color matching functions are tabulated in **Appendix 11.1**. The three color matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ approximately correspond to the eye sensitivity curves of the red, green, and blue cones, respectively.

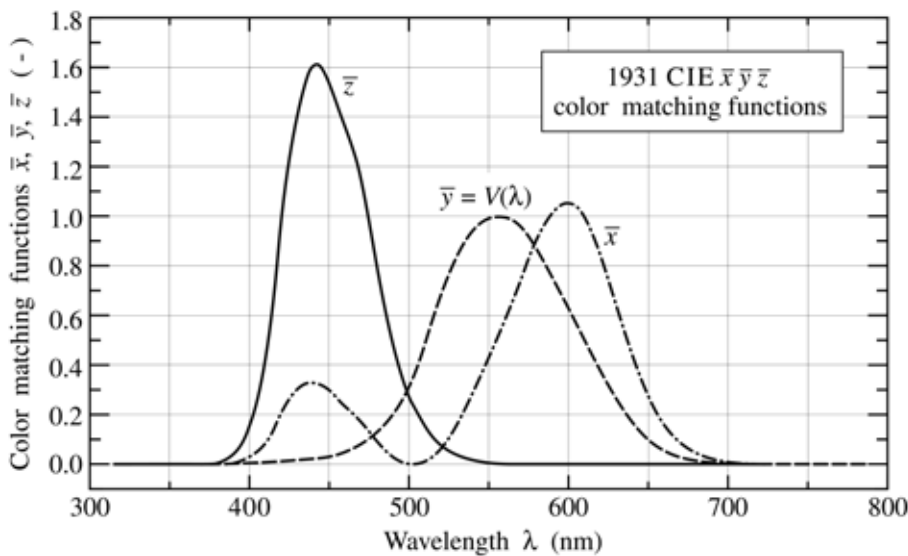


Fig. 11.3. CIE (1931) $\bar{x}\bar{y}\bar{z}$ color matching functions. The \bar{y} color matching function is identical to the eye sensitivity function $V(\lambda)$.

The perception of colored light can be analyzed in terms of the degree to which the light stimulates the three types of cones. The degree of stimulation allows one to characterize *all* possible colors.

The green color matching function has been chosen in such a way that its numerical value is identical to the eye sensitivity function, i.e.

$$\boxed{\bar{y}(\lambda) = V(\lambda)} \quad (11.6)$$

Note that $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$, and $V(\lambda)$ are dimensionless quantities.

The degree of stimulation of the three types of cones is given by

$$X = \int_{\lambda} \bar{x}(\lambda) P(\lambda) d\lambda \quad (11.7)$$

$$Y = \int_{\lambda} \bar{y}(\lambda) P(\lambda) d\lambda \quad (11.8)$$

$$Z = \int_{\lambda} \bar{z}(\lambda) P(\lambda) d\lambda \quad (11.9)$$

where X , Y , and Z are the ***tristimulus values*** that indicate the relative stimulation of each of the three cones. Equations (11.7) – (11.9) suggest that the unit of the tristimulus values is Watt. A prefactor in front of the integral can be included so that the tristimulus values become dimensionless. If only *ratios* of tristimulus values are employed, as below, the prefactors and units cancel and thus become irrelevant.

The ***chromaticity coordinates*** x and y are calculated from the tristimulus values according to

$$x = \frac{X}{X + Y + Z} \quad (11.10)$$

$$y = \frac{Y}{X + Y + Z} . \quad (11.11)$$

Thus, the value of a chromaticity coordinate is the stimulation of one particular cone normalized to the entire stimulation ($X + Y + Z$). The value of the z chromaticity coordinate is calculated analogously, that is

$$z = \frac{Z}{X + Y + Z} = 1 - x - y . \quad (11.12)$$

Note that the z chromaticity value can be obtained from x and y , so that there is no new information provided by the z chromaticity coordinate. Therefore, the z coordinate is redundant and, for this reason, does not need to be used.

The chromaticity diagram is shown in Fig. 11.4. Reddish and greenish colors are found for large values of x and y , respectively. Bluish colors are found for large values of z , which is, according to Eq. (11.12), for low values of x and y , or near the origin of the chromaticity diagram.

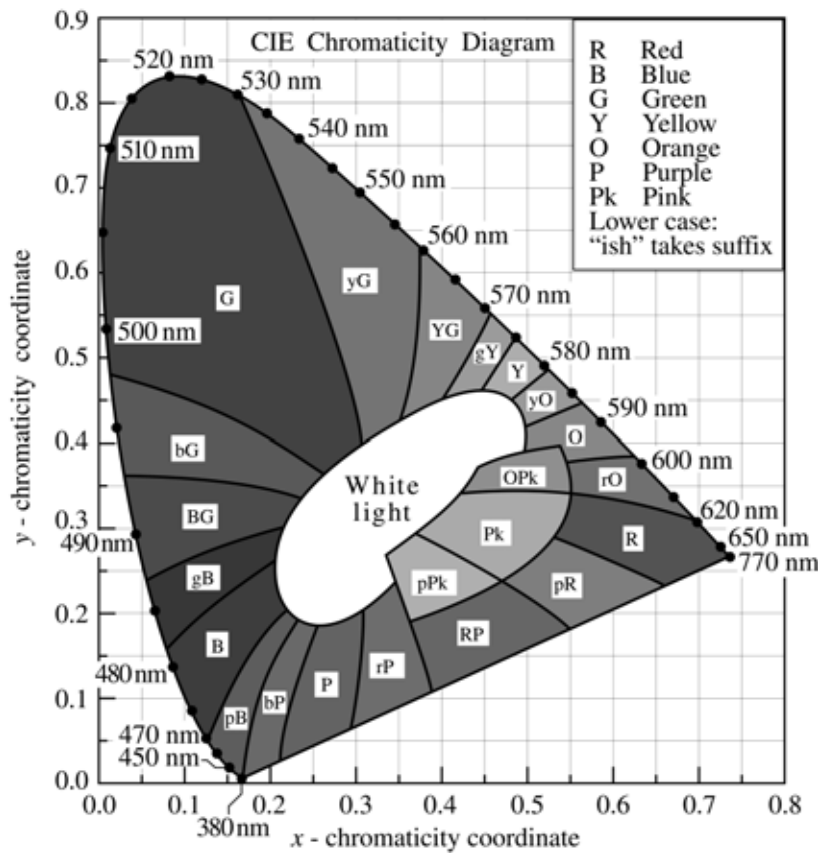


Fig. 11.4. CIE chromaticity diagram. Monochromatic colors are located on the perimeter and white light is located in the center of the diagram (adopted from Gage *et al.*, 1977).

Monochromatic or pure colors are found on the perimeter of the chromaticity diagram. White light is found in the center of the chromaticity diagram. All colors can be characterized in terms of their location in the chromaticity diagram.

There are chromaticity coordinates other than the (x, y) chromaticity coordinates. In 1960, the CIE introduced the (u, v) and in 1976 the (u', v') **uniform chromaticity coordinates**. These coordinates form the **uniform chromaticity diagram**. The uniform chromaticity coordinates are calculated from the tristimulus values according to

$$u = \frac{4X}{X + 15Y + 3Z} \quad v = \frac{6Y}{X + 15Y + 3Z} \quad (\text{CIE, 1960}) \quad (11.13)$$

and

$$u' = \frac{4X}{X + 15Y + 3Z} \quad v' = \frac{9Y}{X + 15Y + 3Z} \quad (\text{CIE, 1976}) . \quad (11.14)$$

The motivation for the introduction of the uniform chromaticity coordinates is as follows: the color differences between two different points in the (x, y) chromaticity diagram is spatially very non-uniform, that is, the color may change much more rapidly in one direction, e.g. the x -direction, compared with the other direction, e.g. the y -direction.

This deficiency of the (x, y) chromaticity diagram is strongly reduced, although not eliminated, in the (u, v) and (u', v') uniform chromaticity diagrams. As a result, the *color difference* between two locations in the uniform chromaticity diagram is directly proportional to the *geometrical distance* between these points.

There are also other methods for characterizing colors. In one such scheme, colors are characterized according to ***hue***, ***saturation***, and ***brightness***. The *hue* gives the *color* of a specific light. The hue can be, for example, red or green. The degree of *saturation* indicates the balance between a single monochromatic color (which has high saturation) and white light (which has low saturation). The color sensation also depends on the *brightness* of the light. Generally the perception of colors decreases or gets totally lost at low light intensities (scotopic vision regime). That is, objects lose their colorful appearance at sufficiently low light levels.

11.5 Color purity

Monochromatic sources ($\Delta\lambda \rightarrow 0$) are located on the perimeter of the chromaticity diagram. However, as the spectral linewidth of a light source gets broader, the color location in the chromaticity diagram moves towards the center of the chromaticity diagram. If the spectral width of a light source becomes comparable to the entire visible range, the light source is *white* and thus located near the center of the chromaticity diagram.

The ***dominant wavelength*** of a test light source is defined as the monochromatic color located on the perimeter of the chromaticity diagram that appears to be closest to the color of the test light source. The dominant wavelength is determined by drawing a straight line between a white reference illuminant through the (x, y) chromaticity coordinate of the test light source to be measured, to the perimeter of the chromaticity diagram. The intersection point is the dominant

wavelength of the light source. The procedure for finding the dominant wavelength is schematically shown in Fig. 11.5.

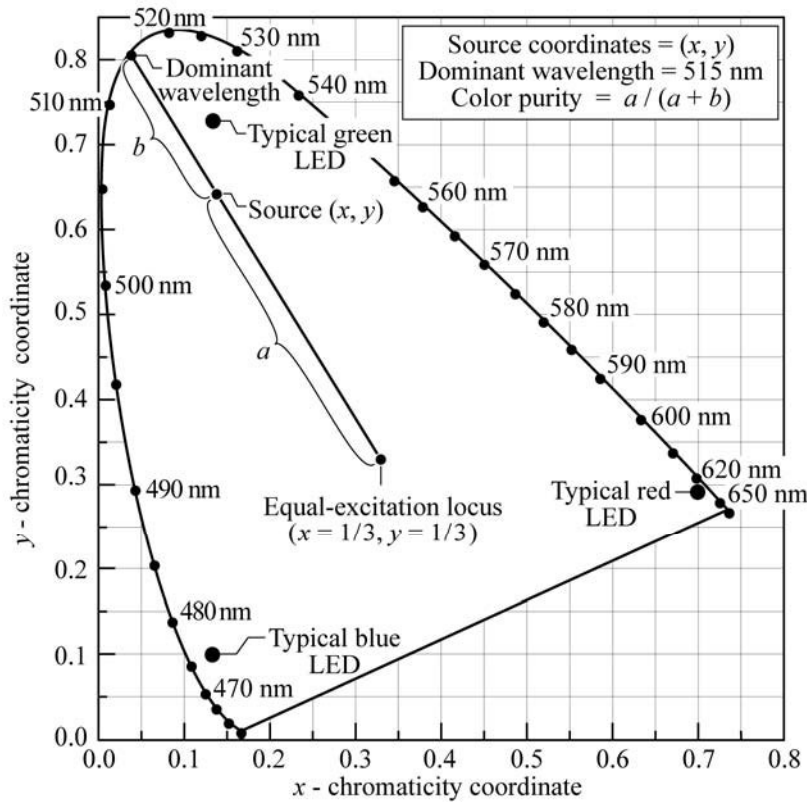


Fig. 11.5. Chromaticity diagram showing the determination of the *dominant wavelength* and *color purity* of a light source with chromaticity coordinates (x, y) using the equal-excitation locus ($x = 1/3, y = 1/3$) as the white-light reference. Also shown are typical locations of blue, green, and red LEDs.

The **color purity** or **color saturation** of a light source is the distance in the chromaticity diagram from the (x, y) color coordinate point to the coordinate of a white reference illuminant normalized by the distance between the reference illuminant to the dominant color. The **equal-excitation** (frequently called **equal-energy**) chromaticity coordinates are used for the white reference illuminant chromaticity coordinates. The color purity is given by

$$\text{color purity} = \frac{a}{a + b} = \frac{\sqrt{(x - x_{ee})^2 + (y - y_{ee})^2}}{\sqrt{(x_d - x_{ee})^2 + (y_d - y_{ee})^2}} \quad (11.15)$$

where a and b are shown in Fig. 11.5 and (x, y) , (x_{ee}, y_{ee}) , and (x_d, y_d) represent the chromaticity coordinates of the light source under test, of the equal-excitation reference illuminant, and of the dominant color, respectively. Thus the color purity is the relative distance of a light source under test from the center of the chromaticity diagram. Generally, the color purity is 100 % for monochromatic light sources ($\Delta\lambda \rightarrow 0$) located on the perimeter of the chromaticity diagram and

0 % for white illuminants.

Also shown in Fig. 11.5 are the typical locations of a GaInN green LED, an AlGaInP red LED, and a GaInN blue LED. The color purity of the red LED is very high, very close to 100 %. The green LED has a lower color purity due to the non-zero spectral width of the LED emission and the strong curvature of the chromaticity diagram in the green region.

11.6 LEDs in the chromaticity diagram

Monochromatic light sources ($\Delta\lambda \rightarrow 0$) are located on the perimeter of the chromaticity diagram. Light emission from LEDs is monochromatic (single color) to the eye but LEDs are not monochromatic in the strict physical sense since LEDs have a spectral linewidth of about $1.8 kT$. Owing to the finite spectral linewidth of LEDs, they are not located on the very perimeter of the chromaticity diagram but are located *close* to the perimeter. When a source emits light distributed over a range of wavelengths, then the chromaticity location moves towards the center of the diagram.

The location of different LEDs on the chromaticity diagram is shown in Fig. 11.6. Inspection of the figure reveals that the location of red and blue LEDs is on the perimeter of the chromaticity diagram. However, blue–green and green LEDs are located off the perimeter closer to the center of the diagram due to the finite linewidth of the emission spectrum and the strong curvature of the chromaticity diagram in the green wavelength range.

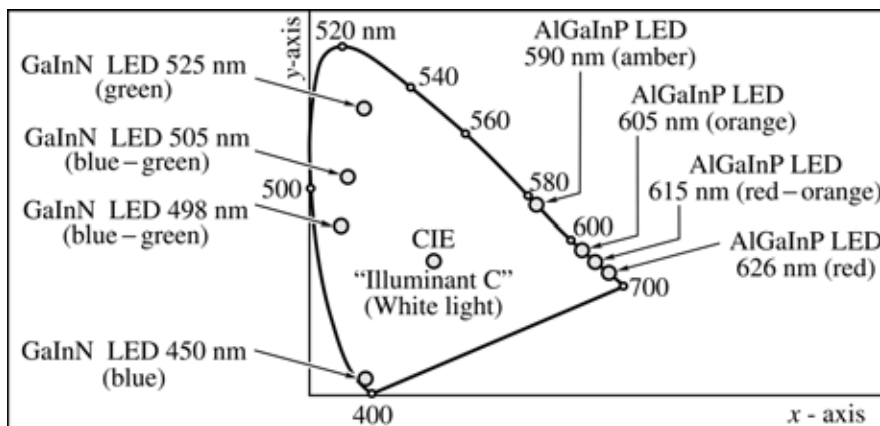


Fig. 11.6. Location of LED light emission on the chromaticity diagram (adopted from Schubert and Miller, 1999).

11.7 White illuminants and color temperature

White light usually has a broad spectrum extending over the entire visible range. The ideal model for white light is sunlight. The Sun's optical spectrum is shown in Fig. 11.7, including the spectrum incident above the Earth's atmosphere, and at sea level. The spectrum of sunlight

extends over the entire visible region. However, the Sun's spectrum depends on the time of day, season, altitude, weather, and other factors.

It is desirable to define an independent standard for white light. The **black-body radiation spectrum** is used as one such standard. The black-body spectrum is characterized by only one parameter, the temperature of the body. The black-body spectrum was first derived by Max Planck (1900) and is given by

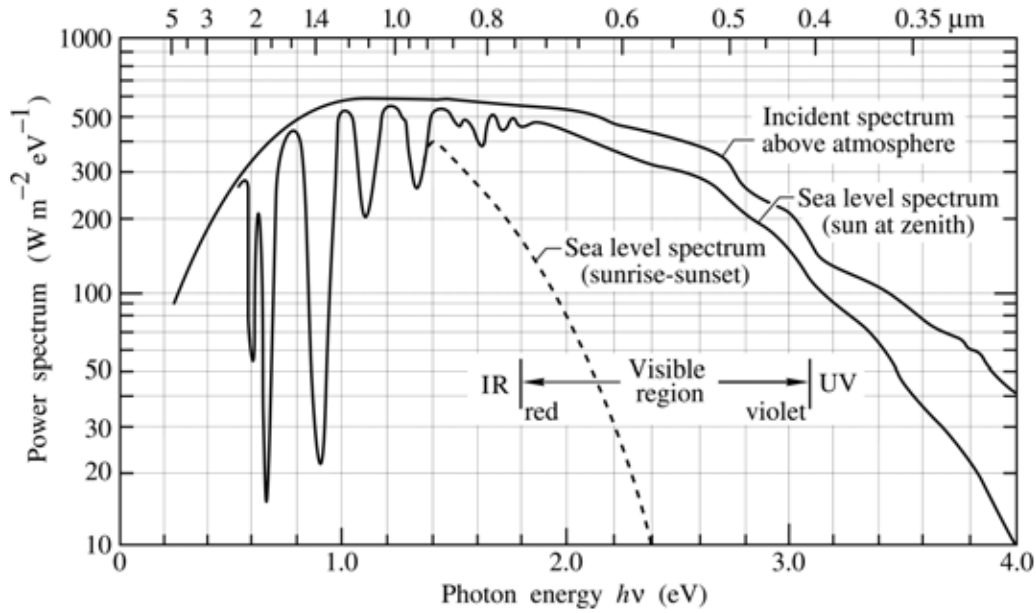


Fig. 11.7. Power spectrum of solar radiation versus photon energy and wavelength for different conditions (adopted from Jackson, 1975).

$$I(\lambda) = \frac{2 h c^2}{\lambda^5 \left[\exp\left(\frac{h c}{\lambda k T}\right) - 1 \right]} \quad (11.16)$$

The maximum intensity of radiation emanating from a black body with temperature T occurs at the wavelength given by **Wien's law**

$$\lambda_{\max} = \frac{2880 \text{ } \mu\text{m K}}{T} . \quad (11.17)$$

At “low” black-body temperatures, e.g. 2000 K, the radiation occurs mostly in the infrared. As the temperature increases, the maximum of the radiation moves into the visible wavelength

range.

The location of the black-body radiation in the chromaticity diagram is shown in Fig. 11.8. As the temperature of the black body increases, the chromaticity location moves from the red wavelength range towards the center of the diagram. Typical black-body temperatures in the white region of the chromaticity diagram range between 2500 and 10 000 K. Also shown in Fig. 11.8 are the locations of several illuminants standardized by the CIE. These standard illuminants include Illuminant A, B, C, and D₆₅.

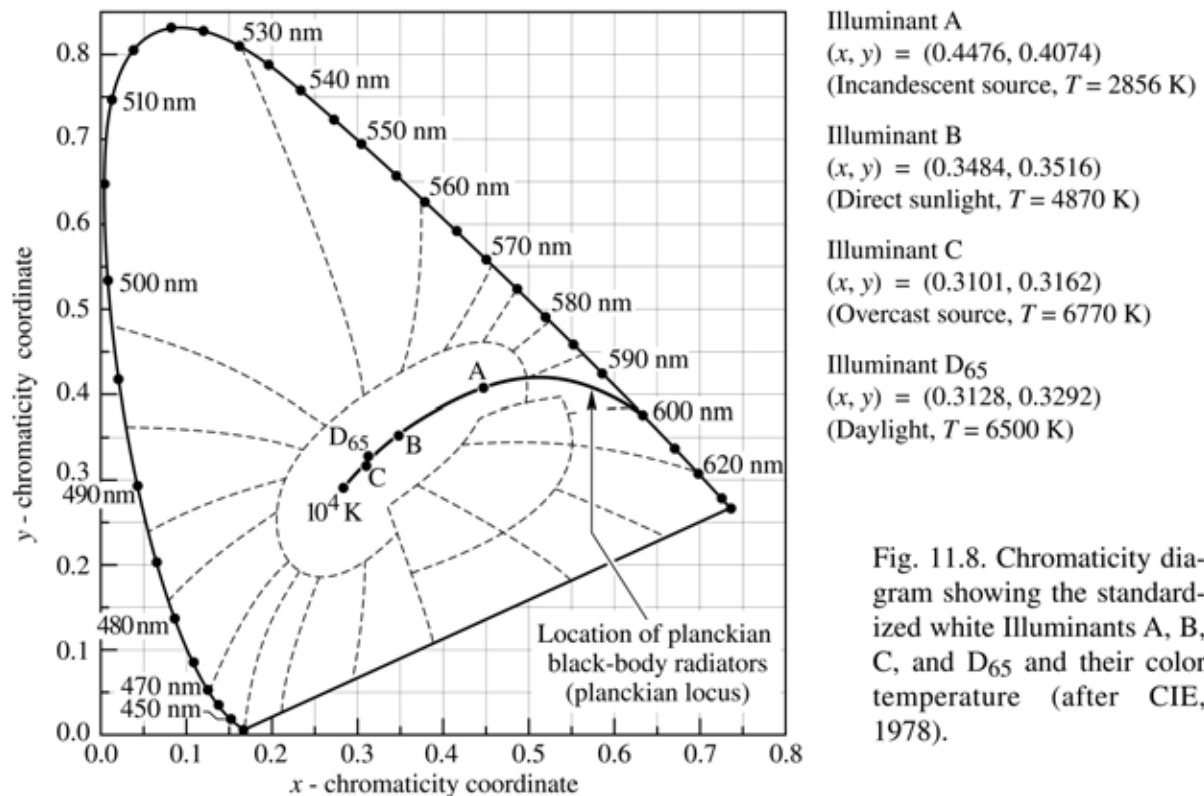


Fig. 11.8. Chromaticity diagram showing the standardized white Illuminants A, B, C, and D₆₅ and their color temperature (after CIE, 1978).

The **color temperature** (CT) of a white light source, given in units of Kelvin, is the temperature of a planckian black-body radiator that has the same chromaticity location as the white light source considered.

If the color of a white light source does not fall on the planckian locus, the **correlated color temperature** (CCT), also given in units of Kelvin, is used. The correlated color temperature of a white light source is defined as the temperature of a planckian black-body radiator, whose color is closest to the color of the white light source.

The correlated color temperature of a light source is determined as follows. On the (u, v) uniform chromaticity diagram, the point on the planckian locus that is *closest* to the chromaticity

location of the light source, is determined. The correlated color temperature is the temperature of the planckian black-body radiator at that point. The determination of the correlated color temperature was discussed in CIE publication No. 17.4 (1987) and by Robertson (1968).

The chromaticity locations of incandescent light sources are very close to, although not exactly on the planckian locus (Ohno, 2001). For such incandescent light sources, the color temperature can be specified. Standard incandescents have color temperatures ranging from 2000 to 2900 K. The common warm incandescent light source has a color temperature of 2800 K. Quartz halogen incandescent lamps have a color temperature ranging from 2800 to 3200 K (Ohno, 1997).

Other light sources, such as metal halide light sources, are further removed from the planckian locus. For such light sources, the correlated color temperature should be used. Bluish white lights have a correlated color temperature of about 8000 K.

Note that white light can be created in several different ways. These ways include the creation of white light by **broad-band** emission, by the employment of a **dichromatic** source (a source emitting two complementary colors), or a **trichromatic** source (a source emitting three colors).

11.8 Additive color mixing

The *combination* or *additive mixing* of two or more light sources is employed in a number of applications. In LED displays, three different types of LEDs, usually emitting in the red, green, and blue, are used. The three colors are mixed so that the observer sees a mixture of the three colors. Another example of color mixing are white-light emitters based on two or three complementary colors.

Next, we determine the chromaticity coordinates of a light source emitting three discrete emission bands. Assume that each emission band is much narrower than any of the three color matching functions. Consider three sources with spectral power density $P_1(\lambda)$, $P_2(\lambda)$, and $P_3(\lambda)$ with peak wavelengths λ_1 , λ_2 , λ_3 . Assume further that the three light sources have the chromaticity coordinates (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) . If the spectral linewidth of the three sources is much narrower than the color matching functions, then the integrals of Eqs. (11.7) – (11.9) can be rewritten as

$$X = \bar{x}(\lambda_1) P_1 + \bar{x}(\lambda_2) P_2 + \bar{x}(\lambda_3) P_3 \quad (11.18)$$

$$Y = \bar{y}(\lambda_1) P_1 + \bar{y}(\lambda_2) P_2 + \bar{y}(\lambda_3) P_3 \quad (11.19)$$

$$Z = \bar{z}(\lambda_1) P_1 + \bar{z}(\lambda_2) P_2 + \bar{z}(\lambda_3) P_3 \quad (11.20)$$

where P_1 , P_2 , and P_3 are the optical powers emitted by the three sources. Using the abbreviations

$$L_1 = \bar{x}(\lambda_1) P_1 + \bar{y}(\lambda_1) P_1 + \bar{z}(\lambda_1) P_1 \quad (11.21)$$

$$L_2 = \bar{x}(\lambda_2) P_2 + \bar{y}(\lambda_2) P_2 + \bar{z}(\lambda_2) P_2 \quad (11.22)$$

$$L_3 = \bar{x}(\lambda_3) P_3 + \bar{y}(\lambda_3) P_3 + \bar{z}(\lambda_3) P_3 \quad (11.23)$$

the chromaticity coordinates of the combined light are calculated according to Eqs. (11.10) and (11.11) and are given by

$$x = \frac{x_1 L_1 + x_2 L_2 + x_3 L_3}{L_1 + L_2 + L_3} \quad (11.24)$$

$$y = \frac{y_1 L_1 + y_2 L_2 + y_3 L_3}{L_1 + L_2 + L_3} . \quad (11.25)$$

Thus, the chromaticity coordinate of the multi-component light is a linear combination of the individual chromaticity coordinates weighted by the L_i factors.

The principle of color mixing is shown in Fig. 11.9. The figure shows the mixing of *two* colors with chromaticity coordinates (x_1, y_1) and (x_2, y_2) . For the case of two colors, $L_3 = P_3 = 0$. The mixed color will be located on the straight line connecting the chromaticity coordinates of the two light sources. Thus *any* color (including white) located between the two chromaticity points can be created by mixing the two colors.

Figure 11.9 also shows the mixing of *three* colors, located in the red, green, and blue region of the chromaticity diagram. The three chromaticity points are typical points for red, green, and blue LEDs. The three points are connected by a dashed line. The region within the dashed line, often referred to as the **color gamut**, represents all colors that can be created by mixing primary colors, e.g. red, green, and blue. The ability to create a great variety of colors is an important quality for displays. It is desirable that the area provided by the three light sources is as large as possible to create brilliant displays.

The color gamut represents the entire range of colors that can be created from a set of primary colors. Color gamuts are polygons positioned in the chromaticity diagram. For the case of *three* primary colors, the color gamut is a *triangle*, as shown in Fig. 11.9. All colors created by

additive mixtures of the vertex points (primary colors) of a gamut, are necessarily located inside the gamut.

The insight now gained on color mixing allows one to understand the location of different LEDs in the chromaticity diagram. The perimeter of the chromaticity diagram in the red spectral region is approximately a *straight line*, so that red LEDs, despite their thermal broadening, are located on the perimeter. In contrast, the perimeter of the chromaticity diagram is strongly curved in the green region, so that green LEDs due to their spectral broadening, are displaced from the perimeter towards the center of the chromaticity diagram.

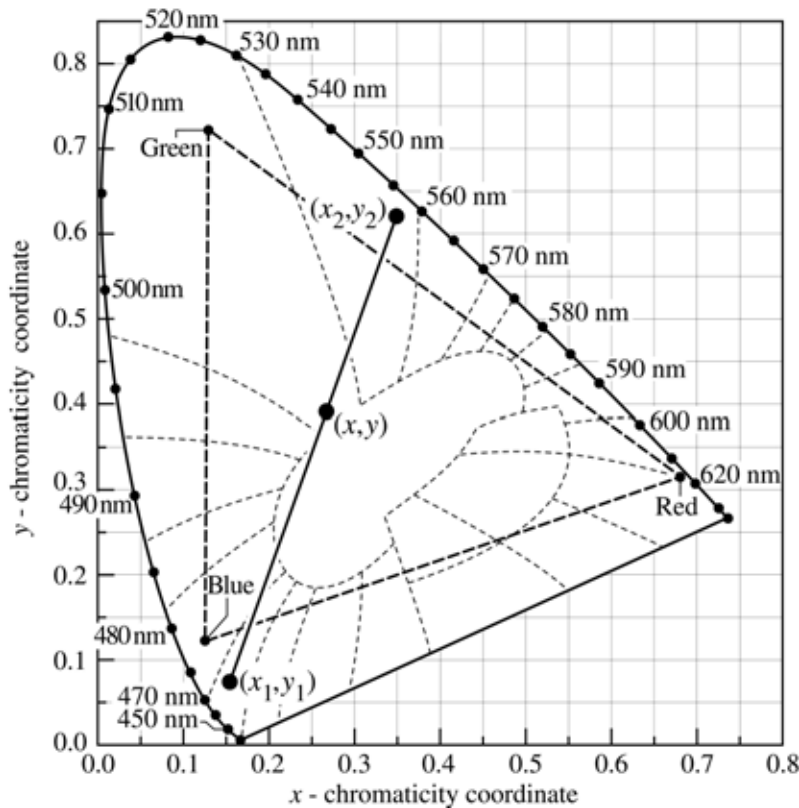


Fig. 11.9. Principle of color mixing illustrated with two light sources with chromaticity coordinates (x_1, y_1) and (x_2, y_2) . The resulting color has the coordinates (x, y) . Also shown is the triangular area of the chromaticity diagram (color gamut) accessible by additive mixing of a red, green, and blue LED.

11.9 Color-rendering index

Another important quality of white light emitters is the *color-rendering index* or *CRI* (Wyszecki and Stiles, 1982; MacAdam, 1993; Berger-Schunn, 1994; CIE, 1995). It is a measure of the ability of an *illuminant*, i.e. an illumination source, to render all the colors of an object illuminated with the light source. The color-rendering characteristics of a white light source are relevant for light sources used for illumination purposes, i.e. when rendering of the colors of an object is important. The color-rendering index is irrelevant for white light sources used in indicator lamp and signage applications.

The color-rendering ability of a *test light source* is evaluated by comparing it with a *reference light source*. The reference light source for calculation of the CRI is a planckian black-body radiator with the same color temperature or correlated color temperature as the test light source, or one of the standardized illuminants. The CIE recommends that the reference light source be a planckian black-body radiator using the correlated color temperature as the black-body temperature (CIE, 1995). Ideally, the test light source and the reference light source have the same chromaticity coordinates and luminous flux.

By definition, the reference light source has ideal color-rendering properties and its color-rendering index is $CRI = 100$. Illuminants other than the chosen reference light source have a color-rendering index lower than 100. Note that the calculation of the CRI depends sensitively on the choice of the reference light source. The selection of the reference light source is therefore critical when comparing the CRIs of different light sources.

Incandescent quartz halogen lamps have one of the best color-rendering properties of artificial light sources. Such lamps are used in locations where color rendering is of prime importance such as in museums, art galleries, and clothing shops. The drawback of quartz halogen lamps is high power consumption.

The CRI calculation has been discussed in detail by Wyszecki and Stiles (1982). The 1976 CIE general CRI is calculated according to

$$CRI_{\text{general}} = \frac{1}{8} \sum_{i=1}^8 CRI_i \quad (11.26)$$

where the CRI_i are the *special color-rendering indices* for eight sample objects. The reflectivity curves of these eight sample objects are shown in Fig. 11.10. The special color-rendering indices are given by

$$CRI_i = 100 - 4.6 \Delta E_i^* \quad (11.27)$$

where ΔE_i^* is the difference in color that occurs when a sample object is illuminated with the reference illumination source and the test illumination source. The special color-rendering index is calculated in such a way that it is 100, if there is no color difference when the sample object is illuminated with the reference source and the test source. At the time Eq. (11.27) was established, the pre-factor 4.6 had been chosen in such a way that the general CRI become equal to about 60 when a “standard warm white” fluorescent lamp was used as a test source and an

incandescent lamp as a reference source. Current fluorescent light sources have higher CRIs, typically in the range 60 – 85 (Kendall and Scholand, 2001).

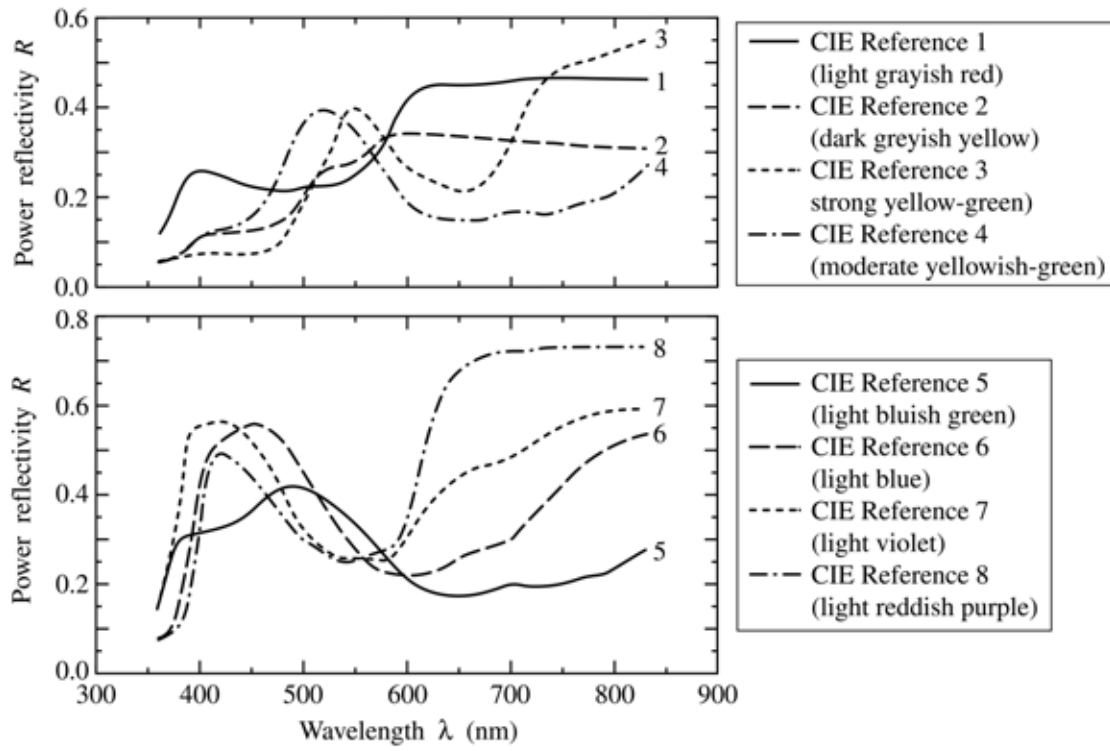


Fig. 11.10. Reflectivity curves of eight sample objects used for the calculation of the general color rendering index (CRI) of light sources used for illumination purposes (after CIE data, 1978).

The color difference is calculated according to

$$\Delta E_i^* = \sqrt{(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2} \quad (11.28)$$

where

$$\Delta L^* = L_{\text{test}}^* - L_{\text{ref}}^* = \left[116 \left(\frac{Y_{\text{test,reflected}}}{Y_{\text{ref,direct}}} \right)^{1/3} - 16 \right] - \left[116 \left(\frac{Y_{\text{ref,reflected}}}{Y_{\text{ref,direct}}} \right)^{1/3} - 16 \right] \quad (11.29)$$

$$\Delta u^* = u_{\text{test}}^* - u_{\text{ref}}^* = 13 L_{\text{test}}^* (u_{\text{test,reflected}} - u_{\text{ref,direct}}) - 13 L_{\text{ref}}^* (u_{\text{ref,reflected}} - u_{\text{ref,direct}}) \quad (11.30)$$

$$\Delta v^* = v_{\text{test}}^* - v_{\text{ref}}^* = 13 L_{\text{test}}^* (v_{\text{test,reflected}} - v_{\text{ref,direct}}) - 13 L_{\text{ref}}^* (v_{\text{ref,reflected}} - v_{\text{ref,direct}}) \quad (11.31)$$

and

$$u = \frac{4 X}{X + 15 Y + 3 Z} \quad \text{and} \quad v = \frac{9 Y}{X + 15 Y + 3 Z} . \quad (11.32)$$

Note that u and v are calculated from the tristimulus values of the reference source spectrum (subscript “ref,direct”), from the reference source spectrum reflected off the sample objects (subscript “ref,reflected”), and from the test source spectrum reflected off the sample objects (subscript “test,reflected”). When calculating the CRI using the equations given above, the chromaticity coordinates and luminous flux of the reference source and the test source should be identical in order to get the highest possible CRI for the test source.

The calculation of the color-rendering index and the choice of the numerical prefactors in Eqs. (11.27) – (11.32) is somewhat arbitrary. These prefactors have been determined in experiments with human subjects. Evidence exists that the current prefactors are not optimum (Wyszecki and Stiles, 1982).

Light source	Color-rendering index
Sunlight	100 (a)
Quartz halogen W filament incandescent light	100 (b)
W filament incandescent light	100 (b)
Fluorescent light	60 – 85 (b)
Trichromatic white light LED	60 – 85 (b, c)
Phosphor-based white LED	55 – 85 (b, c)
Broadened dichromatic white light LED	10 – 60 (b, c)
Hg vapor light coated with phosphor	50 (b)
Hg vapor light	33 (b)
Low and high-pressure Na vapor light	10 and 22 (b)
Green monochromatic light	– 50 (c)

Table 11.5. Color-rendering indices (CRIs) of different light sources. (a) Using sunlight as reference source. (b) Using incandescent light with the same correlated color temperature as the reference light source. (c) Using Illuminant D₆₅ as the reference light source (some data after Kendall and Scholand, 2001).

The color-rendering indices of common light sources are summarized in Table 11.5. The table also includes the color-rendering index of several types of LED sources including dichromatic white LEDs, trichromatic white LEDs, and phosphor-based white LEDs.

Exercise. Color rendering. The color of objects depends strongly on the light source illuminating the objects. Some light sources render the natural colors of an object (true color rendering) and some light sources do not (false color rendering).

- (a) What is the color of a yellow banana when illuminated with a red LED?
- (b) What is the color of a green banana when illuminated with a yellow LED?
- (c) Could it be advantageous for a grocer to illuminate meat with red LEDs, bananas with yellow LEDs, and oranges with orange LEDs?
- (d) Is it possible for two objects of different colors to appear to have the same color under certain illumination conditions?
- (e) Why are low-pressure Na vapor lights used despite their low color-rendering index?
- (f) What would be the advantage and disadvantage of using green LEDs for illumination?

Solution: (a) Red. (b) Yellow. (c) Yes. (d) Yes. (e) Because of high luminous efficiency. (f) High luminous efficacy but low color-rendering properties.

References

- Berger-Schunn A. “Practical color measurement” (John Wiley and Sons, New York, 1994)
- CIE “Commission Internationale de l’Eclairage Proceedings (Cambridge University Press, Cambridge, 1931)
- CIE data of 1931 and 1978 available at < <http://cvision.ucsd.edu> > and < <http://www.cvrl.org> > (1978)
- CIE publication No. 17.4 “International lighting vocabulary”, see < <http://www.cie.co.at> > (CIE, Vienna, Austria, 1987)
- CIE publication No. 13.3 “Method of measuring and specifying color-rendering of light sources”, see < <http://www.cie.co.at> > (CIE, Vienna, Austria, 1995)
- Encyclopedia Britannica, Inc. Illustration of human eye adopted from 1994 edition of the encyclopedia (1994)
- Gage S., Evans D., Hodapp M. W., and Sorensen H. “Optoelectronics applications manual” 1st edition (McGraw Hill, New York, 1977)
- Jackson J. D. “Classical electrodynamics” p. 424 (John Wiley and Sons, New York, 1975)
- Judd D. B. “Report of U. S. Secretariat Committee on Colorimetry and Artificial Daylight” in Proceedings of the 12th Session of the CIE, Vol. 1, p. 11 (Bureau Central de la CIE, Paris, 1951)

- Kendall M. and Scholand M. “Energy savings potential of solid state lighting in general lighting applications” available to the public from National Technical Information Service (NTIS), U. S. Department of Commerce, 5285 Port Royal Road, Springfield, Virginia 22161 (2001)
- MacAdam D. L. (Editor) “Colorimetry – Fundamentals” (SPIE Optical Engineering Press, Bellingham, Washington, 1993)
- Ohno Y. “Photometric standards” Chapter 3 in *OSA/AIP Handbook of Applied Photometry*, 55 (Optical Society of America, Washington DC, 1997)
- Ohno Y. “Photometry and radiometry” Chapter 14 in *OSA Handbook of Optics, Volume III Review for Vision Optics, Part 2, Vision Optics* (McGraw-Hill, New York, 2001)
- Planck M. “On the theory of the law on energy distribution in the normal spectrum (translated from German)” *Verhand. der Deutschen Phys. Gesellschaft* **2**, 237 (1900)
- Robertson R. “Computation of correlated color temperature and distribution temperature” *J. Opt. Soc. Am.* **58**, 1528 (1968)
- Schmid W., Scherer M., Karnutsch C., Plobl A., Wegleiter W., Schad S., Neubert B., and Streubel K. “High-efficiency red and infrared light-emitting diodes using radial outcoupling taper” *IEEE J. Select. Topics Quant. Electronics* **8**, 256 (2002)
- Schubert E. F. and Miller J. N. “Light-emitting diodes – an introduction” *Encyclopedia of Electrical Engineering*, edited by J. G. Webster, Vol. **11**, p. 326 (John Wiley and Sons, New York, 1999)
- Vos J. J. “Colorimetric and photometric properties of a 2-degree fundamental observer” *Color Res. Appl.* **3**, 125 (1978)
- Wyszecki G. and Stiles W. S. “Color science” 2nd edition (John Wiley and Sons, New York, 1982)

Appendix 11.1

Tabulated values of the CIE 1978 color matching functions and eye sensitivity function (after CIE, 1978).

The color matching functions are also called the Judd–Vos-modified CIE 1978 two-degree color matching functions for point sources.

λ	$\bar{x}(\lambda)$ <i>red</i>	$\bar{y}(\lambda) = V(\lambda)$ <i>green</i>	$\bar{z}(\lambda)$ <i>blue</i>			
380 nm	2.6899E-3	2.0000E-4	1.2260E-2	600 nm	1.0550	6.3100E-1
385 nm	5.3105E-3	3.9556E-4	2.4222E-2	605 nm	1.0362	5.6654E-1
390 nm	1.0781E-2	8.0000E-4	4.9250E-2	610 nm	9.9239E-1	5.0300E-1
395 nm	2.0792E-2	1.5457E-3	9.5135E-2	615 nm	9.2861E-1	4.4172E-1
400 nm	3.7981E-2	2.8000E-3	1.7409E-1	620 nm	8.4346E-1	3.8100E-1
405 nm	6.3157E-2	4.6562E-3	2.9013E-1	625 nm	7.3983E-1	3.2052E-1
410 nm	9.9941E-2	7.4000E-3	4.6053E-1	630 nm	6.3289E-1	2.6500E-1
415 nm	1.5824E-1	1.1779E-2	7.3166E-1	635 nm	5.3351E-1	2.1702E-1
420 nm	2.2948E-1	1.7500E-2	1.0658	640 nm	4.4062E-1	1.7500E-1
425 nm	2.8108E-1	2.2678E-2	1.3146	645 nm	3.5453E-1	1.3812E-1
430 nm	3.1095E-1	2.7300E-2	1.4672	650 nm	2.7862E-1	1.0700E-1
435 nm	3.3072E-1	3.2584E-2	1.5796	655 nm	2.1485E-1	8.1652E-2
440 nm	3.3336E-1	3.7900E-2	1.6166	660 nm	1.6161E-1	6.1000E-2
445 nm	3.1672E-1	4.2391E-2	1.5682	665 nm	1.1820E-1	4.4327E-2
450 nm	2.8882E-1	4.6800E-2	1.4717	670 nm	8.5753E-2	3.2000E-2
455 nm	2.5969E-1	5.2122E-2	1.3740	675 nm	6.3077E-2	2.3454E-2
460 nm	2.3276E-1	6.0000E-2	1.2917	680 nm	4.5834E-2	1.7000E-2
465 nm	2.0999E-1	7.2942E-2	1.2356	685 nm	3.2057E-2	1.1872E-2
470 nm	1.7476E-1	9.0980E-2	1.1138	690 nm	2.2187E-2	8.2100E-3
475 nm	1.3287E-1	1.1284E-1	9.4220E-1	695 nm	1.5612E-2	5.7723E-3
480 nm	9.1944E-2	1.3902E-1	7.5596E-1	700 nm	1.1098E-2	4.1020E-3
485 nm	5.6985E-2	1.6987E-1	5.8640E-1	705 nm	7.9233E-3	2.9291E-3
490 nm	3.1731E-2	2.0802E-1	4.4669E-1	710 nm	5.6531E-3	2.0910E-3
495 nm	1.4613E-2	2.5808E-1	3.4116E-1	715 nm	4.0039E-3	1.4822E-3
500 nm	4.8491E-3	3.2300E-1	2.6437E-1	720 nm	2.8253E-3	1.0470E-3
505 nm	2.3215E-3	4.0540E-1	2.0594E-1	725 nm	1.9947E-3	7.4015E-4
510 nm	9.2899E-3	5.0300E-1	1.5445E-1	730 nm	1.3994E-3	5.2000E-4
515 nm	2.9278E-2	6.0811E-1	1.0918E-1	735 nm	9.6980E-4	3.6093E-4
520 nm	6.3791E-2	7.1000E-1	7.6585E-2	740 nm	6.6847E-4	2.4920E-4
525 nm	1.1081E-1	7.9510E-1	5.6227E-2	745 nm	4.6141E-4	1.7231E-4
530 nm	1.6692E-1	8.6200E-1	4.1366E-2	750 nm	3.2073E-4	1.2000E-4
535 nm	2.2768E-1	9.1505E-1	2.9353E-2	755 nm	2.2573E-4	8.4620E-5
540 nm	2.9269E-1	9.5400E-1	2.0042E-2	760 nm	1.5973E-4	6.0000E-5
545 nm	3.6225E-1	9.8004E-1	1.3312E-2	765 nm	1.1275E-4	4.2446E-5
550 nm	4.3635E-1	9.9495E-1	8.7823E-3	770 nm	7.9513E-5	3.0000E-5
555 nm	5.1513E-1	1.0000	5.8573E-3	775 nm	5.6087E-5	2.1210E-5
560 nm	5.9748E-1	9.9500E-1	4.0493E-3	780 nm	3.9541E-5	1.4989E-5
565 nm	6.8121E-1	9.7875E-1	2.9217E-3	785 nm	2.7852E-5	1.0584E-5
570 nm	7.6425E-1	9.5200E-1	2.2771E-3	790 nm	1.9597E-5	7.4656E-6
575 nm	8.4394E-1	9.1558E-1	1.9706E-3	795 nm	1.3770E-5	5.2592E-6
580 nm	9.1635E-1	8.7000E-1	1.8066E-3	800 nm	9.6700E-6	3.7028E-6
585 nm	9.7703E-1	8.1623E-1	1.5449E-3	805 nm	6.7918E-6	2.6076E-6
590 nm	1.0230	7.5700E-1	1.2348E-3	810 nm	4.7706E-6	1.8365E-6
595 nm	1.0513	6.9483E-1	1.1177E-3	815 nm	3.3550E-6	1.2950E-6
				820 nm	2.3534E-6	9.1092E-7
				825 nm	1.6377E-6	6.3564E-7
						1.5476E-10

Appendix 11.2

Reflectivity data of eight reference objects used for the calculation of the color-rendering index (after CIE, 1978).

λ (nm)	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
380 nm	0.219	0.07	0.065	0.074	0.295	0.151	0.378	0.104
385 nm	0.239	0.079	0.068	0.083	0.306	0.203	0.459	0.129
390 nm	0.252	0.089	0.07	0.093	0.31	0.265	0.524	0.17
395 nm	0.256	0.101	0.072	0.105	0.312	0.339	0.546	0.24
400 nm	0.256	0.111	0.073	0.116	0.313	0.41	0.551	0.319
405 nm	0.254	0.116	0.073	0.121	0.315	0.464	0.555	0.416
410 nm	0.252	0.118	0.074	0.124	0.319	0.492	0.559	0.462
415 nm	0.248	0.12	0.074	0.126	0.322	0.508	0.56	0.482
420 nm	0.244	0.121	0.074	0.128	0.326	0.517	0.561	0.49
425 nm	0.24	0.122	0.073	0.131	0.33	0.524	0.558	0.488
430 nm	0.237	0.122	0.073	0.135	0.334	0.531	0.556	0.482
435 nm	0.232	0.122	0.073	0.139	0.339	0.538	0.551	0.473
440 nm	0.23	0.123	0.073	0.144	0.346	0.544	0.544	0.462
445 nm	0.226	0.124	0.073	0.151	0.352	0.551	0.535	0.45
450 nm	0.225	0.127	0.074	0.161	0.36	0.556	0.522	0.439
455 nm	0.222	0.128	0.075	0.172	0.369	0.556	0.506	0.426
460 nm	0.22	0.131	0.077	0.186	0.381	0.554	0.488	0.413
465 nm	0.218	0.134	0.08	0.205	0.394	0.549	0.469	0.397
470 nm	0.216	0.138	0.085	0.229	0.403	0.541	0.448	0.382
475 nm	0.214	0.143	0.094	0.254	0.41	0.531	0.429	0.366
480 nm	0.214	0.15	0.109	0.281	0.415	0.519	0.408	0.352
485 nm	0.214	0.159	0.126	0.308	0.418	0.504	0.385	0.337
490 nm	0.216	0.174	0.148	0.332	0.419	0.488	0.363	0.325
495 nm	0.218	0.19	0.172	0.352	0.417	0.469	0.341	0.31
500 nm	0.223	0.207	0.198	0.37	0.413	0.45	0.324	0.299
505 nm	0.225	0.225	0.221	0.383	0.409	0.431	0.311	0.289
510 nm	0.226	0.242	0.241	0.39	0.403	0.414	0.301	0.283
515 nm	0.226	0.253	0.26	0.394	0.396	0.395	0.291	0.276
520 nm	0.225	0.26	0.278	0.395	0.389	0.377	0.283	0.27
525 nm	0.225	0.264	0.302	0.392	0.381	0.358	0.273	0.262
530 nm	0.227	0.267	0.339	0.385	0.372	0.341	0.265	0.256
535 nm	0.23	0.269	0.37	0.377	0.363	0.325	0.26	0.251
540 nm	0.236	0.272	0.392	0.367	0.353	0.309	0.257	0.25
545 nm	0.245	0.276	0.399	0.354	0.342	0.293	0.257	0.251
550 nm	0.253	0.282	0.4	0.341	0.331	0.279	0.259	0.254
555 nm	0.262	0.289	0.393	0.327	0.32	0.265	0.26	0.258
560 nm	0.272	0.299	0.38	0.312	0.308	0.253	0.26	0.264
565 nm	0.283	0.309	0.365	0.296	0.296	0.241	0.258	0.269
570 nm	0.298	0.322	0.349	0.28	0.284	0.234	0.256	0.272
575 nm	0.318	0.329	0.332	0.263	0.271	0.227	0.254	0.274
580 nm	0.341	0.335	0.315	0.247	0.26	0.225	0.254	0.278
585 nm	0.367	0.339	0.299	0.229	0.247	0.222	0.259	0.284
590 nm	0.39	0.341	0.285	0.214	0.232	0.221	0.27	0.295
595 nm	0.409	0.341	0.272	0.198	0.22	0.22	0.284	0.316
600 nm	0.424	0.342	0.264	0.185	0.21	0.22	0.302	0.348
605 nm	0.435	0.342	0.257	0.175	0.2	0.22	0.324	0.384
610 nm	0.442	0.342	0.252	0.169	0.194	0.22	0.344	0.434

615 nm	0.448	0.341	0.247	0.164	0.189	0.22	0.362	0.482
620 nm	0.45	0.341	0.241	0.16	0.185	0.223	0.377	0.528
625 nm	0.451	0.339	0.235	0.156	0.183	0.227	0.389	0.568
630 nm	0.451	0.339	0.229	0.154	0.18	0.233	0.4	0.604
635 nm	0.451	0.338	0.224	0.152	0.177	0.239	0.41	0.629
640 nm	0.451	0.338	0.22	0.151	0.176	0.244	0.42	0.648
645 nm	0.451	0.337	0.217	0.149	0.175	0.251	0.429	0.663
650 nm	0.45	0.336	0.216	0.148	0.175	0.258	0.438	0.676
655 nm	0.45	0.335	0.216	0.148	0.175	0.263	0.445	0.685
660 nm	0.451	0.334	0.219	0.148	0.175	0.268	0.452	0.693
665 nm	0.451	0.332	0.224	0.149	0.177	0.273	0.457	0.7
670 nm	0.453	0.332	0.23	0.151	0.18	0.278	0.462	0.705
675 nm	0.454	0.331	0.238	0.154	0.183	0.281	0.466	0.709
680 nm	0.455	0.331	0.251	0.158	0.186	0.283	0.468	0.712
685 nm	0.457	0.33	0.269	0.162	0.189	0.286	0.47	0.715
690 nm	0.458	0.329	0.288	0.165	0.192	0.291	0.473	0.717
695 nm	0.46	0.328	0.312	0.168	0.195	0.296	0.477	0.719
700 nm	0.462	0.328	0.34	0.17	0.199	0.302	0.483	0.721

12

White-light LEDs

LEDs can be highly efficient light sources. External quantum efficiencies and electrical-to-optical power conversion efficiencies exceeding 50 % have been demonstrated. As the trend of higher efficiencies in LEDs continues, the number of applications increases as well. A highly interesting new application is *general illumination* where LEDs compete with conventional illumination sources such as incandescent and fluorescent lamps. White LEDs can certainly be used for signage and display applications but they are potentially also suited for general illumination applications. For a comprehensive introduction to lighting technology using solid-state sources, see Zukauskas *et al.* (2002).

12.1 Generation of white light

Light is perceived as *white* light if the three types of cones located on the retina of the human eye are excited in a certain ratio. For the case of white light, the tristimulus values have values such that the location of the color is near the center of the chromaticity diagram.

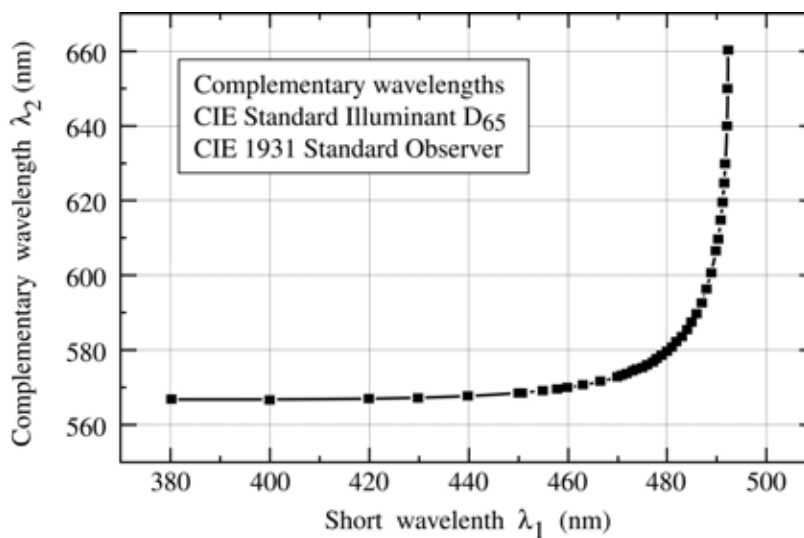


Fig. 12.1. Monochromatic complementary wavelengths resulting in the perception of white light at a certain power ratio (after Wyszecki and Stiles, 1982).

White light can be generated in several different ways. One way of generating white light is the use of two monochromatic colors, called *complementary colors*. Two complementary colors, at a certain power ratio, result in tristimulus values that cause the perception of white light. The wavelengths of complementary colors are shown in Fig. 12.1.

The numerical values for monochromatic complementary wavelengths are given in Table 12.1. Also given in the table is the power ratio required to attain the same chromaticity coordinate as Illuminant D₆₅.

Complementary wavelengths		Power ratio	Complementary wavelengths		Power ratio
λ_1 (nm)	λ_2 (nm)	$P(\lambda_2) / P(\lambda_1)$	λ_1 (nm)	λ_2 (nm)	$P(\lambda_2) / P(\lambda_1)$
380	560.9	0.000642	460	565.9	1.53
390	560.9	0.00955	470	570.4	1.09
400	561.1	0.0785	475	575.5	0.812
410	561.3	0.356	480	584.6	0.562
420	561.7	0.891	482	591.1	0.482
430	562.2	1.42	484	602.1	0.440
440	562.9	1.79	485	611.3	0.457
450	564.0	1.79	486	629.6	0.668

Table 12.1. Wavelengths λ_1 and λ_2 of monochromatic complementary colors with respect to CIE Illuminant D₆₅ and the CIE 1964 Standard Observer. Also given is the required power ratio. The Illuminant D₆₅ has the chromaticity coordinates $x_{D65} = 0.3138$ and $y_{D65} = 0.3310$ (after Wyszecki and Stiles, 1982).

Another way of generating white light is by the emission of *three* colors. If these three colors are emitted at certain wavelengths and with a certain power ratio, the mixed color is perceived as white light. Still another way to generate white light is by a broad-band emitter that emits over the entire visible spectrum. Such an emitter approximates the Sun's spectrum and is perceived as white. There are still other ways to generate white light, for example by the mixture of a broad-band and discrete-line emitters.

Most white-light emitters use an LED emitting at short wavelength and a **wavelength converter**. Some or all of the light emitted by the LED is absorbed in the converter material and then re-emitted as light with a longer wavelength. As a result, the LED emits at least two different wavelengths. The types and characteristics of wavelength converter materials will be discussed below.

The possibility that white light can be generated in different ways raises the question as to

which is the optimum way to generate white light? There are two parameters that need to be considered. *First*, the luminous efficiency and *secondly*, the color rendering index. For *signage applications*, the luminous efficiency is of primary importance and the color rendering index is irrelevant. For *illumination applications*, both the luminous efficiency and the color rendering index are important.

White-light sources employing two monochromatic complementary colors result in the highest possible luminous efficacy. However, the color-rendering index of such a dichromatic light source is lower than that of broad-band emitters.

The maximum luminous efficacy of radiation, attainable for white light created by two complementary monochromatic colors was calculated by MacAdam (1950). MacAdam showed that luminous efficacies exceeding 400 lm/W can be attained using a dichromatic source for white-light generation. The work of MacAdam (1950) was further refined by Ivey (1963) and Thornton (1971). These authors showed that dichromatic white-light sources have high luminous efficacy but low color-rendering properties, making them perfectly suitable for display applications but unsuitable for daylight illumination applications. In addition, Thornton (1971) showed that trichromatic white-light sources, i.e. sources creating white light by additive mixing of three discrete colors, have a color-rendering index suited for most applications. Thornton reported on an experiment in which 60 observers judged the color rendition of meat, vegetables, flowers, complexions, etc., when illuminated with a trichromatic light source with peak wavelengths at 450, 540, and 610 nm. The color rendition in this experiment was found to be “very good if not excellent” illustrating the suitability of trichromatic white-light sources as potent daylight illumination sources.

A white-light source simulating the Sun’s spectrum possesses excellent color rendering properties but the efficacy of such a light source is lower than that of a dichromatic white LED.

12.2 Wavelength converter materials

There are several types of converter materials including *phosphors*, *semiconductors*, and *dyes*. Converter materials have several parameters of interest, including the absorption wavelength, the emission wavelength, and the quantum efficiency. A good converter has near 100 % quantum efficiency. The overall *energy efficiency* of a wavelength converter is given by

$$\eta = \eta_{\text{ext}} (\lambda_1 / \lambda_2) \quad (12.1)$$

where η_{ext} is the external quantum efficiency of the converter. Even if the external quantum efficiency is unity, there is always energy loss associated with the wavelength conversion process, so that the energy conversion efficiency of a wavelength converter is always less than unity.

The most common wavelength converter materials are **phosphors**. Phosphors consist of an inorganic host material doped with an optically active element. A common host is yttrium aluminum garnet (YAG) with the chemical formula $\text{Y}_3\text{Al}_5\text{O}_{12}$. Phosphors having YAG as a host are called **YAG phosphors**. The optically active dopant is a rare-earth element, a rare-earth oxide, or another rare-earth compound. Most rare-earth elements are optically active and well-known rare-earth light emitters include cerium (Ce) used in white-light YAG phosphors, neodymium (Nd) used in lasers (Nd-doped YAG lasers), erbium (Er) used in optical amplifiers, and thorium (Th) oxide used in the mantle of gaslights.

The optical absorption and emission spectrum of a commercial phosphor is shown in Fig. 12.2. The phosphor displays an absorption band and a lower-energy emission band. The emission band is rather broad, making this particular phosphor suitable for white light emission. Phosphors are very stable materials and can have quantum efficiencies close to 100 %. A common phosphor used for white LEDs is Ce-doped YAG phosphor (Nakamura and Fasol, 1997). For Ce-doped phosphors, quantum efficiencies of 75 % have been reported (Schlotter *et al.*, 1999).

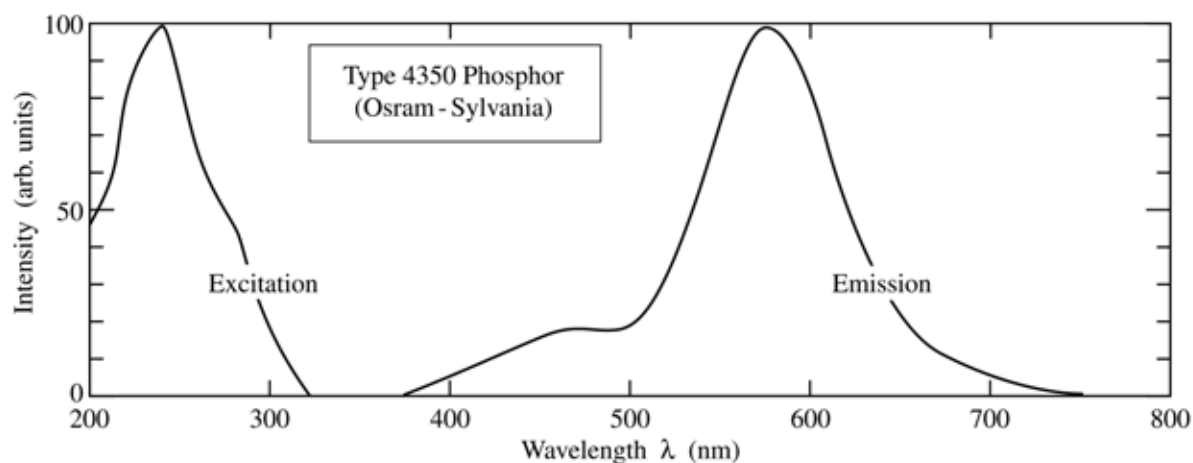


Fig. 12.2. Absorption and emission spectrum of a commercial phosphor (after Osram-Sylvania, 2000).

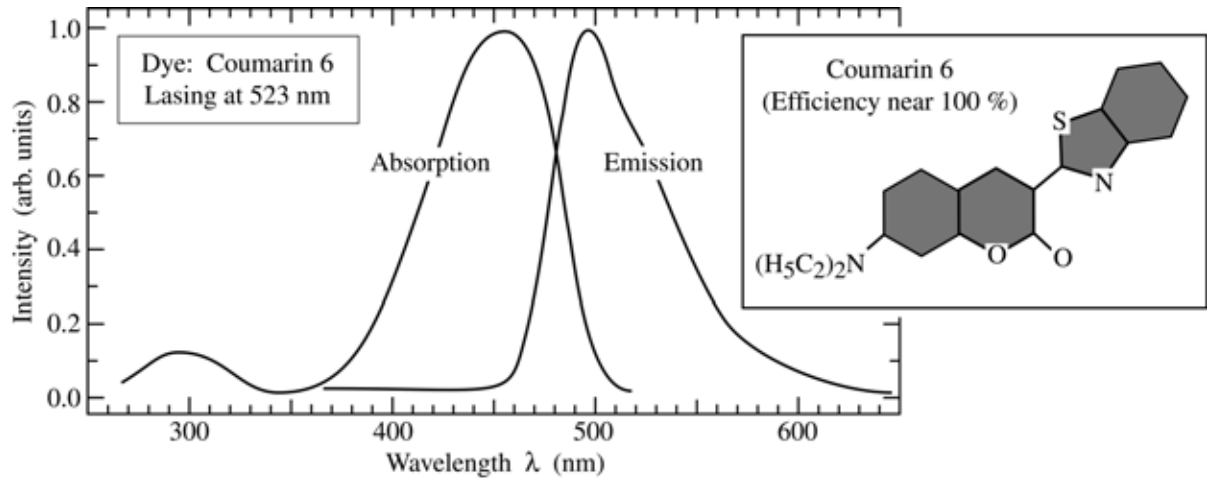


Fig. 12.3. Absorption and emission spectrum of the commercial dye “Coumarin 6”. The inset shows the chemical structure of the dye molecule.

Dyes are another type of wavelength converter. Many different dyes are commercially available. An example of a dye optical absorption and emission spectrum is shown in Fig. 12.3. Dyes can have quantum efficiencies close to 100 %. However, dyes as organic molecules lack the long-term stability afforded by phosphors and semiconductors.

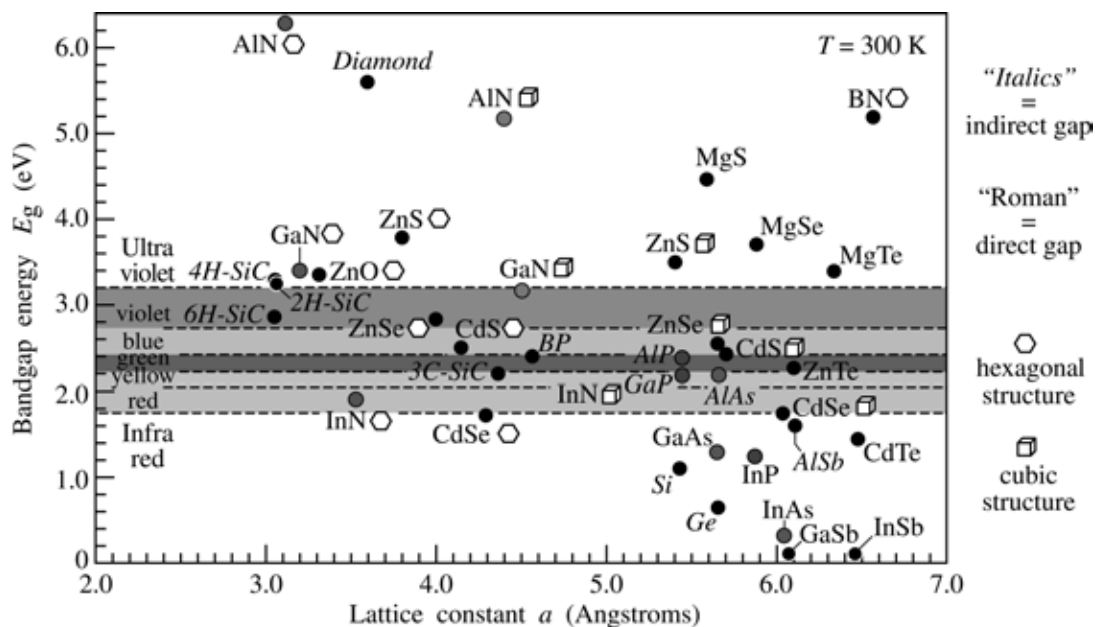


Fig. 12.4. Room-temperature bandgap energy versus lattice constant of common elemental and binary compound semiconductors.

Finally, **semiconductors** are another type of wavelength converter. Semiconductors are characterized by narrow emission lines with linewidths of the order of $2 kT$. The spectral

emission linewidth of semiconductors is narrower than the linewidth of many phosphors and dyes (see Figs. 12.2 and 12.3). Thus, semiconductors allow one to tailor the emission spectrum of a semiconductor wavelength converter with good precision.

As for phosphors and dyes, semiconductors can have internal quantum efficiencies near 100 %. The light escape problem in semiconductor converters is less severe than it is in LEDs due to the fact that the semiconductor converter does not need contacts that could block the light.

As for phosphors and dyes, there are a great variety of semiconductors available. Figure 12.4 shows elemental and binary compound semiconductors versus the semiconductor lattice constant. Using ternary or quaternary alloys, wavelength converters operating at virtually any visible wavelength can be fabricated.

12.3 White LEDs based on phosphor converters

White LEDs using a blue GaInN/GaN LED and a phosphor wavelength converter have been first demonstrated by Bando *et al.* (1996) using a GaInN/GaN LED as reported by Nakamura *et al.* (1995). These white LEDs were reviewed by Nakamura and Fasol (1997). The phosphor used as a wavelength converter is Ce-doped YAG with chemical formula $(Y_{1-a}Gd_a)_3(Al_{1-b}Ga_b)_5O_{12} : Ce$. The exact chemical composition of the host (YAG) and the dopants (e.g. Ce) are usually proprietary and not publicly available.

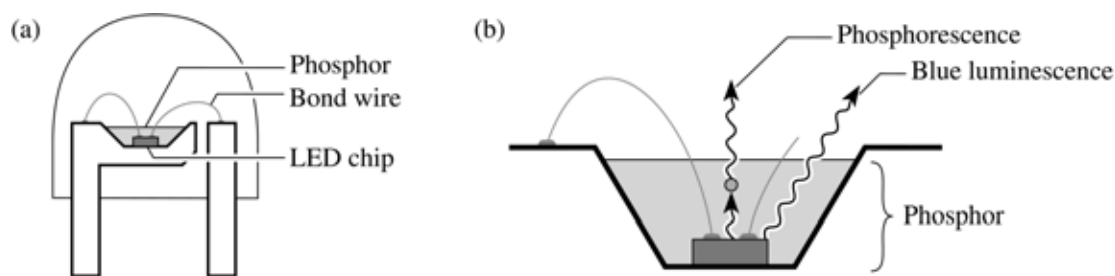


Fig. 12.5. (a) Structure of white LED consisting of a GaInN blue LED chip and a phosphor-containing epoxy encapsulating the semiconductor die. (b) Wavelength-converting phosphorescence and blue luminescence (after Nakamura and Fasol, 1997).

The cross-sectional structure of a white LED is shown in Fig. 12.5 (a). The figure shows the LED die emitting in the blue and the YAG phosphor surrounding the die. The YAG phosphor can be made as a powder and suspended in epoxy resin. During the manufacturing process, a droplet of the YAG phosphor suspended in the epoxy is deposited on the LED die, so that the resin fills the cup-shaped depression in which the LED die is located, as shown in Fig. 12.5 (b).

As indicated in the figure, a fraction of the blue light is absorbed by the phosphor and re-emitted as longer-wavelength light.

The emission spectrum of the phosphor-based white LED thus consists of the blue emission band originating from the semiconductor LED and longer-wavelength phosphorescence, as shown in Fig. 12.6. The thickness of the phosphor-containing epoxy and the concentration of the phosphor suspended in the epoxy determine the relative strengths of the two emission bands. The two bands can thus be adjusted to optimize the luminous efficiency and the color-rendering characteristics of the LED.

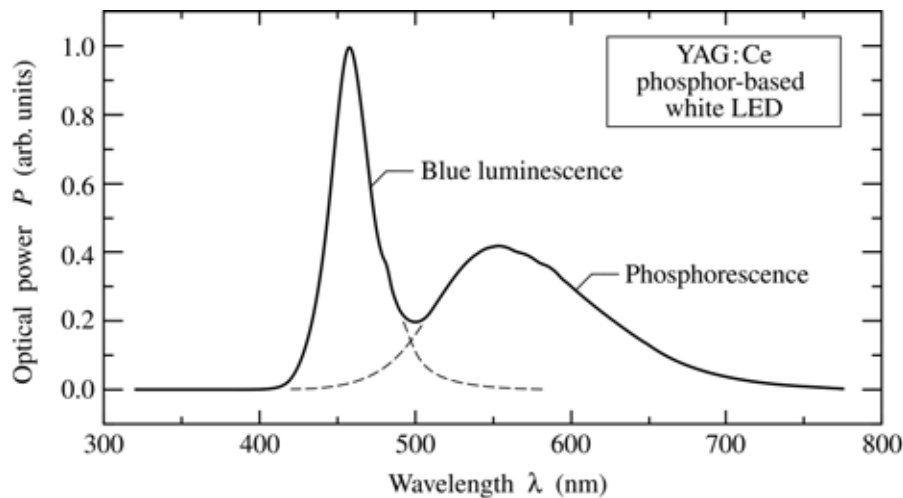


Fig. 12.6. Emission spectrum of a phosphor-based white LED manufactured by the Nichia Chemical Industries Corporation (Anan, Tokushima, Japan).

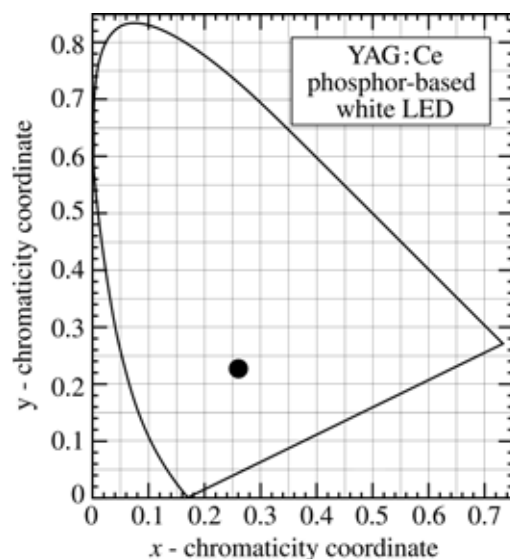


Fig. 12.7. Chromaticity coordinates of a commercial phosphor-based white LED manufactured by Nichia Chemical Industries Corporation (Anan, Tokushima, Japan).

The location of the white LED in the chromaticity diagram is shown in Fig. 12.7. The location suggests that the emission color is white with a bluish tint. This contention is confirmed when looking at the LED. The light emitted from the LED has a bluish appearance.

There are several commercial entities developing white LEDs for illumination applications. These companies include LumiLeds, a joint venture of the Phillips and Agilent Corporations, GELcore, a joint venture of the Emcore and GE Lighting Corporations, Nichia Corporation (Japan), Osram Opto Semiconductors, Cree Lighting Corporation and other companies.

12.4 UV-pumped phosphor-based white LEDs

White LEDs can also be fabricated by optical excitation of phosphor in the ultraviolet (UV) wavelength range (Karlicek, 1999). LEDs emitting in the UV are required for such white LEDs. The development of AlGaInN-based LEDs emitting in the UV is challenging due to the well-known problems of high-Al content AlGaN LEDs such as the lower efficiency compared with GaInN/GaN structures and the low p-type doping efficiency in AlGaN.

In UV-pumped white LEDs, the entire visible emission originates in the phosphor. Phosphors suitable for such LEDs are similar to the phosphors used in fluorescent light tubes. The phosphors in fluorescent lights are pumped by the UV emission of Hg vapor discharge occurring in the tube. These phosphors are readily available and the color rendering properties of such phosphors are suitable for many applications.

A white LED using a UV AlGaInN LED pump source and a tricolor phosphor blend was reported by Kaufmann *et al.* (2001). The LED pump source emitted at 380 – 400 nm, that is, near the boundary between the visible and UV spectrum. The phosphor blend consisted of three phosphors emitting in the red, green, and blue part of the spectrum. A color-rendering index of 78 was reported for the LED.

The advantages of UV-pumped phosphor-based white LEDs are the availability of phosphors with high color rendering properties. The color-rendering index (CRI) of these phosphors typically ranges between 60 and 80. Furthermore, such LEDs exhibit independence of the emission spectrum on the exact emission wavelength of the UV pump, since the visible emission is solely due to the phosphor. Consequently, UV-pumped white LEDs are expected to have a highly reproducible optical spectrum so that “binning” of such LEDs will probably not be required.

A fundamental disadvantage of UV-pumped white LEDs is the relatively high energy loss incurred when converting UV light to white light. The luminous efficiency of UV-pumped white

LEDs is therefore markedly lower than white LEDs with excitation in the blue wavelength range.

12.5 White LEDs based on semiconductor converters (PRS-LED)

LEDs using semiconductor wavelength converters have been demonstrated by Guo *et al.* (1999). The schematic structure of the photon-recycling semiconductor LED (PRS-LED) is shown in Fig. 12.8. The figure indicates that a fraction of the light emitted by the blue GaInN LED is absorbed by a AlGaInP secondary active region and re-emitted (or “recycled”) as lower-energy photons. In order to obtain white light, the intensity of the two light sources must have a certain ratio that will be calculated below. The schematic power budget of the device is shown in Fig. 12.9. It is assumed that the electrical input power is P_0 , and the output powers in the blue and amber spectral range are P_1 and P_2 , respectively. The power conversion efficiency of the blue LED and the photon-recycling semiconductor are assumed to be η_1 and η_2 , respectively. The calculation of efficiency and luminous efficiency of the device is given below.

The energy loss occurring in the photon recycling process must be taken into account when determining the optimum choice of wavelengths for highest efficiency. Note that energy is lost even if the recycling process occurs with unit quantum efficiency. To calculate the optimum wavelength of operation, we represent white light by the Illuminant C standard, for which the chromaticity coordinates are $x_c = 0.3101$, $y_c = 0.3163$, $z_c = 0.3736$. Using these chromaticity coordinates, the pairs of complementary wavelengths can be determined (see Fig. 12.1 or Wyszecki and Stiles, 1982).

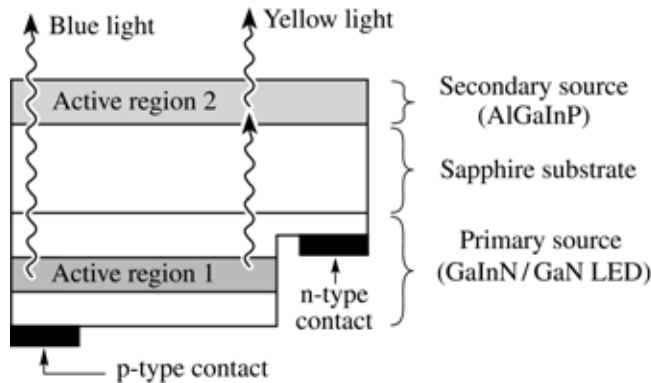


Fig. 12.8. Schematic structure of a photon-recycling semiconductor LED with one current-injected active region (Active region 1) and one optically excited active region (Active region 2) (after Guo *et al.*, 1999).

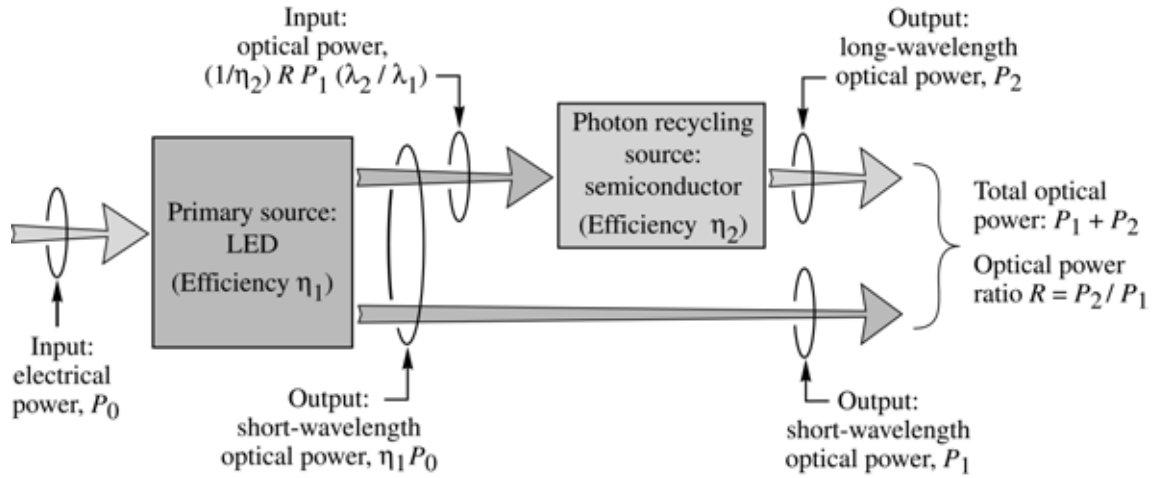


Fig. 12.9. Photon-recycling semiconductor LED power budget with electrical input power P_0 and optical output power P_1 and P_2 .

12.6 Calculation of the power ratio of PRS-LED

Next, we calculate the light power ratio between two sources required for the emission of white light and the luminous efficiency of the photon-recycling semiconductor LED. We refer to λ_1 and λ_2 as the primary (short) and secondary (long) wavelength, respectively. For white light, λ_1 and λ_2 are pairs of complementary wavelengths. We define the color masses of the two light sources as

$$m_1 = \bar{x}_1 + \bar{y}_1 + \bar{z}_1 \quad \text{and} \quad m_2 = \bar{x}_2 + \bar{y}_2 + \bar{z}_2 \quad (12.2)$$

where \bar{x}_1 , \bar{y}_1 , \bar{z}_1 , \bar{x}_2 , \bar{y}_2 , and \bar{z}_2 are color matching functions at the two emission wavelengths λ_1 and λ_2 , respectively (Judd, 1951; Vos, 1978; MacAdam, 1950, 1985). We define the power ratio of the two light sources as

$$R = P_2 / P_1 \quad (12.3)$$

where P_1 and P_2 are the optical powers of the short-wavelength source (λ_1) and the long-wavelength source (λ_2), respectively. The chromaticity coordinates of the newly generated color are then given by

$$y_{\text{new}} = \frac{P_1 \bar{y}_1 + P_2 \bar{y}_2}{P_1 m_1 + P_2 m_2} = \frac{\bar{y}_1 + R \bar{y}_2}{m_1 + R m_2} \quad (12.4)$$

and

$$x_{\text{new}} = \frac{\bar{x}_1 + R \bar{x}_2}{m_1 + R m_2} . \quad (12.5)$$

For a white-light emitter, x_{new} and y_{new} can be chosen to coincide with the chromaticity coordinates of the Illuminant C standard ($x_c = 0.3101$, $y_c = 0.3162$; CIE, 1932; Judd, 1951), i.e., $x_{\text{new}} = x_c = 0.3101$ and $y_{\text{new}} = y_c = 0.3162$. Solving Eq. (12.5) for the power ratio R yields

$$R = \frac{\bar{y}_1 - y_c m_1}{y_c m_2 - \bar{y}_2} . \quad (12.6)$$

The power ratio as calculated from Eq. (12.6) is shown in Fig. 12.10.

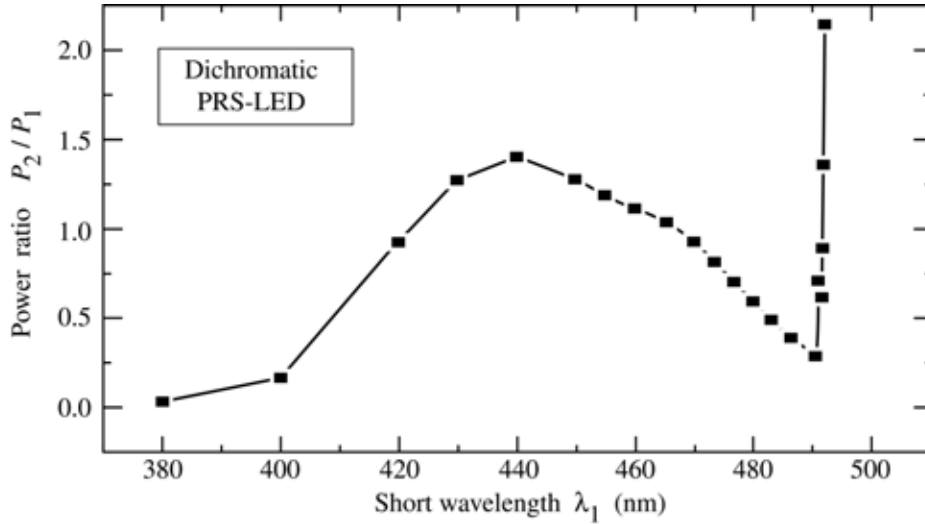


Fig. 12.10. Calculated power ratio between the two optical output powers P_1 and P_2 required to obtain white light emission (after Guo *et al.*, 1999).

12.7 Calculation of the luminous efficiency of PRS-LED

To produce the optical power P_2 at the wavelength of λ_2 through the recycling of photons from the primary source with wavelength λ_1 , the optical power required from the primary source is given by

$$\frac{P_2}{\eta_2} \frac{\lambda_2}{hc} \frac{hc}{\lambda_1} = \frac{P_2 \lambda_2}{\eta_2 \lambda_1} \quad (12.7)$$

where η_2 is the optical-to-optical conversion efficiency of the photon recycling light source. If P_0 is the electrical input power, then the optical power emitted by the primary LED source is $\eta_1 P_0$, where η_1 is the electrical-to-optical power conversion efficiency of the primary LED. Thus, the optical power emitted by the primary LED is given by

$$P_1 + \frac{P_2 \lambda_2}{\eta_2 \lambda_1} = \eta_1 P_0. \quad (12.8)$$

Solving the equation for the electrical input power and using $P_2 = R P_1$ yields

$$P_0 = P_1 \left(\frac{1}{\eta_1} + \frac{R \lambda_2}{\eta_1 \eta_2 \lambda_1} \right). \quad (12.9)$$

The total optical output power of the PRS-LED is given by

$$P_{\text{out}} = P_1 + P_2 = (1 + R) P_1 \quad (12.10)$$

so that the overall electrical-to-optical power efficiency of the photon-recycling dichromatic light source is given by

$$\eta = \frac{P_{\text{out}}}{P_0} = \frac{P_1 (1 + R)}{P_1 \left(\frac{1}{\eta_1} + \frac{R \lambda_2}{\eta_1 \eta_2 \lambda_1} \right)} = \frac{1 + R}{\frac{1}{\eta_1} + \frac{R \lambda_2}{\eta_1 \eta_2 \lambda_1}}. \quad (12.11)$$

The luminous flux Φ_{lum} of the new color, is given by

$$\Phi_{\text{lum}} = 683 \frac{\text{lm}}{\text{W}} (\bar{y}_1 P_1 + \bar{y}_2 P_2) = 683 \frac{\text{lm}}{\text{W}} (\bar{y}_1 + \bar{y}_2 R) P_1. \quad (12.12)$$

Then the luminous efficacy (measured in lumens per optical Watt) of the photon-recycling semiconductor LED is given by

$$\frac{\Phi_{\text{lum}}}{P_{\text{out}}} = 683 \frac{\text{lm}}{\text{W}} \frac{\bar{y}_1 + \bar{y}_2 R}{1 + R}. \quad (12.13)$$

Thus, the luminous efficiency (measured in lumens per electrical Watt) of the PRS-LED is given by

$$\frac{\Phi_{\text{lum}}}{P_0} = 683 \frac{\text{lm}}{\text{W}} \eta \frac{\bar{y}_1 + \bar{y}_2 R}{1 + R} . \quad (12.14)$$

Using this formula, we calculate the luminous efficiency as a function of the primary wavelength. The result of the calculation is shown in Fig. 12.11 for *ideal* sources, i.e. for $\eta_1 = \eta_2 = 100\%$.

The maximum luminous efficiency occurs if the primary wavelength source emits at 440 nm. A theoretical luminous efficiency of 336 lm/W is obtained for this wavelength. Note that we assume in the calculation that both light sources emit monochromatic light. However, the spontaneous emission from semiconductors has a $1.8 kT$ spectral width. Taking into account a finite linewidth, the expected luminous efficiency is slightly lower.

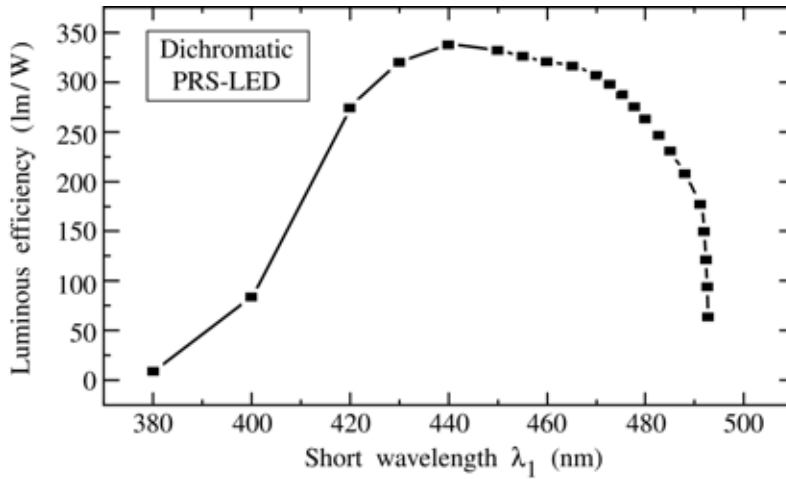


Fig. 12.11. Calculated luminous efficiency of a dichromatic PRS-LED versus its primary emission wavelength (after Guo *et al.*, 1999).

12.8 Spectrum of PRS-LED

PRS-LEDs have been demonstrated using a GaInN/GaN LEDs emitting in the blue and an electrically passive AlGaInP photon recycling semiconductor emitting in the red part of the spectrum (Guo *et al.*, 1999). The emission spectrum of the device is shown in Fig. 12.12. It shows the emission spectrum of the primary LED at 470 nm and a second emission line at 630 nm due to absorption of the 470 nm light in the AlGaInP layer and re-emission of light at 630 nm. The recycling semiconductor used in this experiment is an AlGaInP/GaAs double

heterostructure. The photon-recycling semiconductor is planar and no surface texturing was performed.

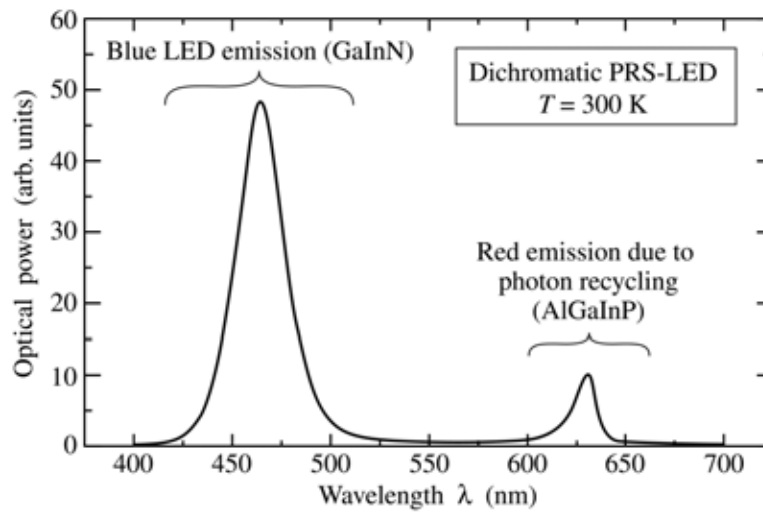


Fig. 12.12. Emission spectrum of dichromatic PRS-LED with current-injected GaInN blue LED primary source and AlGaInP photon recycling wafer (secondary source) emitting in the red (after Guo *et al.*, 2000).

To avoid absorption of light in the GaAs substrate, the GaAs substrate of the AlGaInP epitaxial layer was removed. First, the AlGaInP/GaAs recycling semiconductor was mounted on a glass slide. Subsequently, the GaAs substrate was removed by polishing and selective wet chemical etching. Then, the primary LED wafer and the photon-recycling wafer were bonded together.

The theoretical luminous efficiency of the PRS-LED and several other devices is given in Table 12.2. The results show that the dichromatic light source has the highest luminous efficiency as compared to spectrally broader emitters.

Type of LED	Luminous efficiency (lm/W)	Chromaticity coordinates (x, y)
Dichromatic LED	300 – 340	(0.31, 0.32)
Broadened dichromatic LED	280 – 320	(0.31, 0.32)
Trichromatic LED	240 – 290	(0.31, 0.32)
Phosphor-based LED	200 – 280	(0.31, 0.32)

Table 12.2. Theoretical luminous efficiency of different types of white LEDs assuming unit quantum efficiency for the devices and the absence of resistive power losses.

Generally, dichromatic white LEDs have a higher luminous efficacy but lower color-rendering index (CRI) compared with trichromatic white LEDs. It can be shown that there is a fundamental trade-off between color rendering and the luminous efficacy of light-emitting devices (Walter, 1971). In order to improve the general CRI of dichromatic devices such as the PRS-LED, two possibilities can be considered. *First*, the emission lines can be intentionally broadened, e.g. by compositional grading. *Secondly*, a second photon-recycling semiconductor can be added thus creating a tri-chromatic PRS-LED. However, any broadening of the two emission lines or the addition of an emission line will decrease the luminous efficacy and efficiency of the device.

12.9 White LEDs based on dye converters

White LEDs can also be fabricated using organic dye molecules as wavelength converter materials. The dyes can be incorporated in the epoxy encapsulant (Schlotter *et al.*, 1997). Dyes can also be incorporated in optically transparent polymers.

A drawback of organic dyes are their finite lifetime. Dye molecules “bleach out”, i.e. become optically inactive, after a certain number of photon absorption events. Typically a dye molecule is stable for about $10^4 - 10^6$ optical transitions (Jones, 2000). The molecular stability of dyes is a serious drawback. The lifetime of dyes is considerably shorter than the lifetime of semiconductor or phosphor wavelength converters.

Dyes have a relatively small difference between the absorption and the emission band (Stokes shift). For example, the Stokes shift for the dye *Coumarin 6* is just 50 nm, as shown in Fig. 12.3. This shift is smaller than the Stokes shift required for dichromatic white LEDs that need typical wavelength shifts of 100 nm or more, as inferred from Fig. 12.1.

References

- Bando K., Noguchi Y., Sakano K., and Shimizu Y. (in Japanese) *Tech. Digest, Phosphor Res. Soc.*, 264th Meeting, November 29 (1996)
- CIE *Commission Internationale de l'Eclairage Proceedings, 1931* (Cambridge University Press, Cambridge, 1932)
- Guo X., Graff J. W., and Schubert E. F. “Photon-recycling semiconductor light-emitting diode” *IEDM Technical Digest*, **IEDM-99**, 600 (1999)
- Guo X., Graff J. W., and Schubert E. F. “Photon-recycling for high brightness LEDs” *Compound Semiconductors* **6**, May/June (2000)

- Ivey H. F. "Color and efficiency of luminescent light sources" *J. Opt. Soc. America* **53**, 1185 (1963)
- Jones G., personal communication (2000)
- Judd D. B. "Report of US secretariat committee on colorimetry and artificial daylight" in *Proceedings of the Twelfth Session of the CIE, Stockholm* Vol. **1**, p. 11 (Bureau central de la CIE, Paris, 1951)
- Karlicek Jr. R. F., personal communication (1999)
- Kaufmann U., Kunzer M., Köhler K., Obloh H., Pletschen W., Schlotter P., Schmidt R., Wagner J., Ellens A., Rossner W., and Kobusch M. "Ultraviolet pumped tricolor phosphor blend white emitting LEDs" *Phys. Stat. Sol. (a)* **188**, 143 (2001)
- MacAdam D. L. "Maximum attainable luminous efficiency of various chromaticities" *J. Opt. Soc. Am.* **40**, 120 (1950)
- MacAdam D. L. "Color measurement: theme and variations" (Springer, New York, 1985)
- Nakamura S., Senoh M., Iwasa N., Nagahama S., Yamada T., and Mukai T. "Superbright green InGaN single-quantum-well-structure light-emitting diodes" *Jpn. J. Appl. Phys. (Letters)* **34**, L1332 (1995)
- Nakamura S. and Fasol G. "The blue laser diode" (Springer, Berlin, 1997)
- Osram-Sylvania Corporation. Data sheet on type 4350 phosphor (2000)
- Schlotter P., Schmidt R., and Schneider J. "Luminescence conversion of blue light emitting diodes" *Appl. Phys. A* **64**, 417 (1997)
- Schlotter P., Baur J., Hielscher C., Kunzer M., Obloh H., Schmidt R., and Schneider J. "Fabrication and characterization of GaN/InGaN/AlGaIn double heterostructure LEDs and their application in luminescence conversion LEDs" *Materials Sci. Eng.* **B59**, 390 (1999)
- Thornton W. A. "Luminosity and color-rendering capability of white light" *J. Opt. Soc. America* **61**, 1155 (1971)
- Vos, J. J. "Colorimetric and photometric properties of a 2-degree fundamental observer" *Color Research and Applications* **3**, 125 (1978)
- Walter W. "Optimum phosphor blends for fluorescent lamps" *Appl. Opt.* **10**, 1108 (1971)
- Wyszecki G. and Stiles W. S. "Color science: concepts and methods, quantitative data and formulae" 2nd edition (John Wiley and Sons, New York 1982)
- Zukauskas A., Shur M. S., and Gaska R. "Introduction to solid-state lighting" (John Wiley and Sons, New York, 2002)

13

Optical communication

LEDs are used in communication systems transmitting low and medium data rates (< 1 Gbit/s) over short and medium distances (< 10 km). These communication systems are based on either *guided light waves* (Hecht, 2001; Keiser, 1999; Mynbaev and Scheiner, 2001) or *free-space waves* (Carruthers, 2002; Heatley *et al.*, 1998; Kahn and Barry, 2001). In guided-wave communication, individual optical fibers or fiber bundles are used as the transmission medium and LED-based optical communication links are limited to distances of a few kilometers. Optical fiber systems include *silica* and *plastic* optical fibers. Free-space communication is usually limited to a room, even though longer distances are possible. In this chapter we discuss the characteristics of transmission media used for LED communication.

13.1 Types of optical fibers

The cross section of optical fibers consists of a circular core region surrounded by a cladding region. The core region has a higher refractive index than the cladding region. Typically, the core refractive index is about 1 % higher than the cladding refractive index. Light propagating in the core is guided inside the core by means of *total internal reflection*. The condition of total internal reflection can be inferred from Snell's law. A light ray is *totally internally reflected* whenever it is incident at the core–cladding boundary. In a ray-optics picture, light rays propagating inside the core follow a *zigzag path*.

There are three types of optical fibers used in communication systems. These types are the (i) *step-index multimode* fiber, (ii) *graded index multimode* fiber, and (iii) *single-mode* fiber. The three types of fibers are shown in Fig. 13.1 along with the refractive index profile.

Step-index multimode fibers have a relatively large core diameter. Typical core diameters are 50, 62.5, and 100 μm for silica fibers used in communication systems. Plastic optical fibers have larger core diameters, typically 1 mm. An important advantage of multimode fibers is *easy coupling* of the light source to the fiber. Usually, a ± 5 μm accuracy of alignment is sufficient for

multimode fibers with a core diameter of 50 μm . The main disadvantage of multimode fibers is the occurrence of *modal dispersion*.

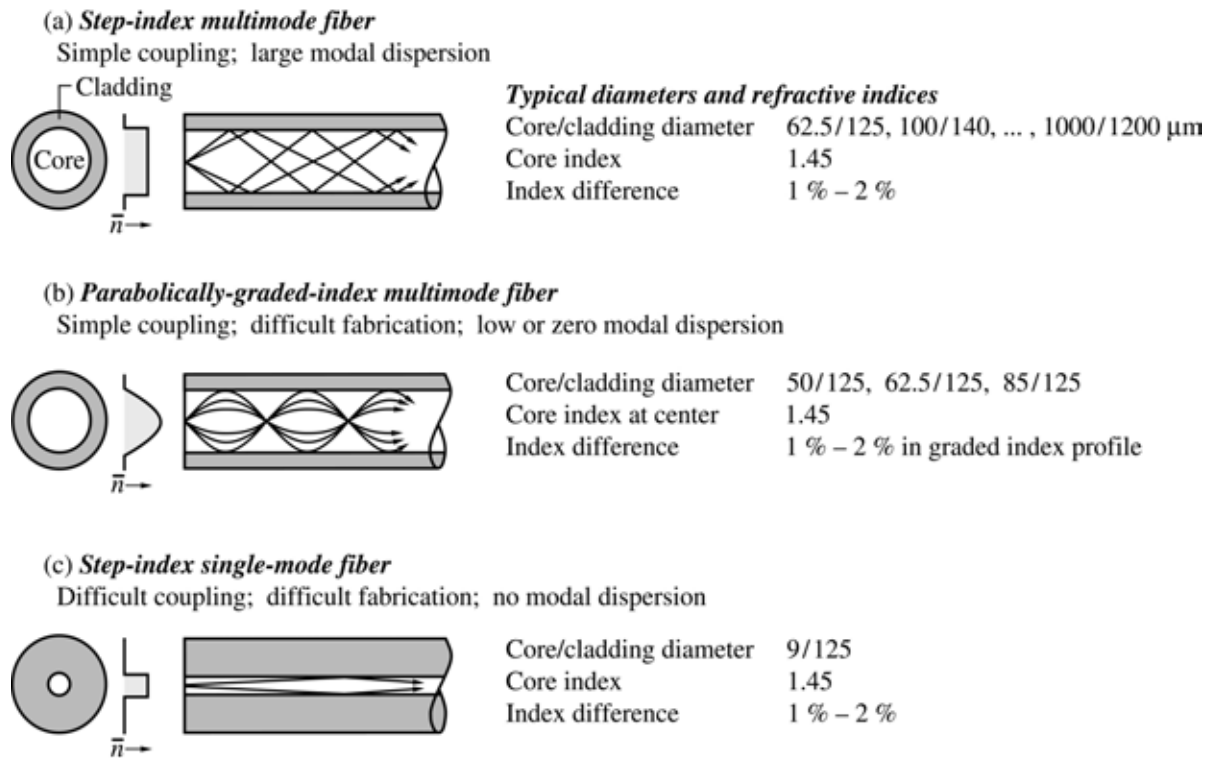


Fig. 13.1. (a) Step-index multimode fibers allow for the propagation of several optical modes. (b) Parabolically graded-index multimode fibers allow for the propagation of several modes with similar propagation constant. Graded-index multimode fibers have a lower modal dispersion than step-index multimode fibers. (c) Step-index single-mode fibers have a small core diameter and no modal dispersion.

Because the core diameter in multi-mode fibers is much larger than the operating wavelength, several optical modes can propagate in the waveguide. These optical modes have different propagation constants so that different modes arrive at the end of the fiber at different times, even if they were launched at the same time. This leads to a broadening of the optical pulse and limits the maximum bit rate that can be transmitted over a fiber of a given length.

Modal dispersion is reduced by *graded-index* multimode fibers. Graded-index multimode fibers have a parabolically graded core index leading to a reduction in modal dispersion.

Single-mode fibers have such a small core diameter, that only a single optical mode can propagate in the fiber. Typical single-mode core diameters are 5 – 10 μm . The main advantage of single-mode fibers is the *lack of modal dispersion*. The main disadvantage of single-mode fibers is *difficult coupling* due to the small core diameter. A small core diameter requires light sources

with high brightness such as lasers. However, LEDs, in particular edge-emitting LEDs and superluminescent LEDs are also occasionally used with single-mode fibers. Coupling of light into single-mode fibers requires precise alignment with tolerances of a few micrometers.

If the *optical power* to be transmitted over an optical fiber is of prime interest, the core diameter should be as large as possible and the core–cladding index difference should also be as large as possible. Speciality fibers with core diameters of > 1 mm are available. Such fibers are not suitable for communication applications due to the large modal dispersion.

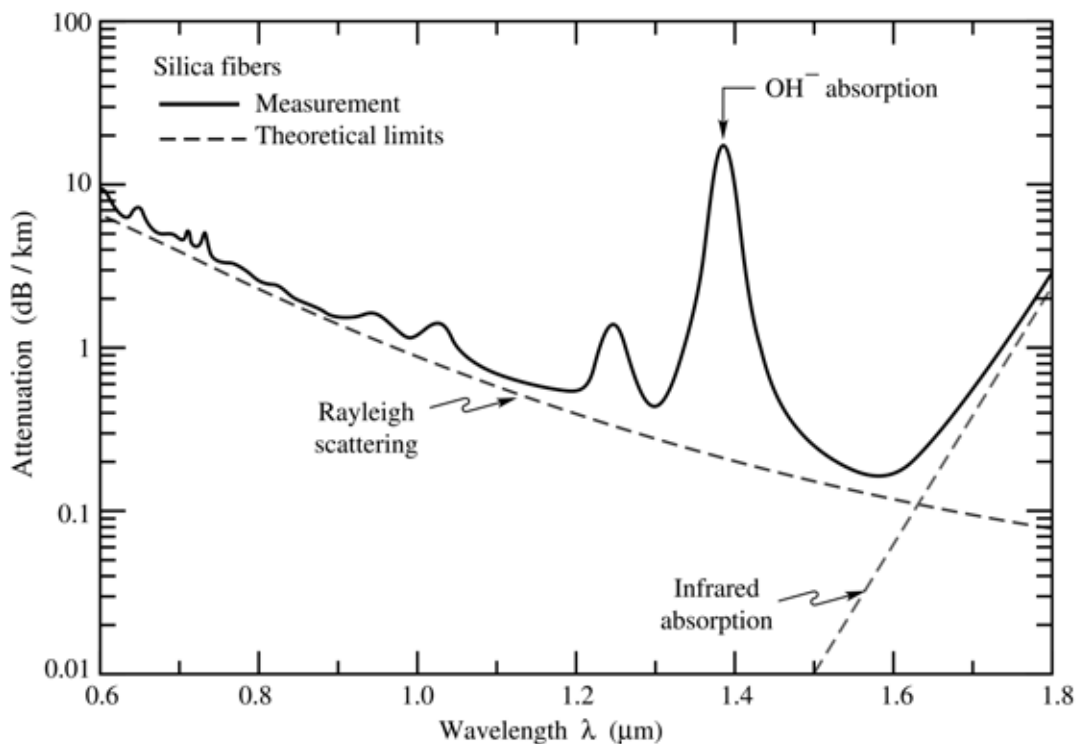


Fig. 13.2. Measured attenuation in silica fibers (solid line) and theoretical limits (dashed lines) given by Rayleigh scattering in the short-wavelength region, and by molecular vibrations (infrared absorption) in the infrared spectral region.

13.2 Attenuation in silica and plastic optical fibers

Silica (SiO_2) has excellent optical properties including great long-term stability. A large variety of glasses and fibers are available. The attenuation of silica fibers is shown in Fig. 13.2. Inspection of Fig. 13.2 reveals that a minimum loss of 0.2 dB occurs at a wavelength of 1.55 μm .

There are several optical “windows” for communication over silica fibers. These communication windows are at 0.85, 1.3, and 1.55 μm . The 0.85 μm communication window is suitable for communication with GaAs-based LEDs and lasers. However, this window is limited to short distances due to the material dispersion and due to the high attenuation of silica fibers at

that wavelength. The $1.3\ \mu\text{m}$ communication window is also suited for communication with LEDs and lasers. This window has relatively low loss and zero dispersion, allowing for high-bit-rate transmission, in particular in graded-index and single-mode fibers. The $1.55\ \mu\text{m}$ communication window is characterized by the lowest loss of all three windows. Consequently, this window is used for long-distance high-bit-rate communication. To allow for high bit rates, single-mode fibers must be used. Since it is difficult to efficiently couple light emerging from an LED into a single-mode fiber, lasers are preferred over LEDs at $1.55\ \mu\text{m}$.

Plastic optical fibers are becoming increasingly popular for short-distance communication. However, plastic fibers have losses that are about 1000 times greater than the losses in silica fibers. Therefore, the transmission distances are limited to just a few meters to a few hundred meters, e.g. for communication within an automobile or airplane.

The attenuation in plastic fibers is shown in Fig. 13.3. The preferred communication window of plastic fibers is at $650\ \text{nm}$, where the loss is of the order of $0.1 - 0.2\ \text{dB}$ per meter. At even shorter wavelengths, the attenuation in plastic fibers decreases. However, the material dispersion increases, thus making the $650\ \text{nm}$ wavelength the preferred communication wavelength in plastic optical fibers.

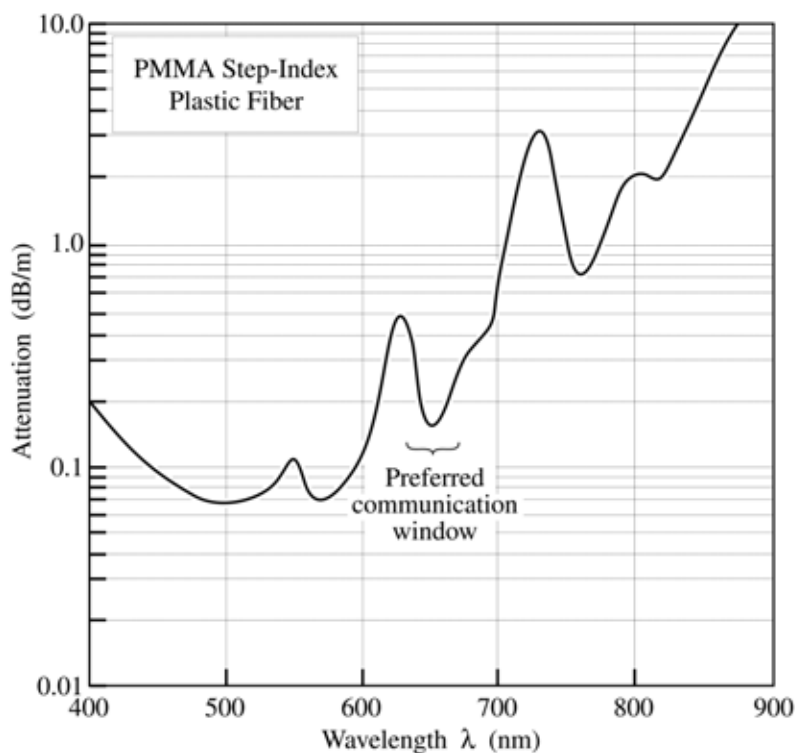


Fig. 13.3. Attenuation of a PMMA step-index plastic optical fiber. At $650\ \text{nm}$, the preferred communication wavelength, the attenuation is about $150\ \text{dB/km}$ (after data sheet of Toray Industries Ltd.).

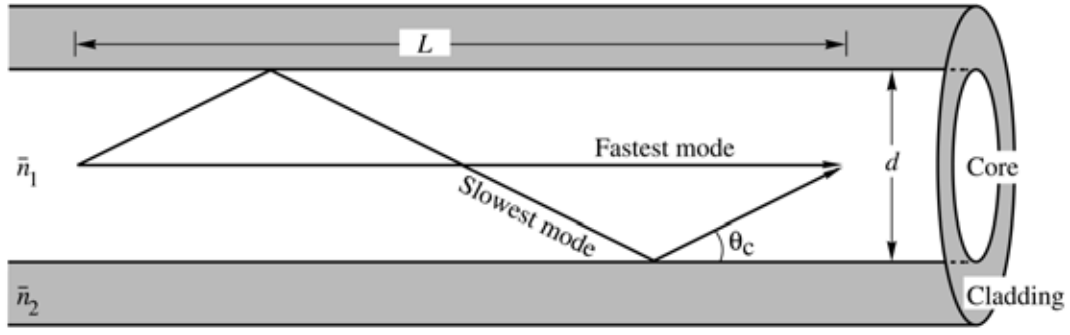


Fig. 13.4. Geometry used for calculation of the modal dispersion in a multimode fiber waveguide.

13.3 Modal dispersion in fibers

Modal dispersion occurs in multimode fibers that have a larger core diameter or a larger index difference between the core and the cladding than single-mode fibers. Typical core diameters range from 50 to 1000 μm for multimode fibers and 5 – 10 μm for single-mode fibers. In the ray optics model, *different optical modes* correspond to light rays propagating at *different angles* in the core of the waveguide. The derivation of propagation angles in multimode fibers would go beyond the scope of this chapter. Here an approximate calculation will be performed to obtain the modal dispersion.

Consider a fiber waveguide with refractive indices of the core and cladding of \bar{n}_1 and \bar{n}_2 , respectively. Assume that the waveguide supports the propagation of more than one optical mode. Two of these modes are shown schematically in Fig. 13.4. Owing to the difference in optical path length, the mode with the smaller propagation angle θ will arrive earlier at the end of the multimode fiber. The **modal dispersion** is the time delay between the fastest and the slowest optical mode normalized to the length L of the waveguide.

In the calculation, assume that the phase and group velocity are given by $v_{\text{ph}} = c / \bar{n}_1 \approx v_{\text{gr}}$. The fastest mode has the smallest propagation angle and we approximate the smallest angle by $\theta_{m=0} \approx 0^\circ$. The slowest mode has the largest propagation angle and we approximate the largest angle by $\theta_m \approx \theta_c$, where θ_c is the critical angle of total internal reflection. This approximation can be made without loss of accuracy for multimode fibers which carry many modes.

The propagation times for the fastest and slowest mode per unit length of the fiber are given by

$$\tau_{\text{fast}} = \frac{L}{c / \bar{n}_1} \quad \tau_{\text{slow}} = \frac{L / \cos \theta_c}{c / \bar{n}_1} \quad (13.1)$$

where the critical angle for total internal reflection can be derived from Snell's law and is given by

$$\theta_c = \arccos(\bar{n}_2 / \bar{n}_1). \quad (13.2)$$

The time delay per unit length, or modal dispersion, is then given by

$$\frac{\Delta\tau}{L} = \frac{\tau_{\text{slow}} - \tau_{\text{fast}}}{L} = \frac{\bar{n}_1}{c} \left(\frac{1}{\cos \theta_c} - 1 \right). \quad (13.3)$$

A waveguide supporting *many* modes has a large time delay between the fastest and slowest mode. Thus modal dispersion increases with the number of optical modes supported by the waveguide.

Exercise. Modal dispersion in waveguides. Calculate the time delay between the slowest and the fastest mode, and the maximum possible bit rate for a 1 km long multimode fiber waveguide with core refractive index $\bar{n}_1 = 1.45$ and cladding refractive index $\bar{n}_2 = 1.4$.

Solution: Using Snell's law (Eq. 13.2), one obtains $\theta_c \approx 15^\circ$. The time delay calculated from Eq. (13.3) amounts to $\Delta\tau = 170$ ns. The minimum time required to transmit one bit of information is given by $\Delta\tau$. This yields an approximate maximum bit rate of $f_{\text{max}} = 1/170 \text{ ns} = 5.8 \text{ Mbit/s}$. The calculation shows that modal dispersion can be a significant limitation in optical communication. Graded-index multimode fibers or single-mode fibers are therefore required for high-speed communication systems.

13.4 Material dispersion in fibers

Material dispersion is another mechanism limiting the capacity of optical fibers. Material dispersion is due to the dependence of the refractive index on the wavelength. Figure 13.5 shows, as a function of wavelength, the phase refractive index and the group refractive index of silica. The indices are defined as

$$\bar{n} = \frac{c}{v_{\text{ph}}} \quad (\text{phase refractive index}) \quad (13.4)$$

and

$$\bar{n}_{\text{gr}} = \frac{c}{v_{\text{gr}}} \quad (\text{group refractive index}) \quad (13.5)$$

where v_{ph} and v_{gr} are the phase and group velocity in silica, respectively. The *phase refractive index* and the *group refractive index* are related by

$$\bar{n}_{\text{gr}} = \bar{n} - \lambda \frac{d\bar{n}}{d\lambda} = \bar{n} - \lambda_0 \frac{d\bar{n}}{d\lambda_0} . \quad (13.6)$$

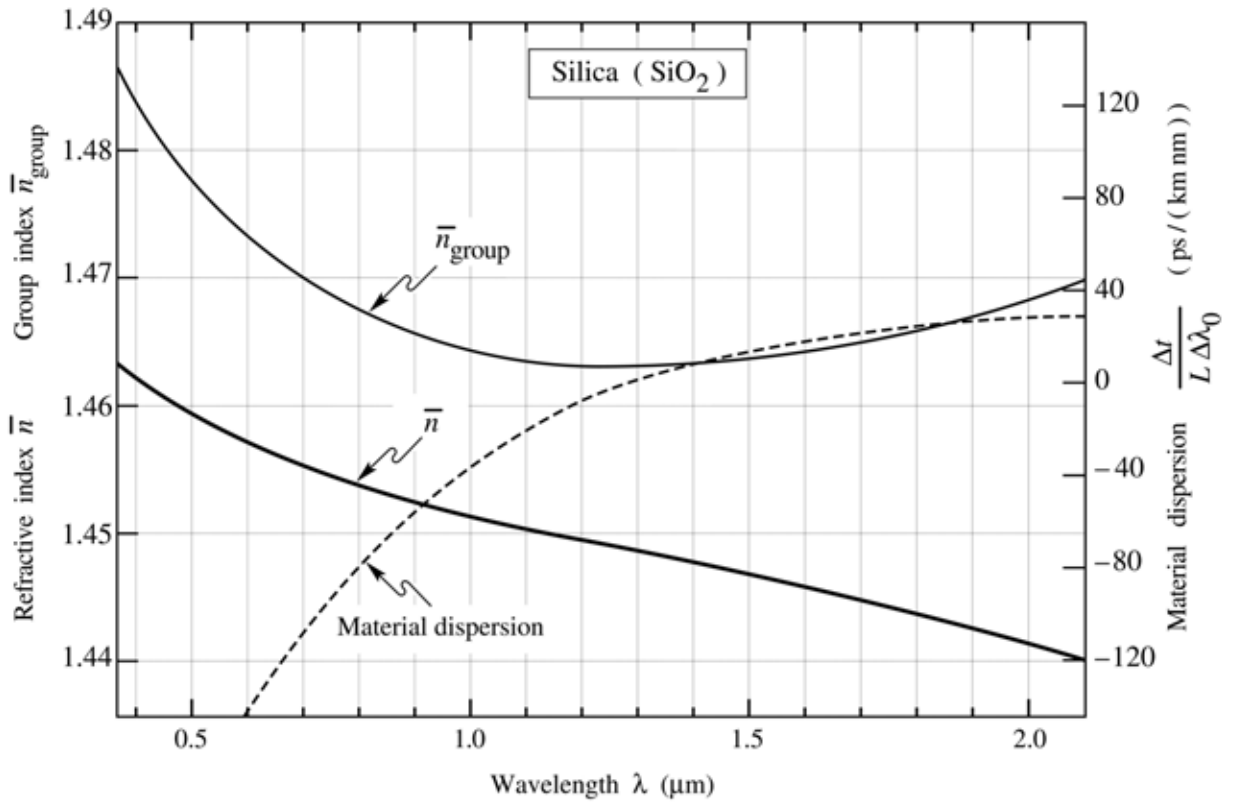


Fig. 13.5. Refractive index, group index, and material dispersion of a silica fibers for an optical signal spectral width $\Delta\lambda_0$ in vacuum. The material dispersion of regular silica fibers is zero at $\lambda \approx 1.3 \mu\text{m}$.

If the fiber is dispersive, the difference in group velocity of the “slowest color” and the “fastest color” of the optical signal is given by

$$\left| \Delta v_{\text{gr}} \right| = \frac{c}{\bar{n}_{\text{gr}}^2} \frac{d\bar{n}_{\text{gr}}}{d\lambda} \Delta\lambda \quad (13.7)$$

where $\Delta\lambda$ is the spectral width of the optical signal.

The time delay between the leading edge and trailing edge of an optical signal after travelling in the fiber for a length L , called the **material dispersion**, is given by

$$\Delta\tau = \frac{L}{v_{\text{gr}}^2} \Delta v_{\text{gr}} = \frac{L}{c} \frac{d\bar{n}_{\text{gr}}}{d\lambda} \Delta\lambda = \frac{L}{c} \frac{d\bar{n}_{\text{gr}}}{d\lambda_0} \Delta\lambda_0. \quad (13.8)$$

The material dispersion is measured in ps/(nm km) and it is illustrated for silica fibers in Fig. 13.5. LEDs have a broad emission linewidth. Therefore material dispersion is, along with modal dispersion, the bandwidth-limiting factor in optical fiber communication systems operated with LEDs.

Exercise. Material dispersion in waveguides. Derive Eqs. (13.6) and (13.7). Why does material dispersion have a much smaller significance for semiconductor lasers than for LEDs?

Substantial material dispersion exists in plastic fibers at all wavelengths of interest. These wavelengths are the local loss minimum at 650 nm and the low-loss region of 500 – 600 nm. The material dispersion is given in Table 13.1. The data indicates that 650 nm is the wavelength of least dispersion, making 650 nm the preferred communication wavelength in plastic optical fibers.

Wavelength	525 nm	560 nm	650 nm
Material dispersion	700 ps/(nm km)	500 ps/(nm km)	320 ps/(nm km)

Table 13.1. Material dispersion in PMMA plastic optical fibers (courtesy of R. Marcks von Wurtemberg, *Mitel* Corporation, 2000).

Exercise. Comparison of material and modal dispersion. Consider a 62.5 μm core diameter multimode step-index fiber of 3 km length with a core index of $\bar{n}_1 = 1.45$ and a cladding index of $\bar{n}_2 = 1.4$. Assume that the fiber inputs come from either an LED or a laser emitting at 850 nm.

Assume that the LED and the laser have a linewidth of 50 and 5 nm, respectively. Calculate the material and the modal dispersion for each case and explain the result.

13.5 The numerical aperture of fibers

Owing to the requirement of total internal reflection, the only light rays that can propagate losslessly in the core of an optical fiber are those that have a propagation angle smaller than the critical angle for total internal reflection. Light rays for which the propagation angle is too large will consequently not couple into the fiber. Here we consider the coupling of light from an LED light source into an optical fiber. We assume that the fiber end has a polished planar surface normal to the optical axis of the fiber, as shown in Fig. 13.6.

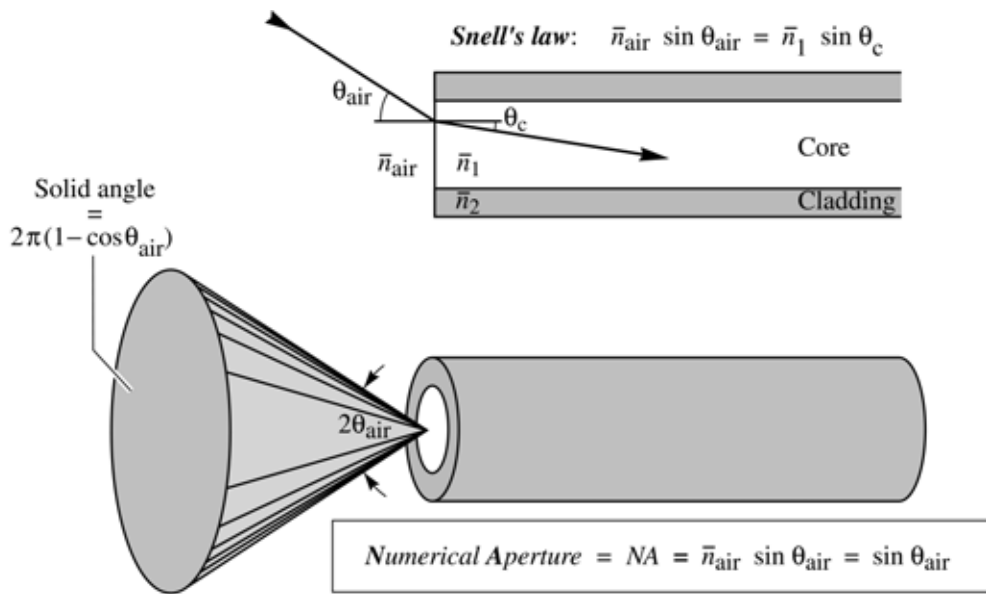


Fig. 13.6. Illustration of the *numerical aperture* (NA) of a fiber. For example, the light acceptance angle in air is $\theta_{\text{air}} = 11.5^\circ$ for a numerical aperture $NA = 0.2$.

As a consequence of the requirement of total internal reflection, only a range of angles will be “accepted” by the fiber for lossless propagation. Outside the **acceptance angle** range, light rays will be refracted into the cladding layer where they will incur losses.

The range of allowed angles can be inferred from Snell’s law. As illustrated in Fig. 13.6, the maximum angle for acceptance in the fiber is given by

$$\bar{n}_{\text{air}} \sin \theta_{\text{air}} = \bar{n}_1 \sin \theta_c . \quad (13.9)$$

Since the refractive index of air is approximately unity, the maximum acceptance angle in air is given by

$$\theta_{\text{air}} = \arcsin(\bar{n}_1 \sin \theta_c). \quad (13.10)$$

The maximum acceptance angle defines a cone of allowed angles, as shown in Fig. 13.6. Light rays incident on the core of the optical fiber with propagation angles within the cone can propagate without loss.

Another way to express the acceptance cone is the **numerical aperture** of the fiber. The numerical aperture (NA) is defined as

$$NA = \bar{n}_1 \sin \theta_c = \bar{n}_{\text{air}} \sin \theta_{\text{air}} = \sin \theta_{\text{air}} \approx \theta_{\text{air}} \quad (13.11)$$

where the approximation $\sin \theta_{\text{air}} \approx \theta_{\text{air}}$ is valid for *small* numerical apertures. Typical NA s for silica single-mode fibers are 0.1 and typical NA s for silica multimode fibers are 0.15 – 0.25. Plastic optical fibers can have higher NA s, typically 0.2 – 0.4.

The solid angle corresponding to a certain NA is given by

$$\text{Solid angle} = \Omega = 2\pi (1 - \cos \theta_{\text{air}}) = 2\pi [1 - \cos(\arcsin NA)] \approx \pi NA^2 \quad (13.12)$$

where the small-angle approximations $\sin \theta_{\text{air}} \approx \theta_{\text{air}}$ and $\cos \theta_{\text{air}} \approx 1 - (1/2)\theta_{\text{air}}^2$ have been used. The power emitted by an LED is proportional to the solid angle. Thus the power coupled to an LED is proportional to the NA squared of the fiber in the small-angle approximation.

Exercise. Coupling efficiency of a fiber butt coupled to an LED. Consider an LED with a point-like emission region that emits an optical power of 1 mW into the hemisphere. For simplicity, assume that the intensity emitted by the LED is independent of the emission angle. What is the maximum acceptance angle of a single-mode fiber with $NA = 0.1$ and multimode fiber with $NA = 0.25$? What is the power that can be coupled into the two fibers?

Solution: The maximum acceptance angle of the single-mode and multimode fibers in air are $\theta_{\text{air}} = 5.7^\circ$ and 14.5° , respectively. The solid angle defined by an acceptance angle θ_{air} is given by $\Omega = 0.031$ and 0.20 for the single-mode and multimode fiber, respectively. Since the entire

hemisphere has a solid angle of 2π , the power coupled into the single-mode and multimode fibers is given by 0.0049 and 0.032 mW, respectively.

13.6 Coupling with lenses

The low coupling efficiency of LEDs to optical fibers can be improved with convex lenses, if the light-emitting region of the LED is smaller than the optical fiber core. In this case, the light-emitting region can be imaged on the fiber core, thereby reducing the angle of incidence. The light source is *adapted* to the *NA* of the fiber (“*NA-matching*”).

A convex lens can produce an image with height I of a light-emitting object with height O . If the image is larger than the object, the angles of the light incident from the lens on the image are less divergent than the light emanating from the object towards the lens. The *smaller divergence* obtained for *magnified* images allows one to increase the coupling efficiency to fibers. The principle of coupling with a convex lens is shown in Fig. 13.7.

The condition for a focused image (minimum image size) is given by the lens equation

$$\frac{1}{d_O} + \frac{1}{d_I} = \frac{1}{f} \quad (13.13)$$

where d_O and d_I are the distance of the object and the image from the lens, respectively, and f is the focal length of the lens.

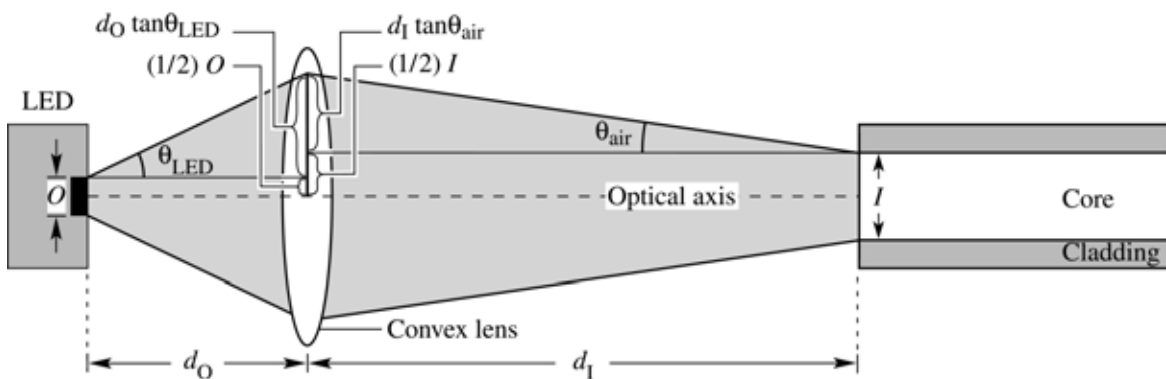


Fig. 13.7. Schematic illustration of coupling with a lens by imaging the light-emitting region of an LED onto the core of an optical fiber. The LED has a circular emission region with diameter O (Object). The emission region is imaged onto the fiber core with diameter I (Image) using a convex lens with focal length f .

The magnification of the image of the LED light source on the core of the fiber is given by

$$M = \frac{I}{O} = \frac{d_I}{d_O} . \quad (13.14)$$

As shown in Fig. 13.7, it is

$$\frac{1}{2} O + d_O \tan \theta_{\text{LED}} = \frac{1}{2} I + d_I \tan \theta_{\text{air}} . \quad (13.15)$$

If the LED and the core of the optical fiber are much smaller than the diameter of the lens and if the angles are relatively small, then Eq. (13.15) can be approximated by

$$\theta_{\text{LED}} = \frac{d_I}{d_O} \theta_{\text{air}} = \frac{I}{O} \theta_{\text{air}} . \quad (13.16)$$

Since d_I is larger than d_O , the acceptance angle for light emanating from the LED is *larger* than that of the fiber, implying *increased* coupling efficiency. Thus, we can define the numerical aperture of the LED, NA_{LED} , which defines the angle of light emanating from the LED that is accepted by the fiber. Using Eq. (13.14) and the small-angle approximation for NA , NA_{LED} is given by

$$NA_{\text{LED}} = \frac{I}{O} NA . \quad (13.17)$$

Since the coupling efficiency is proportional to NA^2 (see Eq. 13.12), the coupling efficiency is increased to

$$\text{Coupling efficiency} \propto NA_{\text{LED}}^2 = \left[(I/O) NA \right]^2 \quad (13.18)$$

The result shows that high coupling efficiencies are obtained for LEDs with small-diameter light-emitting regions, large fiber-core diameters, and large- NA fibers.

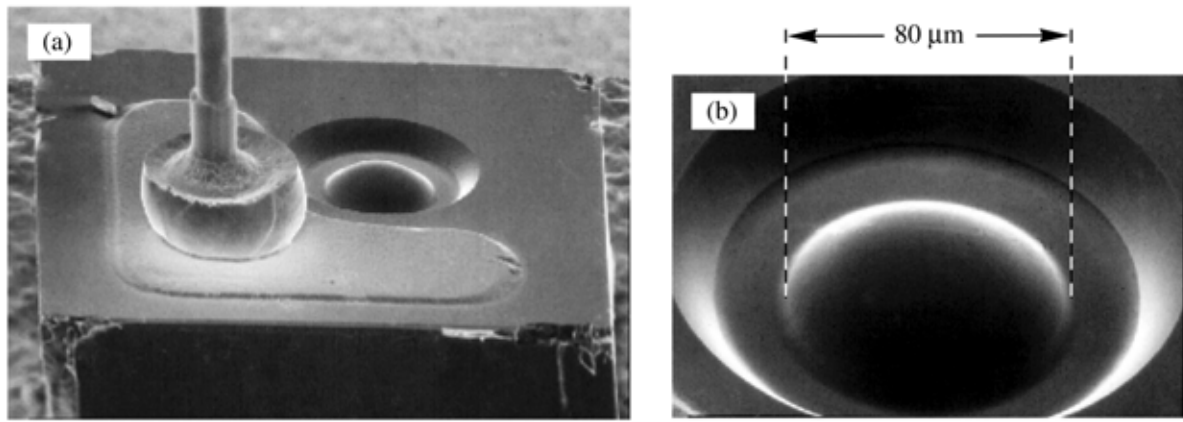


Fig. 13.8. (a) Commercial communication LED chip with integrated lens. (b) Detailed picture of the lens etched by a photochemical process into the GaAs substrate (AT&T ODL product line, 1995).

Exercise. Coupling efficiency of a fiber coupled to an LED with a lens. Consider an LED circular emission region with diameter $20\ \mu\text{m}$ coupled to a silica multimode fiber with $NA = 0.2$ and a core diameter of $62.5\ \mu\text{m}$. The LED emits a power of $1\ \text{mW}$ into the hemisphere lying above the planar LED surface. For simplicity, assume that the LED emission intensity is independent of the emission angle. What is the maximum power that can be coupled into the multimode fiber?

Solution: Improved coupling can be obtained by imaging the LED emission region on to the core of the optical fiber. For maximum coupled power, a convex lens with magnification $M = 62.5\ \mu\text{m} / 20\ \mu\text{m} = 3.125$ can be used. Using the lens, the acceptance angle of the fiber is increased from $\theta_{\text{air}} = 11.5^\circ$ to $\theta_{\text{LED}} = 35.9^\circ$. The solid angle defined by the LED acceptance angle θ_{LED} is given by $\Omega = 1.19$. Since the LED emits $1\ \text{mW}$ into the entire hemisphere (with solid angle $\Omega = 2\pi$), the power coupled into the fiber is given by $0.189\ \text{mW}$.

Lensed LEDs are frequently used in communication applications. The micrograph of an LED with a monolithically integrated lens is shown in Fig. 13.8. The light-emitting region of the LED is $20\ \mu\text{m}$ and the lens shown in Fig. 13.8 has a diameter of about $80\ \mu\text{m}$.

13.7 Free-space optical communication

Free-space optical communication (Carruthers, 2002; Heatley *et al.*, 1998; Kahn and Barry, 2001) is suitable for low to medium bit rates. The most common application of free-space optical communication is the remote control of consumer appliances such as stereos and television sets.

Other applications are the remote control of automobile door locks and the cordless interface between computers and peripheral devices such as a mouse, keyboard and printer.

Free-space optical communication is limited to line-of-sight applications since obstacles such as walls and floors will block the path of light. Furniture may also block the path of light. However, a light beam may be reflected from the ceiling so that communication may still be possible even if there is no direct line of sight connection between the optical transmitter and the receiver.

The wavelength of choice for free-space optical communication is the near infrared. GaAs light-emitting diodes emitting with good efficiency are readily available. Infrared light is preferred over visible light sources because the former does not provide a distraction to anyone near the optical transmitter.

Eye safety considerations limit the maximum power of optical transmitters. At a wavelength of 870 nm, the optical power is limited to typically a few mW. Other wavelengths, such as 1500 nm, allow for higher optical powers. The 1500 nm wavelength range is termed “*eye safe*”, since the cornea absorbs 1500 nm light thus preventing light from reaching the sensitive retina. The wavelength 1500 nm thus allows for higher optical powers than 870 nm sources.

If we restrict our considerations to small distances, the transmission medium air can be considered to be totally lossless. However, the optical signal strength decreases for uncollimated light beams due to spatial divergence. For *isotropic emitters*, the intensity decreases with the square of the radius, i.e.

$$I = P / (4\pi r^2) \quad (13.19)$$

where P is the optical power emitted by the source and r is the distance from the source. The decrease in intensity thus has a very different dependence compared with the intensity in fiber communication.

The rapidly decreasing intensity limits the maximum range of optical communication. Collimated light beams can overcome this problem. Transmission distances of several km are possible without significant loss provided that atmospheric conditions are good, i.e. in the absence of fog or precipitation. Semiconductor lasers are used for such collimated transmission systems due to the ability to form collimated beams with very little spatial dispersion.

Multipath distortion or ***multipath time delay*** severely limits the data rate in free-space optical communication systems. A schematic illustration of multipath distortion is shown in

Fig. 13.9. A light beam emanating from the optical transmitter may take several different paths from the transmitter to the receiver. This is especially true for rooms with high-reflectivity surfaces such as white ceilings, walls, or mirrors. As an approximate rule, the longest path is assumed to be twice as long as the shortest path between the transmitter and the receiver. This approximate rule leads to a multipath distortion time delay of

$$\Delta\tau = L / c \quad (13.20)$$

where L is the transmitter–receiver distance and c is the velocity of light. The maximum data rate is then limited to

$$f_{\max} \approx 1 / \Delta\tau . \quad (13.21)$$

For a room size of 5 m, the multipath delay is about $\Delta\tau = 17$ ns. Thus the data rate will be limited to about 60 MHz.

Another limitation of free-space optical communication is the detector noise. Sunlight and incandescent light sources have strong emission in the infrared. Thus a large DC photocurrent is generated in the detector, especially under direct sunlight conditions. The detector noise can be reduced by limiting the bandwidth of the receiver system. By reducing the bandwidth of the receiver system, and thereby also the system data rate, the detector noise is reduced, since the noise spectrum is much wider than the system bandwidth.

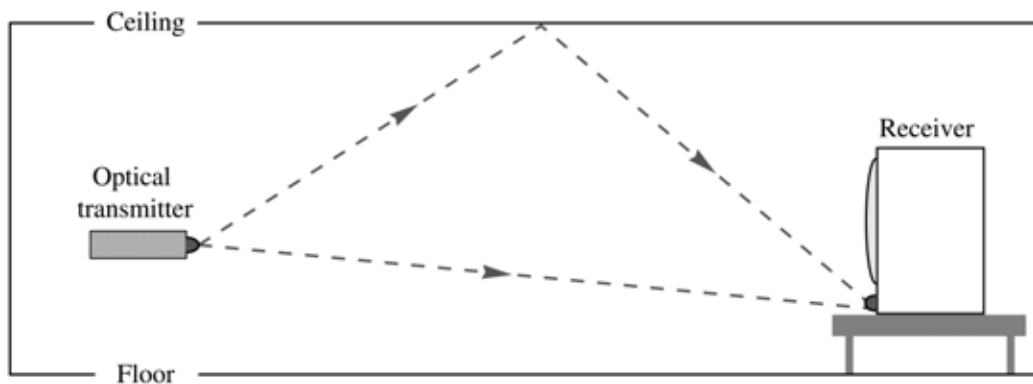


Fig. 13.9. Illustration of multipath distortion of a free-space optical signal limiting the data rate of the signal.

The detector noise due to ambient light sources can also be reduced by using optical band-pass filters, long-wavelength-pass filters, or a combination of both filters. Such filters prevent unwanted ambient light from reaching the detector.

References

- Carruthers J. B. “Wireless infrared communications” in “Wiley Encyclopedia of Telecommunications” edited by J. G. Proakis (John Wiley and Sons, New York, 2002)
- Heatley D. J. T., Wisely D. R., Neild I., and Cochrane P. “Optical wireless: the story so far” *IEEE Comm. Mag.* **36**, 72 (1998)
- Hecht J. “Understanding fiber optics” (Prentice Hall, Upper Saddle River, New Jersey, 2001)
- Kahn J. M. and Barry J. R. “Wireless infrared communications” *Proc. IEEE* **85**, 265 (2001)
- Keiser G. “Optical fiber communications” 3rd edition (McGraw-Hill, New York, 1999)
- Mynbaev D. K. and Scheiner L. L. “Fiber-optic communications technology” (Prentice Hall, Upper Saddle River, New Jersey, 2001)

Communication LEDs

LEDs can be used for either free-space communication or for fiber communication applications. *Free-space communication applications* include the remote control of appliances such as television sets and stereos, and data communication between a computer and peripheral devices. LEDs used in optical *fiber communication applications* are suited for distances of a few km and bit rates up to about 1 Gbit/s. Most fibers used with LEDs are multi-mode (step index and graded index) fibers. However, some applications employ single-mode fibers.

14.1 LEDs for free-space communication

Free-space communication LEDs are commonly made with GaAs or GaInAs active regions and are grown on GaAs substrates. The GaInAs layer is pseudomorphic, i.e. sufficiently thin so that it is coherently strained, and no dislocations are generated. The emission wavelength of GaAs and coherently strained GaInAs LEDs is limited to wavelengths in the IR ranging from 870 nm (for GaAs active regions) to about 950 nm (for GaInAs active regions).

The wavelength of free-space communication LEDs is in the infrared so that the light emitted is invisible to the human eye and does not distract. Since free-space communication usually involves transmission distances of less than 100 m, the transmission medium (air) can be considered, to a good approximation, to be lossless and dispersionless.

The *total light power* is an important figure of merit in free-space communication LEDs, so that the internal efficiency and the extraction efficiency need to be maximized. The emission pattern (far field) is another important parameter. The emission pattern should be wide to reduce the requirement of aiming the emitter towards the receiver.

14.2 LEDs for fiber-optic communication

Light-emitting diodes are the light source of choice for local area low and medium bit rate optical communication. Owing to the spontaneous emission lifetime of about 1 ns in highly excited semiconductors, the maximum bit rates attainable with LEDs is limited to rates

≤ 1 Gbit/s. Thus, multi-Gbit/s transmission rates are not feasible with LED sources. Transmission rates of several hundred Mbit/s are fully sufficient for many local-area communication applications.

LEDs used for fiber communication applications, are very different from LEDs used in lamp applications. In communication LEDs, high coupling efficiency of the light emanating from the LED to the fiber is essential. Only the light emanating from *one* surface, namely the surface of the LED abutting the optical fiber, can be coupled into the fiber. Therefore it is essential to maximize the light emission from *one* surface of the LED. In LEDs used for fiber communication, the *power emitted per unit area* is a useful figure of merit. This is in contrast to free-space communication LEDs where the *total power* emitted by the LED is the appropriate figure of merit.

In order to maximize the LED–fiber coupling efficiency, the light emitting spot should be smaller than the core diameter of the optical fiber. Typical are circular emission regions with diameters of 20 – 50 μm for devices used with multimode fibers. Silica multimode fibers have typical core diameters of 50 – 100 μm .

Plastic optical fibers, on the other hand, can have core diameters as large as 1 mm. Accordingly, LEDs with larger light-emitting areas can be used with plastic fibers.

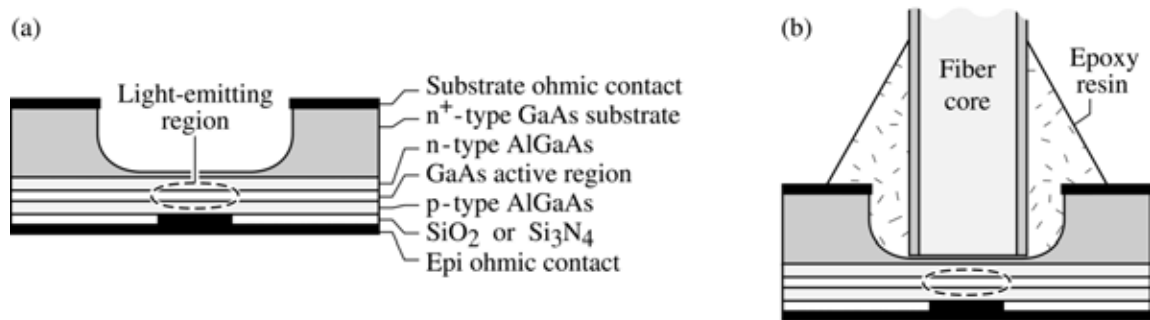


Fig. 14.1. (a) Burrus-type $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ DH LED with the opaque GaAs substrate removed above the active region by wet chemical etching. The Burrus-type LED is mounted substrate-side up. (b) Optical fiber coupled to a Burrus-type LED.

14.3 Surface-emitting Burrus-type communication LEDs emitting at 870 nm

One of the first LED structures suitable for optical fiber communication was developed by Charles Burrus of AT&T Bell Laboratories (Burrus and Miller, 1971; Saul *et al.* 1985). The Burrus-type LED is shown in Fig. 14.1 (a) and consists of a double heterostructure with a GaAs active region grown lattice-matched on a GaAs substrate. The original structure proposed and

demonstrated by Burrus was just a homojunction. However, such homojunction LEDs are no longer in use due to unwanted light reabsorption in the layers adjoining the active region.

The Burrus-type structure has several features making it suitable for communication applications. *First*, the light is generated in an active region of small lateral extent. The lateral size of the active region is determined by the p-type ohmic contact size of the LED. If the p-type confinement layer is sufficiently thin, no current spreading occurs, and the lateral extent of the light-emitting region is the same as the contact size. By design, the lateral extent of the active region is smaller than the core diameter of the optical fiber to maximize coupling efficiency. *Secondly*, the opaque GaAs substrate is partially removed, as shown in Fig. 14.1, to reduce absorption of light in the substrate. The substrate can be thinned by chemically assisted mechanical polishing to about 150 μm thickness followed by a wet chemical etch. If the GaAs membrane created by etching is too thin, it tends to break easily during fiber coupling. On the other hand, a thick membrane reduces radiative efficiency due to absorption.

Figure 14.1 (b) shows a typical coupling arrangement of the Burrus-type LED to an optical fiber. Epoxy is used to permanently attach the fiber to the LED. Note that the p-type contact also serves as a sink for the heat generated in the active region. The heat sink is particularly efficient if the p-type contact includes a thick layer of electroplated gold.

14.4 Surface-emitting communication LEDs emitting at 1300 nm

Communication LEDs emitting at 1300 nm, when used with graded-index silica fibers are suited for high-speed data transmission. A communication LED structure emitting at 1300 nm is shown in Fig. 14.2 (Saul *et al.*, 1985). The light is emitted through the InP substrate, which is transparent at 1300 nm. Accordingly, the device is mounted epi-side down in the LED package.

The device has a GaInPAs active region lattice-matched to the InP substrate. No current spreading layer is used so that the light-emitting region is located directly above the contact. At the light-exit point, an optical lens collimates the light to improve the LED-to-fiber coupling efficiency. The lens is etched into the InP substrate by a photo-electrochemical process (Ostermayer *et al.*, 1983).

The emission spectrum measured from the surface and from the edge of the device is shown in Fig. 14.3. The two emission spectra are markedly different. The spectrum emitted towards the edge has a smaller spectral width due to self-absorption. During self-absorption, predominantly photons of the high-energy part of the spectrum are reabsorbed.

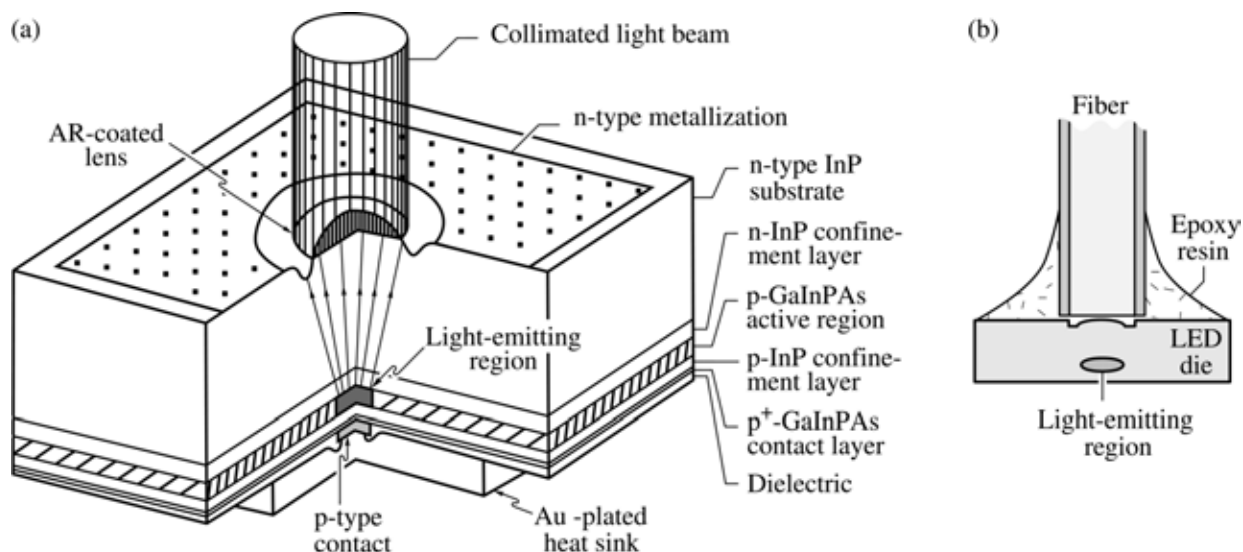


Fig. 14.2. (a) Structure of a communication LED emitting at 1300 nm with a GaInPAs active region lattice-matched to InP. The light generated in the active region is transmitted through the transparent InP substrate. The lateral dimension of the light-emitting region is defined by current injection under the circular ohmic contact with a diameter of 20 μm . An anti-reflection-coated (AR) lens, etched into the substrate, collimates the light beam. (b) Illustration of LED-to-fiber coupling using epoxy resin.

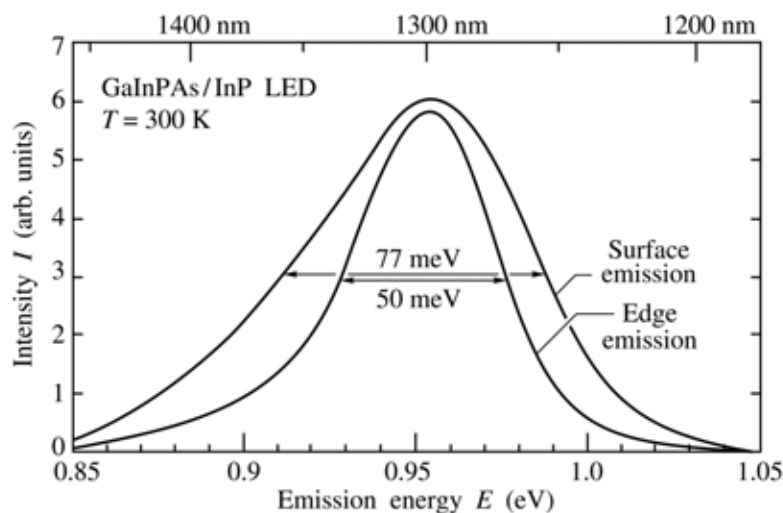


Fig. 14.3. Emission spectrum along the edge and surface of GaInPAs/InP communication LED emitting at 1300 nm. The spectrum emitted along the edge of the LED is narrower due to self-absorption.

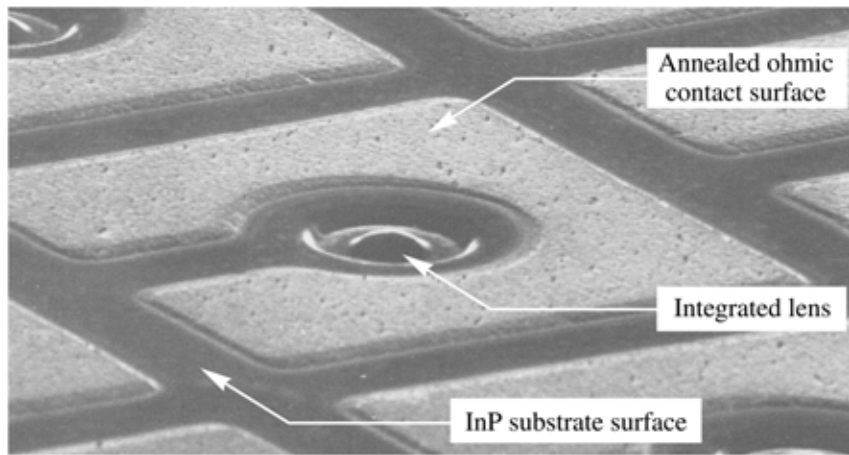


Fig. 14.4. GaInPAs communication LED grown on an InP wafer. The LED has an integrated semiconductor lens. The ohmic contact metal surface has a textured appearance due to the annealing process (after Ostermeyer *et al.*, 1983).

A scanning electron micrograph of a GaInPAs/InP LED wafer is shown in Fig. 14.4. The surface of contacts displays roughness due to the annealing process that follows contact deposition. To reduce Fresnel reflection losses at the semiconductor–air boundary, the lens is anti-reflection coated.

Note that the wavelength 1500 nm is of interest for long-distance silica-fiber communication. Long-distance communication fibers must have a small core diameter to be single mode and avoid modal dispersion. Consequently, LEDs emitting at 1500 nm are not used as sources for silica fiber communication.

14.5 Communication LEDs emitting at 650 nm

LEDs emitting at 650 nm are useful for plastic optical fiber communication. These fibers have a loss minimum and relatively low dispersion at 650 nm. Communication LEDs emitting in the 600 – 650 nm range are based on the $(\text{AlGa})_{0.5}\text{In}_{0.5}\text{P}$ material system just as for 650 nm visible-spectrum LED lamps.

Typical LED structures used for 650 nm plastic fiber communication applications are shown in Figs. 14.5 (a) and (b). Both LEDs are top-emitting devices due to the opaqueness of the GaAs substrate. The LEDs employ current-blocking layers that guide the current to the active region. Light emitted from this region is not obstructed by the metallic ring contact.

The figure shows two types of **current blocking regions**. Figure 14.5 (a) shows a current-blocking pnp structure in the transparent top layer. A Zn-diffused region in the center region of the LED overcompensates donors in the current-blocking layer making it p-type. As a result,

current flows only through the Zn-diffused layer. Figure 14.5 (b) shows a current-blocking layer fabricated by epitaxial regrowth. After growth of an n-type blocking layer on top of the p-type confinement layer, the wafer is taken out of the growth system for patterning and etching. Subsequently, the wafer is re-introduced into the growth system for resumption of epitaxial growth. Epitaxial regrowth is a more expensive process and usually the device yield decreases for structures employing regrowth. The transparent window layers can consist of AlGaAs, as shown in Fig. 14.5, GaP, or another transparent semiconductor.

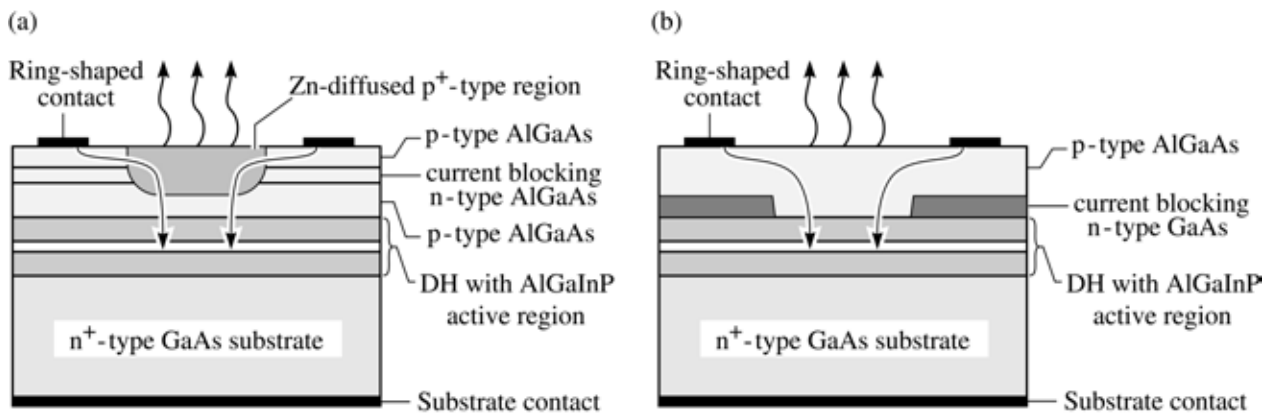


Fig. 14.5. AlGaInP/GaAs LED structures emitting at 650 nm for plastic optical fiber communications. Both LED structures funnel the current to the center of the active region where the emitted light is not obstructed by the top metal contact ring. (a) Structure using an n-type AlGaAs current blocking layer and a p⁺-type diffusion region. (b) Structure fabricated by epitaxial regrowth using an n-type GaAs current blocking layer.

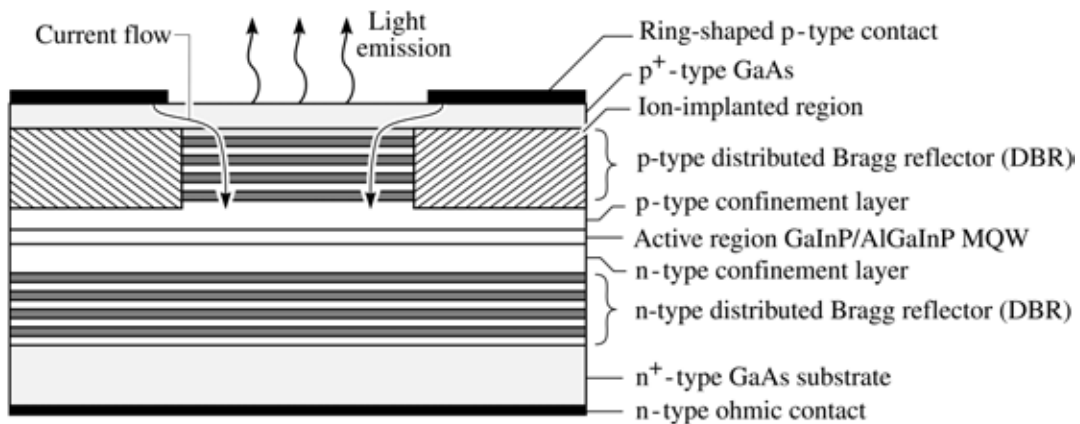


Fig. 14.6. RCLED emitting at 650 nm. Two distributed Bragg reflectors (DBRs) form the cavity. The active region is a GaInP/AlGaInP multiple-quantum well structure (after Whitaker, 1999).

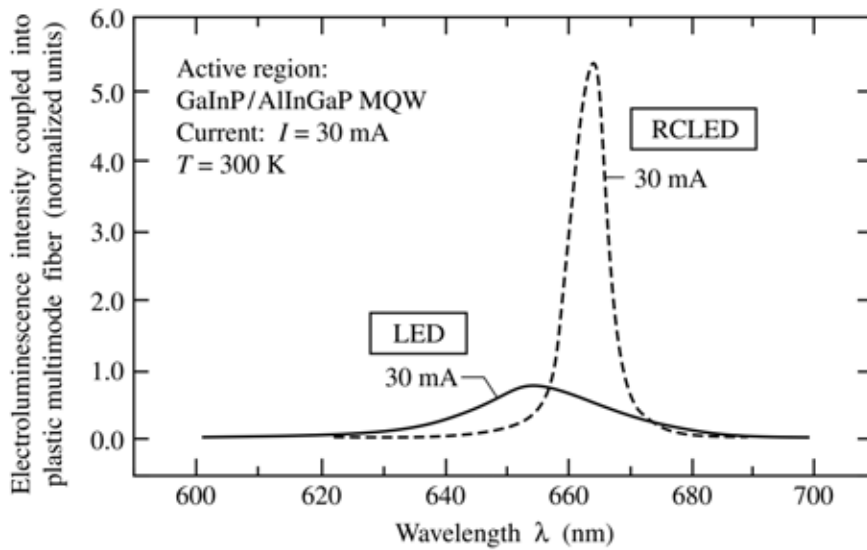


Fig. 14.7. Fiber-coupled ($NA = 0.275$) emission spectrum of RCLED and conventional LED at injection current of 30 mA. The microcavity effect of the RCLED enhances the emission intensity and reduces the emission linewidth, especially for low NA fibers (after Whitaker, 1999).

Resonant-cavity LEDs (RCLEDs) have several advantages over conventional LEDs including high brightness and a narrow spectral width (Schubert *et al.*, 1992, 1994; Schubert and Hunt, 1999). An RCLED emitting at 650 nm is shown in Fig. 14.6 (Streubel and Stevens, 1998; Streubel *et al.*, 1998; Mitel, 2000). As in the two previous structures, the top contact is ring-shaped and the current is guided to the center opening of the ring. An ion-implanted region is used for current blocking. Hydrogen and oxygen have been employed to render the implanted region insulating. Oxygen implants are more stable than hydrogen implants since small hydrogen atoms tend to easily diffuse out of the semiconductor at moderate annealing temperatures.

The emission spectra of an RCLED and a conventional LED are shown in Fig. 14.7. The spectra shown are the spectra of the fiber-coupled light. Inspection of the spectra reveals two features. *First*, the RCLED exhibits a higher spectral purity thereby reducing chromatic dispersion. *Secondly*, the fiber-coupled intensity of the RCLED is higher due to the more directed emission pattern. High speed transmission of 250 Mbit/s over plastic optical fibers have been demonstrated with 650 nm RCLEDs (Streubel and Stevens, 1998).

14.6 Edge-emitting superluminescent diodes (SLDs)

Edge-emitting LEDs are motivated by the need for high-brightness LEDs that allow for high-efficiency coupling to optical fibers. Edge-emitting LEDs comprise an optical waveguide region that guides light emitted along the waveguide by total internal reflection.

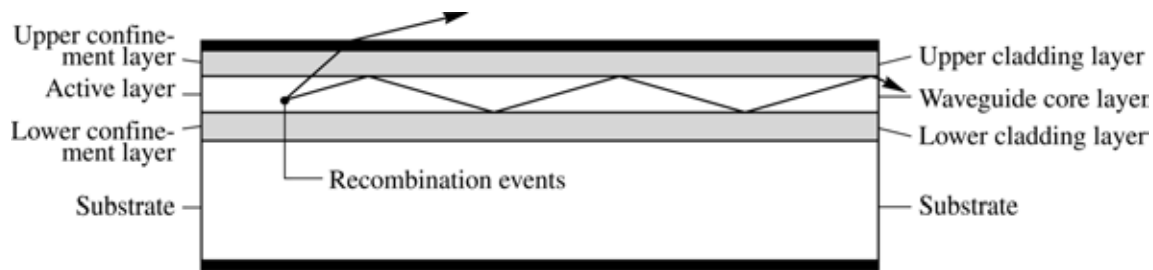


Fig. 14.8. Waveguide geometry showing guided light rays in the core layer with low angles of incidence.

Superluminescent diodes are broad-band high-intensity emission sources that emit incoherent light. Incoherent light does not result in “speckle patterns” obtained from coherent light sources such as lasers. SLDs are suitable as communication devices used with single-mode fibers and also as high-intensity light sources for the analysis of optical components (Liu, 2000).

Light is guided in the *core region* of the waveguide. Total internal reflection occurs at the boundaries between the core region and the upper and lower *cladding layers* as shown in Fig. 14.8. In order to make waveguiding possible, the core layer must have a higher refractive index than the cladding layers. Photons emitted with a sufficiently small angle of incidence at the core–cladding interface will be guided by the waveguide, as indicated in Fig. 14.8.

Since the light is guided in the waveguide, the light intensity emitted by the LED increases linearly with the length of the waveguide. Thus increasing the length of an edge-emitting LED allows one to obtain a higher light output intensity. However, the electrical current required to drive the LED also increases with the stripe length.

Superluminescent light-emitting diodes or *superluminescent diodes* (SLDs) are edge-emitting LEDs that are pumped at such high current levels that *stimulated emission* occurs. In the stimulated emission process, one photon stimulates the recombination of an electron–hole pair and the emission of another photon. The photon created by the stimulated emission process has the same propagation direction, phase, and wavelength as the stimulating photon. Thus SLDs have greater coherence compared with LEDs. In the stimulated emission regime, spontaneous emission towards the top surface of the LED is reduced and emission into waveguide modes is enhanced.

SLDs are quite similar to semiconductor laser structures with one important difference: SLDs lack the optical feedback provided by the reflectors of a semiconductor laser. Two typical SLDs structures are shown in Fig. 14.9. The SLD structure shown in Fig. 14.9 (a) has a reflective backside reflector facet; the front-side facet, however, is coated with an anti-reflection (AR)

coating. To prevent lasing, the front-side facet must have a reflectivity of $\leq 10^{-6}$ (Liu, 2000; Saul *et al.*, 1985). Exceeding the required reflectivity results in unwanted lasing of the device. Owing to the high-quality AR coating requirement, SLDs with an AR coating are expensive to manufacture.

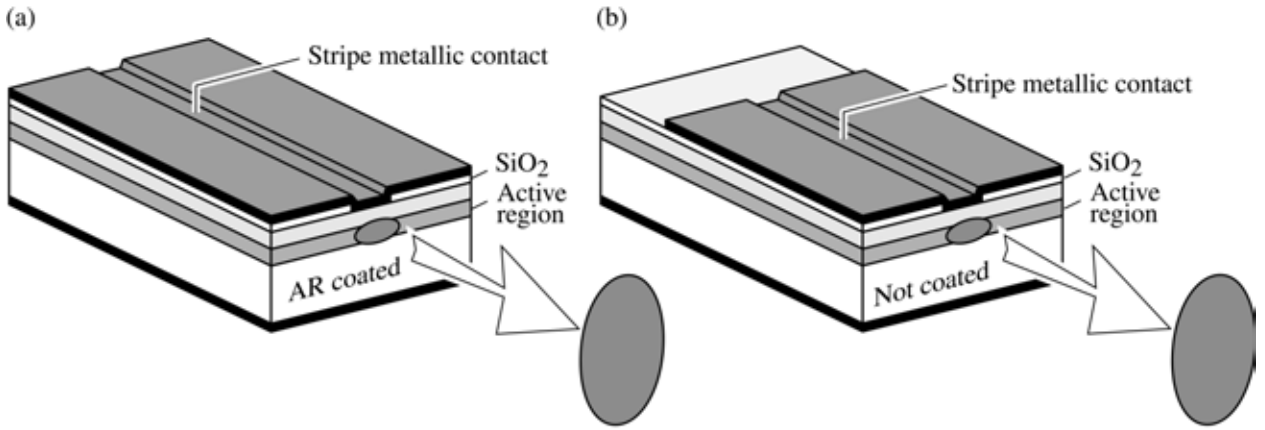


Fig. 14.9. Common structures of superluminescent diodes (SLDs). (a) SLD with cleaved facets coated with anti-reflection (AR) coatings. (b) SLD with cleaved, reflecting facets and stripe contact injecting current over the partial length of the device.

A lower-cost alternative SLD structure is shown in Fig. 14.9 (b). This structure uses a *lossy region* near the backside facet of the diode. The lossy region is not covered by the top metal contact and thus is not pumped by the injection current. Practically no feedback occurs, if the length of the lossy region is much longer than the absorption length of the core region, i.e.

$$\text{length of lossy region} \gg \alpha^{-1}, \quad (14.1)$$

where α is the absorption coefficient of the core region. The absorption coefficient in III–V semiconductors near the band edge is $\alpha \approx 10^4 \text{ cm}^{-1}$. Thus for lossy-region lengths exceeding several tens of micrometers, the optical feedback from the backside facet is negligibly small.

In addition to absorption losses, diffraction losses occur in the region not pumped by the electrical current. Gain guiding occurs in the region injected by the electrical current but not in the lossy region. Thus both, absorption and diffraction losses, prevent this type of SLD from lasing.

Emission spectra of an LED, SLDs and a laser are shown in Fig. 14.10. The LED has a broad spontaneous emission spectrum. The spectrum of an SLD with a residual small facet reflectivity

exhibits periodic oscillations in the emission spectrum due the Fabry–Perot cavity enhancement. An ideal SLD has a smooth spectrum and does not exhibit any oscillations. The spectral width of SLDs is narrower than that of LEDs due to increased coherence caused by stimulated emission. Also shown is the spectrum of a Fabry–Perot laser with several laser modes.

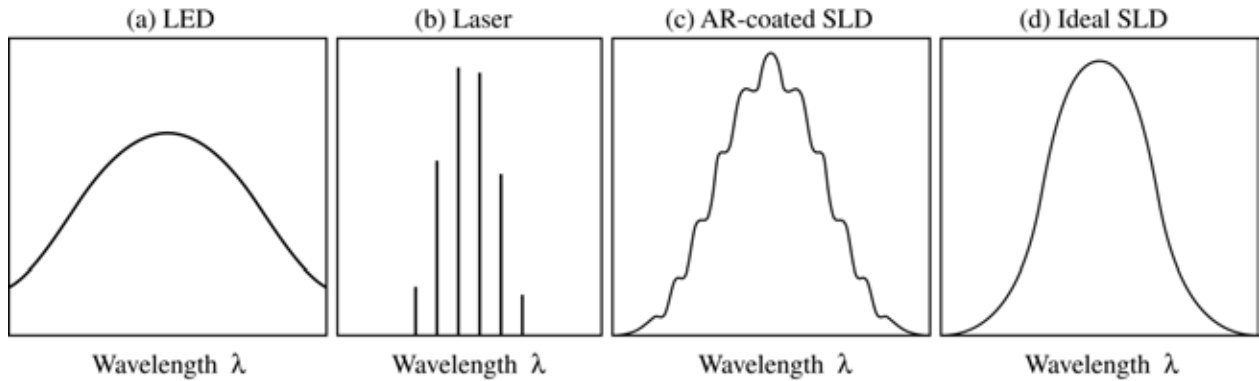


Fig. 14.10. Spectra of (a) a regular LED with a theoretical linewidth of $1.8 kT$, (b) a multi-mode semiconductor laser, (c) a superluminescent diode (SLD) fabricated by AR coating of a multimode laser, (d) an ideal SLD with linewidth less than kT (after Liu, 2000).

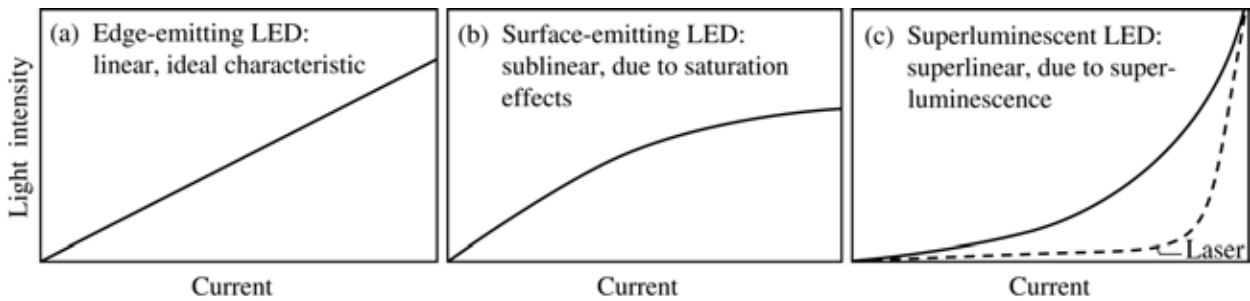


Fig. 14.11. Light-versus-current ($L-I$) characteristic of different LEDs. (a) Edge-emitting LED with little or no saturation effects. (b) Surface-emitting LED with small active area exhibiting saturation effects due to carrier overflow. (c) Superluminescent LED. Also shown is the $L-I$ characteristic of a laser that exhibits a distinct threshold current.

A comparison of the $L-I$ curves of an LED, an SLD, and a laser is shown in Fig. 14.11. Surface-emitting LEDs with a small light-emitting region diameter, tend to have **sublinear** $L-I$ characteristics. At high injection current densities, the small active volume of surface-emitting LED are swamped with carriers leading to saturation. Edge-emitting LEDs operating in the spontaneous emission regime have **linear** $L-I$ characteristics, as expected for ideal LEDs. SLDs have a **superlinear** $L-I$ characteristic due to stimulated emission. In the stimulated emission

regime, an increasing number of photons are guided by the waveguide. The number of photons emitted into waveguide modes increases with injection current as stimulated emission becomes dominant. As for SLDs, semiconductor lasers have superlinear emission characteristics. However, the L – I curve of lasers exhibits a more distinct threshold than that of SLDs.

References

- Burrus C. A. and Miller B. I. “Small-area double heterostructure AlGaAs electroluminescent diode sources for optical fiber transmission lines” *Optical Communication* **4**, 307 (1971)
- Liu Y. “Passive components tested by superluminescent diodes” February issue of *WDM Solutions* p. 41 (2000)
- Mitel Corporation. RCLEDs were first manufactured by the Mitel Corporation. RCLED emitting at 650 nm has Part # 1A466. (2000). Large scale production of RCLEDs was started in 2001 by Osram Optosemiconductors Corporation, *see* Wirth R., Huber W., Karnutsch C., and Streubel K. “High-efficiency resonant-cavity LEDs emitting at 650 nm” *Compound Semiconductors* **8**, 49 (2002)
- Ostermayer Jr. F. W., Kohl P. A., and Burton R. H. “Photoelectrochemical etching of integral lenses on GaInPAs/InP light-emitting diodes” *Appl. Phys. Lett.* **43**, 642 (1983)
- Saul R. H., Lee T. P., and Burrus C. A. “Light-emitting-diode device design” in “Lightwave communications technology” edited by W. T. Tsang *Semiconductors and Semimetals* Vol. **22** Part C, (Academic Press, San Diego, 1985)
- Schubert E. F. and Hunt N. E. J. “Enhancement of spontaneous emission in microcavities” in “Vertical cavity surface emitting lasers” edited by C. Wilmsen, H. Temkin, and L. A. Coldren (Cambridge University Press, Cambridge, UK, 1999)
- Schubert E. F., Wang Y.-H., Cho A. Y., Tu L.-W., and Zydzik G. J. “Resonant-cavity light-emitting diode” *Appl. Phys. Lett.* **60**, 921 (1992)
- Schubert E. F., Hunt N. E. J., Micovic M., Malik R. J., Sivco D. L., Cho A. Y., and Zydzik G. J. “Highly efficient light-emitting diodes with microcavities” *Science* **265**, 943 (1994)
- Streubel K. and Stevens R. “250 Mbit/s plastic fibre transmission using 660 nm resonant cavity light emitting diode” *Electronics Lett.* **34**, 1862 (1998)
- Streubel K., Helin U., Oskarsson V., Backlin E., and Johansson A. “High brightness visible (660 nm) resonant-cavity light-emitting diode” *IEEE Photonics Technol. Lett.* **10**, 1685 (1998)
- Whitaker T. “Resonant Cavity LEDs” *Compound Semiconductors* **5** (4), 32 (May 1999)

15

LED modulation characteristics

LEDs are the most commonly used light source for local-area communication systems operating from very short (< 1 m) to medium distances (5 km). Typical bit rates are several 10 Mbit/s up to about 1 Gbit/s.

LEDs are non-linear devices and as such the series resistance, shunt resistance, and capacitance depend strongly on the applied voltage. A thorough analysis must take into account these non-linearities. However, much can be learned by considering the LED as a linear device, even though some important aspects of the LED modulation behavior cannot be inferred from the linear model. In this chapter, we first analyze LEDs as linear devices and subsequently discuss modulation characteristics, including some non-linear modulation characteristics.

15.1 Rise and fall times, 3 dB frequency, and bandwidth in linear circuit theory

A simple RC circuit is shown in Fig. 15.1 (a). When subjected to a step-function input pulse, the output voltage increases according to

$$V_{\text{out}}(t) = V_0 [1 - \exp(-t/\tau_1)] \quad (15.1)$$

where $\tau_1 = RC$ is the time constant of the RC circuit. When the input voltage returns to zero, the output voltage decreases according to

$$V_{\text{out}}(t) = V_0 \exp(-t/\tau_2) . \quad (15.2)$$

For an RC circuit, it is $\tau_1 = \tau_2$. The rise and fall times are defined as the time difference between the 10 % and 90 % points of the voltage, as shown in Fig. 15.1 (b). The rise and fall time of the signal are related to the time constants τ_1 and τ_2 by

$$\tau_r = (\ln 9) \tau_1 \approx 2.2 \tau_1 \quad \text{and} \quad \tau_f = (\ln 9) \tau_2 \approx 2.2 \tau_2 . \quad (15.3)$$

The *voltage* transfer function $H(\omega)$ is given by

$$H(\omega) = (1 + i \omega \tau)^{-1} . \quad (15.4)$$

The **bandwidth** of the system, Δf , corresponds to the frequency at which the power transmitted through the system is reduced to half of its low-frequency value. This condition can be written as $|H(2\pi f)|^2 = 1/2$. Thus the bandwidth of the RC circuit is given by

$$\Delta f = f_{3\text{dB}} = \frac{1}{2\pi\tau} = \frac{\ln 9}{2\pi\tau_r} = \frac{\ln 9}{\pi(\tau_r + \tau_f)} \approx \frac{0.70}{(\tau_r + \tau_f)} . \quad (15.5)$$

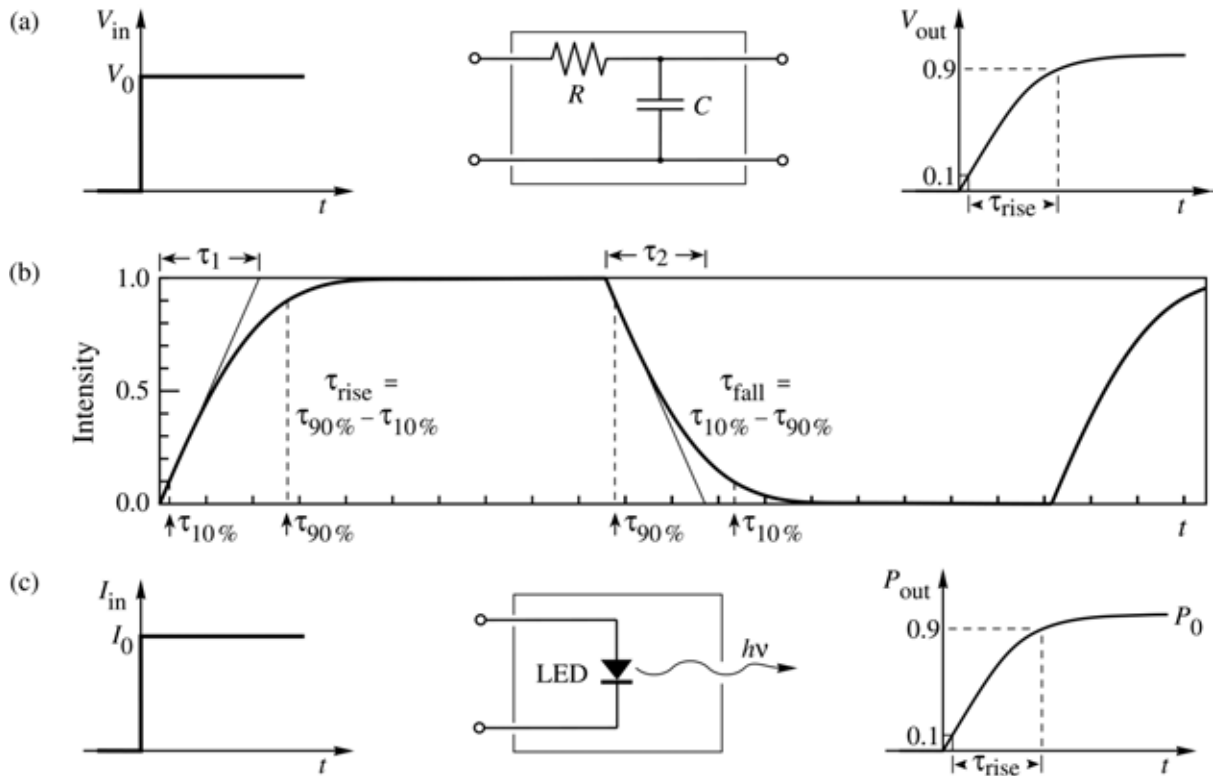


Fig. 15.1. (a) Illustration of the system response of a linear RC system with rise time τ_r . (b) Rise and fall time of a signal with an exponential time dependence and the time constants τ_1 and τ_2 . (c) Illustration of the light output power as a function of time for an LED with a rise time of τ_r .

The bandwidth is also called the **3 dB frequency**, since the power transmitted at this frequency is reduced by 3 dB compared with its low-frequency value.

Next, consider an LED with a rise time τ_r as illustrated in Fig. 15.1 (c). As a step-function input current is applied to the LED, the optical output power increases according to

$$P_{\text{out}}(t) = P_0 [1 - \exp(-t/\tau_r)] . \quad (15.6)$$

In analogy to Eqs. (15.1) and (15.4), the *power* transfer function is given by

$$H_{\text{LED}}^2(\omega) = (1 + i\omega\tau)^{-1} . \quad (15.7)$$

The absolute value of the power transfer function is reduced to half at the 3 dB frequency of the LED. Thus the 3 dB frequency of an LED is given by

$$\Delta f = f_{3\text{dB}} = \frac{\sqrt{3}}{2\pi\tau} = \frac{\sqrt{3} \ln 9}{2\pi\tau_r} = \frac{\sqrt{3} \ln 9}{\pi(\tau_r + \tau_f)} \approx \frac{1.2}{(\tau_r + \tau_f)} . \quad (15.8)$$

Exercise. Derivation of equations. Derive Eqs. (15.3), (15.4), (15.5) and (15.8).

Exercise. Rise and fall time and 3 dB frequency. Consider an LED with a rise time of 1.75 ns. Assume that the fall time of the LED is identical to the rise time. What is the 3 dB frequency of the device? Give the physical reasons as to why Eq. (15.8) gives only an *approximate* value of the 3 dB frequency.

Solution: A 3 dB frequency of 343 MHz is expected on the basis of Eq. (15.8). In practice the 3 dB frequency can be lower or higher than the calculated value since the rise and fall are frequently not exponential. As a practical rule, the numerical factor 1.2 in the numerator of Eq. (15.8) can vary between 1.0 and 1.5.

15.2 Rise and fall time in the limit of large diode capacitance

In diodes used for solid-state lamp applications, the current-injected p-n junction area is large, sometimes as large as the entire LED die. Such diodes have a *large* capacitance. Denoting the diode capacitance as C and the overall series resistance of the drive circuit and the diode as R ,

the rise and the fall time of the diode are equal and these times are given by the RC time constant.

In communication LEDs, the current-injected active region is much smaller, so that the spontaneous lifetime rather than the diode capacitance limits the maximum modulation frequency. As a result, communication LEDs can be modulated. Since LEDs do not exhibit strictly exponential changes in power, as predicted by Eq. (15.6), Eq. (15.8) is only an approximation.

Consider an LED in which the p-n junction region extends over the entire area of the die. The LED has a small contact area that determines the size of the light-emitting spot. Such LEDs are used for communication applications. At zero bias, the capacitance of the LED is given by the *depletion* capacitance (space-charge capacitance) of the diode. Since the area of the diode is large (e.g. $250\ \mu\text{m} \times 250\ \mu\text{m}$), the capacitance is large and can amount to 200 – 300 pF.

As the diode is turned on, the current *crowds* in the area below the contact. When the p-n junction is forward biased, the capacitance of the LED is given by the *diffusion* capacitance. The relevant area is, however, not the entire LED die but just the region injected with the diode current.

The reduction of the diode capacitance increases the LED modulation bandwidth. The *depletion* capacitance can be reduced by mesa etching. However, the mesa should be larger than the contact size in order to avoid surface recombination effects.

There is no viable way to reduce the *diffusion* capacitance. The diffusion capacitance can be reduced by purposely introducing defects that act as luminescence killers. Such defects reduce the minority carrier lifetime and thereby the diffusion capacitance. Such LEDs can be modulated at several GHz. However, the light output intensity decreases as well so that the overall benefit of such lifetime killers is questionable.

15.3 Rise and fall time in the limit of small diode capacitance

Next, we discuss the rise and fall time in the limit of small diode capacitance. This consideration applies, for example, to surface-emitting communication LEDs that have a small-area active region. Consider an LED driven by a constant current that is switched on at $t = 0$. Electrons are injected into the active region and the carrier concentration builds up. At the same time, the optical output intensity of the LED increases. In the case of the monomolecular recombination model, the light output intensity is directly proportional to the injected minority carrier concentration.

The monomolecular rate equation is given by

$$\frac{I}{e A d} = \frac{dn_a}{dt} = \frac{n_a}{\tau} \quad (15.9)$$

where n_a is the carrier concentration in the active region, A is the current-injected area of the active region, and d is the thickness of the active region. The steady-state current flow of magnitude I causes a steady-state minority carrier concentration $n_a = I \tau / (e A d)$. The mean lifetime of the carriers is the spontaneous recombination lifetime τ .

Next, consider that the diode is initially in the “off” state and, starting at $t = 0$, the diode is injected with a constant current I . When the diode is in the “off” state, the minority carrier concentration in the active region is very low and we approximate the concentration with $n_a \approx 0$.

Solving the differential Eq. (15.9) for the initial condition $n_a \approx 0$ at $t = 0$ for a constant injection current yields that the carrier concentration in the active region increases according to

$$n_a(t) = n_a \left(1 - e^{-t/\tau} \right) = \frac{I \tau}{e A d} (1 - e^{-t/\tau}) . \quad (15.10)$$

The equation reveals that it takes the spontaneous recombination time τ to fill the active region with the steady-state carrier concentration. The light output intensity follows the minority carrier concentration directly. Thus, the rise time is given by the spontaneous recombination time.

A similar consideration applies to the fall time of the diode. Once the diode has been switched off, the decay constant for emission is, of course, the spontaneous recombination lifetime. Thus, the fall time of an LED is given by the spontaneous recombination lifetime.

In the case of an undoped active region, the monomolecular recombination equation no longer applies and the bimolecular recombination equation must be used to describe the carrier dynamics. Also, the carrier lifetime is no longer a constant. In this case, the *shortest* lifetime, i.e. the lifetime that applies when the carrier concentration is at the *highest* level, can be used.

It should be noted that there are methods to *reduce* both the rise and the fall time *below* the limit of the spontaneous recombination lifetime. The rise time can be reduced by *current shaping*. The fall time can be reduced by *carrier sweep-out*. Both methods will be discussed below.

15.4 Voltage dependence of the rise and fall time

The measurement of the rise time and fall time of an LED is shown in Fig. 15.2 (Schubert *et al.*, 1996). As indicated in the figure, the rise and fall time are measured from the 10 – 90 % values of the optical signal. The photocurrent of a p-n junction photodetector is used in the measurement. It must be ensured in the measurement that both the pulse generator and the photodiode are much faster than the LED rise and fall time. The measured rise and fall times include the time constant of the pulse generator, the LED, the detector, detector amplifier circuit, and the oscilloscope. However, the time constant of the LED is the longest and hence the dominant time constant. The time constants shown in Fig. 15.2 are *upper limits* to the true time constants of the LED.

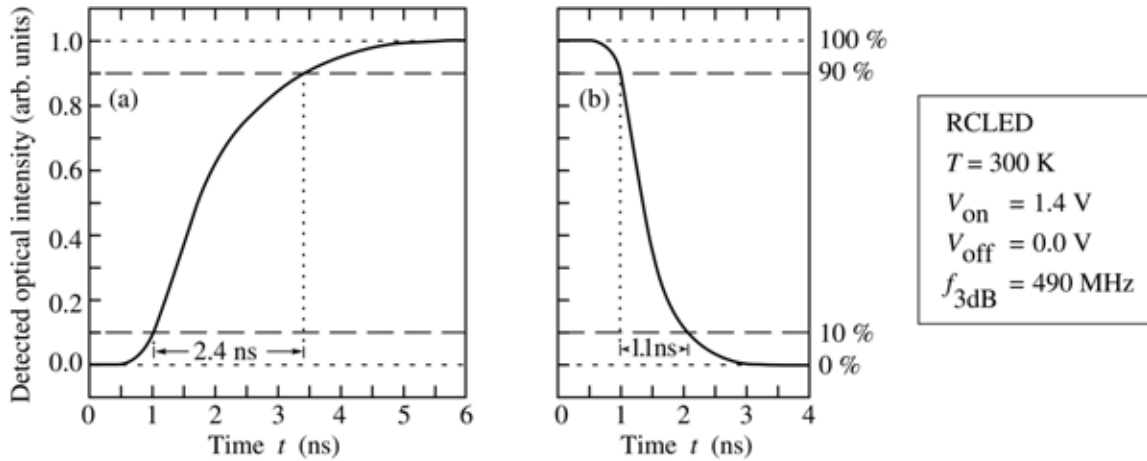


Fig. 15.2. Measured (a) rise time ($\tau_{\text{rise}} = 2.4 \text{ ns}$) and (b) fall time ($\tau_{\text{fall}} = 1.1 \text{ ns}$) of an RCLED. The photodetector used in the measurement is much faster than the LED, so that the measured times are essentially the rise and fall times of the LED.

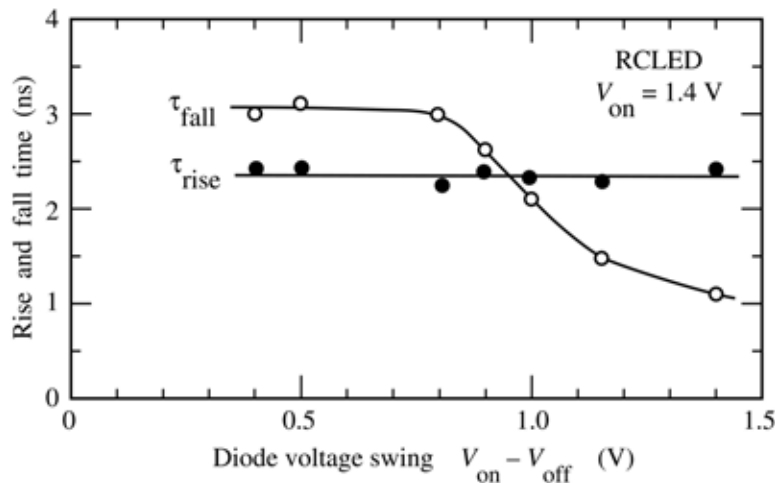


Fig. 15.3. Rise time and fall times as a function of the voltage swing. The fall time of the diode decreases with increasing voltage swing due to sweep-out of carriers out of the active region.

Inspection of Fig. 15.2 reveals that the rise time is much longer than the fall time. The large difference between the rise and the fall time displayed in Fig. 15.2 is not expected based on the theoretical model discussed above.

To gain a better understanding of the difference between the rise and the fall time, the times have been measured as a function of the diode bias conditions. In the “on” state, the diode is biased with a voltage of 1.4 V. However, a range of voltages can be chosen for the “off” state, since a p-n junction diode does not emit light for voltages even slightly below the turn-on voltage.

The experimental results are shown in Fig. 15.3. Whereas the “on” voltage is kept constant at $V_{\text{on}} = 1.4$ V, the “off” voltage is varied from 0 to 1.0 V. Inspection of Fig. 15.3 reveals a strong voltage dependence of the fall time. The voltage dependence is caused by carrier sweep-out of the active region. In contrast, the rise time is practically independent of voltage.

15.5 Carrier sweep-out of the active region

The voltage dependence of the fall time shown in Fig. 15.3 can be explained by voltage-dependent carrier sweep-out of the active region. Figure 15.4 shows the active region band diagram in the “off” state for small (a) and large (b) voltage swings. For the cases of a small voltage swing, carriers essentially remain in the active region until they recombine. As a result, it will take the spontaneous lifetime for the carriers to recombine and the light intensity to decay.

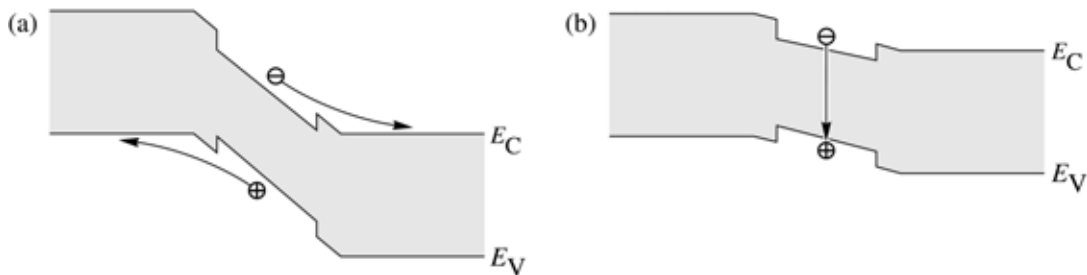


Fig. 15.4. Illustration of two mechanisms determining the fall time. (a) Return to zero bias results in carrier sweep-out of the active region. The sweep-out time can be very short, $\ll 1$ ns. (b) If the modulating voltage amplitude is smaller, carriers are *not* swept out of the active region, so that the intensity decay is determined by the spontaneous recombination lifetime.

The situation is quite different for large voltage swings. At zero bias, the band diagram of the active region is highly sloped due to the built-in electric field of the p-n junction. As a result, free carriers are swept out of the active region into the neutral n- and p-type confinement regions of

the semiconductor. The carrier sweep-out is most efficient for large voltage swings, i.e. when a high electric field is created in the space-charge region of the p-n junction. The sweep-out time can be much shorter than the spontaneous lifetime. Thus, the fall time is determined *not* by the spontaneous recombination lifetime but by the shorter sweep-out time. Considering the magnitude of the built-in electric field and the carrier mobility, the sweep-out time is estimated to be in the picosecond range.

Exercise. Calculation of carrier sweep-out time. Calculate the carrier sweep-out time for typical values of the electric field in the p-n junction depletion region, typical carrier velocity, and an active region thickness of $0.1 - 1 \mu\text{m}$.

Solution: The carrier sweep-out time can be very short. For typical diode parameters, the carrier sweep-out time is about $1 - 100 \text{ ps}$, i.e. much shorter than the spontaneous recombination time.

15.6 Current shaping

A common method to reduce the rise and the fall time is *current shaping* (Lee, 1975; Zucker, 1978; Saul *et al.*, 1985). The diagram of a current-shaping circuit is shown in Fig. 15.5 (a). The current-shaping circuit is essentially a capacitor and a resistor in series with the LED. The capacitor creates a current transient when the LED is switched on or off, as shown in Fig. 15.5 (b). During the switch-on transient, the excess current flowing through the capacitor helps in reaching the steady-state carrier concentration in the active region within a time shorter than the spontaneous lifetime. During the switch-off period ($V = 0$), the capacitor biases the diode in the reverse direction, thus aiding the current sweep-out of the active region. Parameters entering the current transient include the power supply internal resistance, the resistor and capacitor of the current-shaping circuit, and the diode series resistance.

A reasonable design criterion for the current-shaping circuit is that the RC time constant of the circuit introduced by the RC current-shaping circuit should be equal to the spontaneous recombination lifetime and that the initial current at the beginning of the voltage pulse should be twice the magnitude of the steady-state current. In this case, the value of the resistor R of the current-shaping circuit is chosen to equal the differential resistance of the diode. The capacitance of the circuit can be chosen so that the RC time constant of the entire circuit coincides with the LED rise time (Schubert *et al.* 1996). As a result of the current-shaping circuit, the 3 dB bandwidth increases.

Note, that the diode series resistance is strongly voltage-dependent, and therefore linear circuit theory can only provide *estimates* rather than accurate values. Therefore, experimental or numerical methods are required to optimize the RC current-shaping circuit.

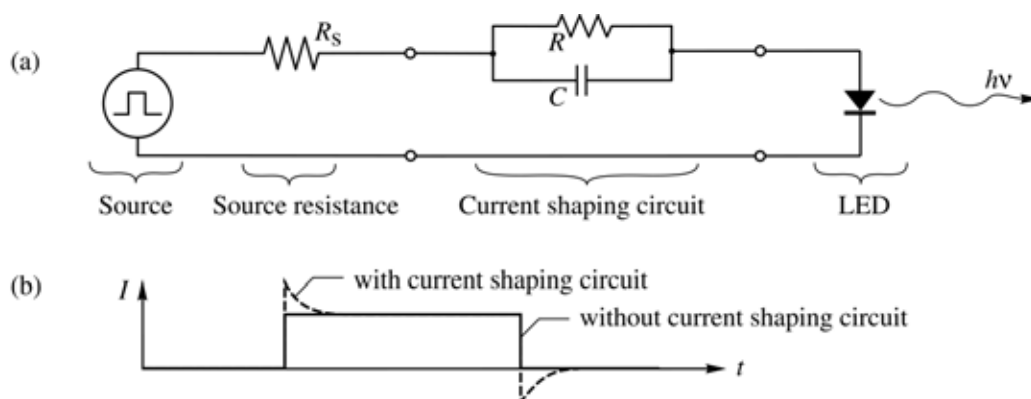


Fig. 15.5. (a) Illustration of an RC “current-shaping circuit” used to decrease the rise time of LEDs. (b) Diode current pulse versus time in the limit of small diode capacitance (solid line) and effect of current shaping circuit on diode current (dashed line).

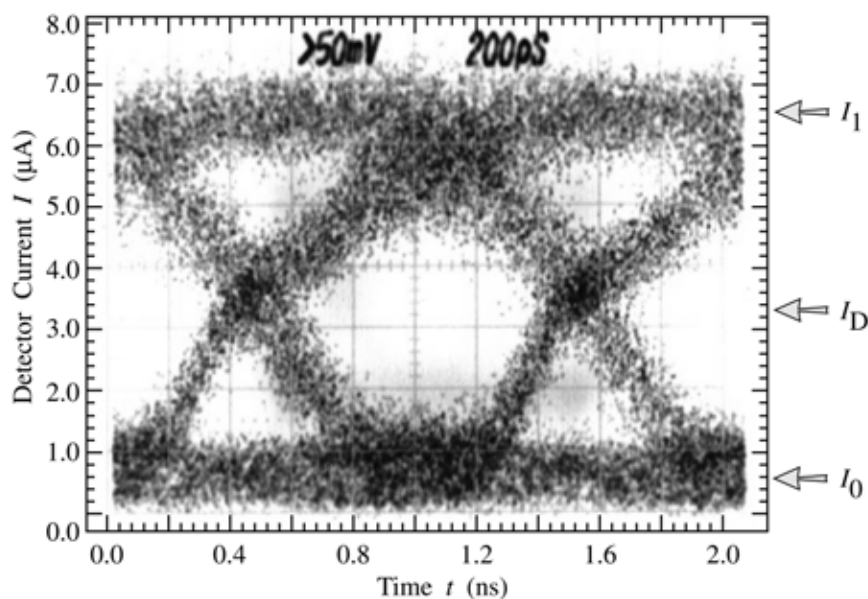


Fig. 15.6. “Eye diagram” of the received optical signal of an RCLED. The optical signal is measured as the photocurrent of a Si photodiode measured with a sampling oscilloscope. The data rate of the RCLED is 622 Mbit/s. Also indicated are the three current levels for the “0” state (I_0), the “1” state (I_1), and the decision current (I_D).

The current-shaping circuit requires an increased operating voltage and this reduces the overall efficiency of the drive circuit. However, the power efficiency is usually of little relevance due to the low overall power consumption of communication LEDs.

15.7 3 dB frequency

The 3 dB frequency of an LED can be determined by measuring the frequency at which the detector signal decreases to one-half of the low-frequency value. The frequency response of the detector needs to be taken into account in the measurement. A 3 dB frequency of about 500 MHz was determined for the LED for which the rise and fall times are shown in Fig. 15.2. Comparison of the 3 dB frequencies for different LED structures revealed that the highest 3 dB frequencies are attained with devices having low parasitic resistances and capacitances. This can be achieved by a small p-type contact area, a thick SiO₂ isolation layer on the p-type side of the device, a small area of the bonding pad on the p-type side of the device, and a mesa etch limiting the p-n junction depletion capacitance at zero bias.

15.8 Eye diagram

The *eye diagram* allows one to estimate the overall performance of an optical communication system. The eye diagram is the receiver signal of a randomly generated digital signal. An eye diagram of an LED operating at 622 Mbit/s is shown in Fig. 15.6. A data rate of 622 Mbit/s is used in the well-known “synchronous optical network” (SONET) standard. The figure indicates the level of the “1” state and of the “0” state and of the *decision level*, i.e. the boundary between what is interpreted by the receiver as a “0” and “1”. The figure also reveals the “eye”. An “open eye” such as the one shown in the figure, allows for a low bit-error rate. The “eye” shown in Fig. 15.6 is wide open, indicating that low bit-error-rate data transmission is possible at that frequency using LEDs. The “on” and “off” pulse-generator voltages of the diode were 1.4 and 1.1 V, respectively. A pulse-shaping RC circuit with $R = 20\ \Omega$ and $C = 100\ \text{pF}$ was used for the measurement. Minimizing parasitic elements (e.g. bond pad capacitance) and employment of a RC pulse shaping circuit should make transmission rates of 1 Gbit/s possible.

As the data rate is increased, the eye will close, i.e. the photocurrent of the “0” and “1” level cannot be clearly distinguished. This results in an increase in the bit-error rate.

15.9 Carrier lifetime and 3 dB frequency

Shortening the minority carrier lifetime through either very high doping of the active region or

deliberate introduction of deep traps will increase the maximum modulation frequency. Deep traps have a two fold effect. *First*, they reduce the minority carrier lifetime, thereby increasing the 3 dB frequency. *Secondly*, they reduce the emission intensity and increase the heat generated inside the LED. Very high concentrations of a shallow dopant, on the other hand, will shorten the carrier lifetime and often, but not necessarily, hurt the device efficiency.

We therefore analyze the effect of lifetime reduction on emission intensity and 3 dB frequency. According to Eq. (15.8), the 3 dB frequency of an LED depends on the radiative lifetime according to $f_{3\text{ dB}} = (3)^{1/2} / 2 \pi \tau$, where $\tau^{-1} = \tau_r^{-1} + \tau_{nr}^{-1}$. The internal device efficiency is given by $\eta_{\text{int}} = \tau_{nr} / (\tau_r + \tau_{nr})$. In the limit of small non-radiative lifetimes, it is $f_{3\text{ dB}} \propto \tau_{nr}^{-1}$ and $\eta_{\text{int}} \propto \tau_{nr}$. Thus, although the modulation bandwidth *can* be increased by the deliberate introduction of deep traps, the power-bandwidth product *cannot*. The relation between 3 dB frequency, output power, and radiative and non-radiative lifetime is shown in Fig. 15.7. The 3 dB frequencies and intensity levels are calculated according to the equations stated above. At frequencies much higher than the 3 dB frequency, a linear decrease of the optical intensity on the log-log scale can be assumed (Wood, 1994).

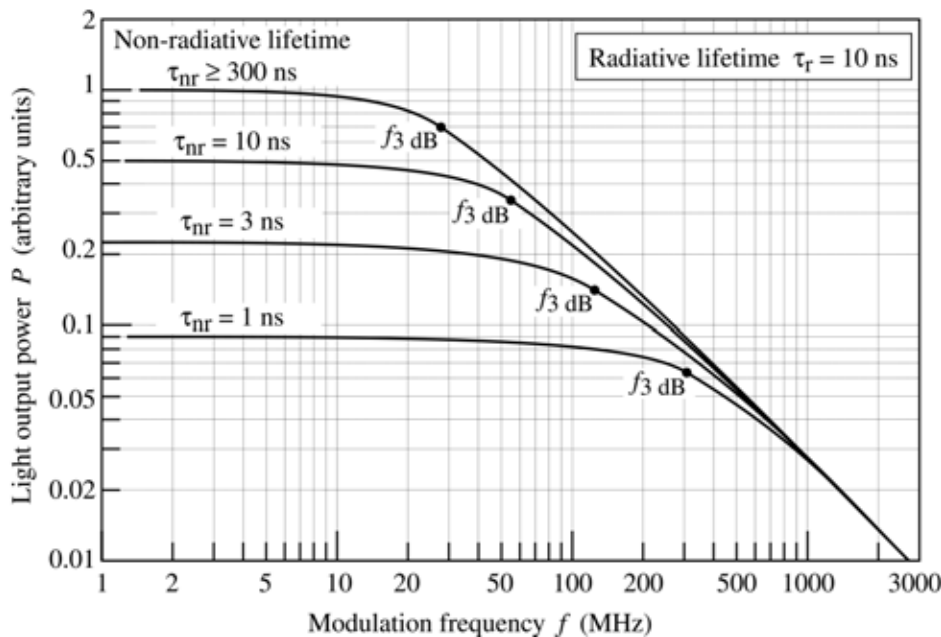


Fig. 15.7. Calculated LED output power versus modulation frequency for different values of the non-radiative lifetime. A radiative lifetime of 10 ns is assumed.

If, however, one were to succeed in decreasing the minority carrier lifetime without affecting the efficiency, the modulation speed of an LED could be increased. Although progress has been made in this direction, the increase has up to now been accompanied by a decrease in device efficiency (Chen *et al.*, 1999).

Chen *et al.* (1999) proposed very high Be doping ($N_A = 2 \times 10^{19}$ to $7 \times 10^{19} \text{ cm}^{-3}$) of a GaAs active region to shorten the lifetime without significantly degrading the internal quantum efficiency. For a test device that was doped at $2 \times 10^{19} \text{ cm}^{-3}$, the authors found a cutoff frequency of 440 MHz and an internal quantum efficiency of 25 – 30 %. The authors found a cutoff frequency of 1.7 GHz and an internal quantum efficiency of 10 % for a device that was doped at $7 \times 10^{19} \text{ cm}^{-3}$.

References

- Chen C. H., Hargis M., Woodall J. M., and Melloch M. R., Reynolds J. S., Yablonovitch E., and Wang W. “GHz bandwidth GaAs light-emitting diodes” *Appl. Phys. Lett.* **74**, 3140 (1999)
- Lee T. P. “Effect of junction capacitance on the rise time of LEDs and on the turn-on delay of injection lasers” *Bell Syst. Tech. J.* **54**, 53 (1975)
- Saul R. H., Lee T. P., and Burrus C. A. “Light-emitting-diode device design” in “Lightwave communications technology” edited by W. T. Tsang *Semiconductor and Semimetals* Vol. **22** Part C (Academic Press, San Diego, 1985)
- Schubert E. F., Hunt N. E. J., Malik R. J., Micovic M., and Miller D. L. “Temperature and modulation characteristics of resonant-cavity light-emitting diodes” *IEEE J. Lightwave Technology* **14**, 1721 (1996)
- Wood D. “Optoelectronic semiconductor devices” p. 98 (Prentice Hall, New York, 1994)
- Zucker J. “Closed-form calculation of the transient behavior of (AlGa)As double heterojunction LEDs” *J. Appl. Phys.* **49**, 2543 (1978)

Appendix 1

Frequently used symbols

<i>Symbol</i>	<i>Explanation</i>	<i>SI unit</i>	<i>Common unit</i>
a_0	lattice constant	m	nm or Å
a_B	Bohr radius	m	nm or Å
A	area	m ²	cm ²
\mathcal{A}	Einstein \mathcal{A} coefficient	s ⁻¹	s ⁻¹
α	absorption coefficient	m ⁻¹	cm ⁻¹
α_g	absorption coefficient at the bandgap energy	m ⁻¹	cm ⁻¹
α_{fc}	free-carrier absorption coefficient	m ⁻¹	cm ⁻¹
α_0	absorption coefficient at $h\nu = 2 E_g$	m ⁻¹	cm ⁻¹
B	bimolecular recombination coefficient	m ³ /s	cm ³ /s
\mathcal{B}	Einstein \mathcal{B} coefficient	m ³ /(J s ²)	m ³ /(J s ²)
c	velocity of light in vacuum	m/s	m/s
C	capacitance	F	F
C-V	capacitance-versus-voltage characteristic	(-)	(-)
d	layer thickness	m	μm
D	diffusion constant of dopants	m ² /s	cm ² /s
\mathcal{D}	dielectric displacement	C/m ²	C/cm ²
D_n, D_p	diffusion constant of electrons, holes	m ² /s	cm ² /s
$\Delta n, \Delta p$	change in electron concentration, hole concentration	m ⁻³	cm ⁻³
$\Delta E_C, \Delta E_V$	band discontinuity of conduction band, valance band	J	eV
e	elementary charge	C	C
\mathcal{E}	electric field	V/m	V/cm
E	energy	J	eV
E_C	energy of conduction band edge	J	eV
E_F	Fermi energy	J	eV
E_{Fi}	Fermi energy in intrinsic semiconductor	J	eV
E_{Fn}	quasi-Fermi energy in n-type region	J	eV
E_{Fp}	quasi-Fermi energy in p-type region	J	eV
E_g	energy of semiconductor bandgap	J	eV
E_{Ryd}	Rydberg energy	J	eV
E_T	energy of trap or deep level	J	eV
E_V	energy of valence band edge	J	eV
E_0	energy of lowest state of a quantum well	J	eV
$\epsilon = \epsilon_r \epsilon_0$	dielectric permittivity	A s/(V m)	A s/(V m)
ϵ_0	absolute dielectric constant	A s/(V m)	A s/(V m)
ϵ_r	relative dielectric constant	(-)	(-)
η	efficiency	%	%
η_{int}	internal quantum efficiency	%	%

η_{ext}	external quantum efficiency	%	%
$\eta_{\text{extraction}}$	extraction efficiency	%	%
η_{power}	power efficiency	%	%
f	frequency	Hz	Hz
F	finesse of cavity	(—)	(—)
Φ	electrostatic potential	V	V
Φ	angle	°	°
Φ_{lum}	luminous flux	lm	lm
ϕ	angle	°	°
ϕ_c	critical angle for total internal reflection	°	°
G	generation rate	m^{-3}/s	cm^{-3}/s
G_0	generation rate in equilibrium	m^{-3}/s	cm^{-3}/s
G_{excess}	excess generation rate	m^{-3}/s	cm^{-3}/s
h	Planck constant	J s	J s or eV s
\hbar	Planck constant divided by 2π	J s	J s or eV s
$h\nu$	photon energy	J	eV
I	optical intensity	W/m^2	W/cm^2
I	current	A	A
I_s	saturation current	A	A
I-V	current-versus-voltage characteristic	(—)	(—)
J	current density	A/m^2	A/cm^2
J_s	saturation current density	A/m^2	A/cm^2
k	wave number ($2\pi/\lambda$)	m^{-1}	cm^{-1}
k	Boltzmann constant	J/K	J/K or eV/K
L	length	m	m
L_{cav}	cavity length	m	μm
L_n, L_p	diffusion length of electrons, holes	m	μm
L-I	light-output-power-versus-current characteristic	(—)	(—)
λ	wavelength	m	nm
λ_{Bragg}	Bragg reflection wavelength	m	nm
m	mass	kg	kg
m_e	free electron mass	kg	kg
m^*	effective mass	kg	kg
m_e^*, m_h^*	effective mass of electron, hole	kg	kg
$m_{\text{lh}}^*, m_{\text{hh}}^*$	effective mass of light hole, heavy hole	kg	kg
$\mu = \mu_r \mu_0$	magnetic permeability	V s/(A m)	V s/(A m)
μ_0	absolute magnetic constant	V s/(A m)	V s/(A m)
μ_r	relative magnetic constant	(—)	(—)
μ_n, μ_p	mobility of electrons, holes	$\text{m}^2/\text{V s}$	$\text{cm}^2/\text{V s}$
n	n-type semiconductor material	(—)	(—)
n^-	lightly doped n-type material	(—)	(—)
n^+	heavily doped n-type material	(—)	(—)
n	electron concentration	m^{-3}	cm^{-3}

n^{2D}	two-dimensional electron concentration	m^{-2}	cm^{-2}
n_0	equilibrium electron concentration	m^{-3}	cm^{-3}
n_{ideal}	diode ideality factor	(–)	(–)
n_p	electron concentration in p-type material	m^{-3}	cm^{-3}
\bar{n}	refractive index	(–)	(–)
\bar{n}_{air}	refractive index of air	(–)	(–)
\bar{n}_s	refractive index of semiconductor	(–)	(–)
\bar{n}_{gr}	group index	(–)	(–)
N_A, N_D	acceptor, donor concentration	m^{-3}	cm^{-3}
N_c	effective density of states at conduction band edge	m^{-3}	cm^{-3}
N_T	concentration of trap or deep level	m^{-3}	cm^{-3}
N_v	effective density of states at valence band edge	m^{-3}	cm^{-3}
ν	frequency of optical radiation	Hz	Hz
p	p-type semiconductor material	(–)	(–)
p^-	lightly doped p-type material	(–)	(–)
p^+	heavily doped p-type material	(–)	(–)
p	momentum	kg m/s	kg m/s
p	hole concentration	m^{-3}	cm^{-3}
p^{2D}	two-dimensional hole concentration	m^{-2}	cm^{-2}
p_0	equilibrium hole concentration	m^{-3}	cm^{-3}
p_n	hole concentration in n-type material	m^{-3}	cm^{-3}
P	power	W	W
$P(\lambda)$	spectral power density	W/m	W/nm
Q	electrical charge	C	C
Q	cavity quality factor	(–)	(–)
r	optical field reflection coefficient	(–)	(–)
R	optical power reflection coefficient	(–)	(–)
R	resistance	Ω	Ω
R	recombination rate	m^{-3}/s	cm^{-3}/s
R_0	recombination rate in equilibrium	m^{-3}/s	cm^{-3}/s
R_{excess}	excess recombination rate	m^{-3}/s	cm^{-3}/s
ρ	charge density	C/m^3	C/cm^3
ρ	resistivity	Ωm	Ωcm
ρ_c	specific contact resistance	Ωm^2	Ωcm^2
ρ_{DOS}^{3D}	three-dimensional density of states	m^{-3}/J	cm^{-2}/eV
ρ_{DOS}^{2D}	two-dimensional density of states	m^{-2}/J	cm^{-2}/eV
S	surface recombination velocity	m/s	cm/s
σ	conductivity	$(\Omega m)^{-1}$	$(\Omega cm)^{-1}$
t	time	s	s
t	layer thickness	m	μm
T	temperature	K	K or $^{\circ}C$
T	optical power transmission coefficient	(–)	(–)

τ	recombination lifetime	s	s
τ_{cav}	recombination lifetime in a cavity	s	s
τ_n, τ_p	recombination lifetimes of electrons, holes	s	s
u, u'	CIE uniform chromaticity coordinate of 1960, 1976	(–)	(–)
v	velocity	m/s	m/s
v_{gr}	group velocity	m/s	m/s
v_{ph}	phase velocity	m/s	m/s
v, v'	CIE uniform chromaticity coordinate of 1960, 1976	(–)	(–)
V	voltage	V	V
V_D	diffusion voltage	V	V
V_j	junction voltage	V	V
V_{th}	threshold voltage	V	V
$V(\lambda)$	eye sensitivity function	(–)	(–)
W	width	m	m
W_D	depletion layer width	m	μm
W_{DH}	double heterostructure width	m	μm
x	chromaticity coordinate	(–)	(–)
\bar{x}	CIE color matching function (red)	(–)	(–)
X	CIE tristimulus value (red)	(–)	(–)
y	chromaticity coordinate	(–)	(–)
\bar{y}	CIE color matching function (green)	(–)	(–)
Y	CIE tristimulus value (green)	(–)	(–)
\bar{z}	CIE color matching function (blue)	(–)	(–)
Z	CIE tristimulus value (blue)	(–)	(–)

Note: This list does not contain some symbols that are used only in the section where they are defined.

Appendix 2

Physical constants

a_B	=	0.5292 Å	Bohr radius	($a_B = 0.5292 \times 10^{-10}$ m)
ϵ_0	=	8.8542×10^{-12} A s/(V m)	absolute dielectric constant	
e	=	1.6022×10^{-19} C	elementary charge	
c	=	2.9979×10^8 m/s	velocity of light in vacuum	
E_{Ryd}	=	13.606 eV	Rydberg energy	
g	=	9.8067 m/s ²	acceleration on earth at sea level due to gravity	
G	=	6.6873×10^{-11} m ³ /(kg s ²)	gravitational constant	($F = G M m/r^2$)
h	=	6.6261×10^{-34} J s	Planck constant	($h = 4.1356 \times 10^{-15}$ eV s)
\hbar	=	1.0546×10^{-34} J s	$\hbar = h/(2\pi)$	($\hbar = 6.5821 \times 10^{-16}$ eV s)
k	=	1.3807×10^{-23} J/K	Boltzmann constant	($k = 8.6175 \times 10^{-5}$ eV/K)
μ_0	=	1.2566×10^{-6} V s/(A m)	absolute magnetic constant	
m_e	=	9.1094×10^{-31} kg	free electron mass	
N_{Avo}	=	6.0221×10^{23} mol ⁻¹	Avogadro number	
$R = k N_{\text{Avo}}$	=	8.3145 J K ⁻¹ mol ⁻¹	ideal gas constant	

Note: The *dielectric permittivity* of a material is given by $\epsilon = \epsilon_r \epsilon_0$, where ϵ_r and ϵ_0 are the *relative* and *absolute* dielectric constant, respectively. The *magnetic permeability* of a material is given by $\mu = \mu_r \mu_0$ where μ_r and μ_0 are the *relative* and *absolute* magnetic constant, respectively.

Useful conversions

$$1 \text{ eV} = 1.6022 \times 10^{-19} \text{ C V} = 1.6022 \times 10^{-19} \text{ J}$$

$$kT = 25.86 \text{ meV} \quad (\text{at } T = 300 \text{ K})$$

Appendix 4

Room-temperature properties of semiconductors

Quantity	Symbol	Ge	Si	InAs	InP	GaAs	GaP	GaN	(Unit)
Crystal structure		D	D	Z	Z	Z	Z	W	—
Gap: direct (<i>D</i>) / indirect (<i>I</i>)									—
Lattice constant	a_0	= 5.64613	5.43095	6.0584	5.8686	5.6533	5.4512	$a_0 = 3.189$ $c_0 = 5.185$	Å
Bandgap energy	E_g	= 0.66	1.12	0.354	1.35	1.42	2.26	3.4	eV
Intrinsic carrier concentration	n_i	= 2×10^{13}	1×10^{10}	7.8×10^{14}	1×10^7	2×10^6	1.6×10^0	1.9×10^{-10}	cm^{-3}
Effective DOS at CB edge	N_c	= 1.0×10^{19}	2.8×10^{19}	8.3×10^{16}	5.2×10^{17}	4.4×10^{17}	1.9×10^{19}	2.3×10^{18}	cm^{-3}
Effective DOS at VB edge	N_v	= 6.0×10^{18}	1.0×10^{19}	6.4×10^{18}	1.1×10^{19}	7.7×10^{18}	1.2×10^{19}	1.8×10^{19}	cm^{-3}
Electron mobility	μ_n	= 3900	1500	33000	4600	8500	110	1500	cm^2/Vs
Hole mobility	μ_p	= 1900	450	450	150	400	75	30	cm^2/Vs
Electron diffusion constant	D_n	= 101	39	858	120	220	2.9	39	cm^2/s
Hole diffusion constant	D_p	= 49	12	12	3.9	10	2	0.75	cm^2/s
Electron affinity	χ	= 4.0	4.05	4.9	4.5	4.07	3.8	4.1	V
Minority carrier lifetime	τ	= 10^{-6}	10^{-6}	10^{-8}	10^{-8}	10^{-8}	10^{-6}	10^{-8}	s
Electron effective mass	m_e^*	= $1.64 m_e$	$0.98 m_e$	0.022	$0.08 m_e$	$0.067 m_e$	$0.82 m_e$	$0.20 m_e$	—
Heavy hole effective mass	m_{hh}^*	= $0.28 m_e$	$0.49 m_e$	0.40	$0.56 m_e$	$0.45 m_e$	$0.60 m_e$	$0.80 m_e$	—
Relative dielectric constant	ϵ_r	= 16.0	11.9	15.1	12.4	13.1	11.1	8.9	—
Refractive index	n	= 4.0	3.3	3.5	3.4	3.4	3.0	2.5	—
Absorption coefficient near E_g	α	= 10^3	10^3	10^4	10^4	10^4	10^3	10^5	cm^{-1}

• D = diamond. Z = zincblende. W = wurtzite. DOS = density of states. VB = valence band. CB = conduction band

• The Einstein relation relates the diffusion constant and mobility in a non-degenerately doped semiconductor: $D = \mu (k T / e)$.

• Minority carrier diffusion lengths are given by $L_n = (D_n \tau_n)^{1/2}$ and $L_p = (D_p \tau_p)^{1/2}$.

• Mobilities and diffusion constants apply to low doping concentrations ($\approx 10^{15} \text{ cm}^{-3}$).

• Minority carrier lifetime τ applies to doping concentrations of 10^{18} cm^{-3} . For other doping concentrations, the lifetime τ is given by $\tau = B^{-1} (n + p)^{-1}$, where $B_{\text{GaAs}} = 10^{-10} \text{ cm}^3/\text{s}$ and $B_{\text{Si}} = 10^{-12} \text{ cm}^3/\text{s}$.

Index

- absorption, 48–51, 114–118, 201, 202, 204, 285
- absorption coefficient, *see* absorption
- absorption constant, *see* absorption coefficient
- acceptance angle, 269, 270
- active region, 99–105, 294, 295
- active region thickness, 101
- adiabatic transport, 68
- AlGaAs, 3–7, 123–128, 146, 160–163, 167–169, 212, 282
- AlGaInP, 18, 19, 123–128, 136–138, 146, 163–165, 167–169, 171–174, 209–211, 253, 281–283
- Al₂O₃ and Al_xO_y, *see* alumina
- alumina (Al₂O₃), 139, 215
- anneal, post-growth, 16
- anti-reflection coating, 138, 139, 284, 285
- attenuation, *see* fiber loss
- Auger coefficient, 40
- Auger recombination, 35, 39, 40
- avalanche multiplication, 2

- bandgap energy, 58, 80, 157, 160, 161, 164, 167
- bandwidth, 288–299
- bimolecular recombination, *see* recombination coefficient, 51–53
- binning, 252
- black-body radiation, 49, 50, 232–234
- blocking layer, *see* electron-blocking layer, *see also* current-blocking layer
- Boltzmann statistics, 57
- Bragg wavelength, 184, 185
- Burrus-type LED, 278, 279
- brightness, 229
- buffer layer, 9

- CaF₂, 139
- candela, definition, 220
- candle, 220
- candlepower, definition, 220
- capacitance, 288–299
- carborundum, 1
- carrier leakage, 68–71, 97, 107
- carrier lifetime, *see* lifetime
- carrier loss, *see* carrier leakage
- carrier overflow, 71–74
- carrier sweep-out, 292–295
- cathodo-luminescence, 109, 110
- cavity quality factor, 182
- characteristic temperature, *see* temperature
- chromaticity coordinates, 227–236, 251, 258
 - uniform, 229, 234
- chromaticity diagram, 226–236
 - and LEDs, 231
- CIE, 226
- CIE illuminant, 233
- CIE reference object, 238, 243, 244
- circular contact, *see* contact
- cladding layer, *see also* confinement layer
 - of fiber, 261–263, 269
 - of planar waveguide, 284
- cluster, 171
- coherence, 286
- coherent strain, *see* strain
- color and wavelength of LED, 223, 228, 230, 231, 236
 - color difference, 229, 238
 - color, dominant, 230
 - color gamut, 235, 236
 - color matching function, 226, 227, 242
 - color mixing, 234–236
 - color, primary, 236
 - color purity, 229–231
 - color rendering, 236–240
 - index, 236–240
 - special, 237
 - general, 237
 - color temperature, 233, 234
 - correlated, 233, 234
- colorimetry, 219
- communication, *see* optical communication
- communication LED, 277–287
- compensation, 4
- complementary wavelength, 245, 246
- composition grading, 64–68
- concentration, critical, 105
- cone cells, 219, 220, 226
- confinement layer, 99–101, 105–107
- confinement region, *see* confinement layer
- contact, circular 130, 131, 135
- contact geometry, 135
- contact reflectivity, 148, 150
- contact, stripe, 43, 44, 128–134
- core of fiber, 261–263, 269
- core region, 284
- correlated color temperature, *see* color temperature
- coupling, 95, 96, 210, 211, 269–273, 278, 280
- critical concentration, *see* concentration
- critical thickness, *see* thickness
- cross-hatch pattern, 109, 110

- current-blocking layer, 147, 148, 281, 282
- current crowding, 131–135, 291
- current shaping, 295–297
- current spreading, 123–131
 - length, 129–131, 133–135
- current–voltage characteristic, 55–83, 173
- dangling bonds, 40, 41, 109, 110
- data rate, *see* maximum data rate
- decay of luminescence, 33, 34, 293–295
- deep levels, 35, 108
- defect, 35, 38, 39, 45
- depletion region, 55, 56
- dichromatic source, 234, 247
- die shaping, 119–123
- diffusion, 281, 282
 - capacitance, 291
 - length, 62, 64
 - voltage, 56
- diode equation, *see* Shockley equation
- diode ideality factor, *see* ideality factor
- diode voltage, *see* forward voltage
- direct–indirect transition, 7, 18, 156–159, 161
- distributed Bragg reflector 140–147, 183, 184, 203, 204
- dominant color, *see* color
- donor–acceptor pair recombination, *see* recombination
- doping, GaN p-type, 16
- doping of active region, 102–105
- doping of confinement layer, 105–108
- double heterostructure, 63, 64, 99–101, 118, 161, 162
- drive circuit, 81–83
- dye, 249, 259
- edge-emitting LED, 263, 283–287
- efficiency, *see* extraction efficiency, luminous efficiency, quantum efficiency, energy, 247, 248
 - power, 85, 225
- Einstein A coefficient, 53, 54
- Einstein B coefficient, 53, 54
- Einstein model, 48, 53, 54
- Einstein relation, 62
- electroluminescence, 1
- electron-blocking layer, 75–77
- electron-beam irradiation, 16
- emission
 - by avalanche multiplication, 2
 - by minority carrier injection, 2
 - energy, 62, 86
 - linewidth, 88, 89
 - pattern, 92–96, 225
 - spectrum, 85–89, 171, 251, 258, 280, 285, 286
 - temperature dependence, 97, 98, 172
- encapsulant, 96, 97, 139, 140
- energy gap, *see* bandgap energy
- epoxy resin encapsulant, *see* encapsulant
- epoxy resin dome, *see* encapsulant
- equal-energy point, 230
- equal-excitation point, 230
- escape cone, *see* light escape cone
- exponential decay, 33, 34
- extended defect, *see* defect
- external quantum efficiency, *see* quantum efficiency
- extraction, *see* light extraction
- extraction efficiency, 84, 96, 97, 114, 226
- eye diagram 296, 297
- eye, human, 219
- eye sensitivity function, 222, 223, 226, 242
- Fabry–Perot resonators, 180–183, 187, 286
- fall time, 288–297
- far-field pattern, *see* emission pattern
- feedback, *see* optical feedback
- fiber
 - coupling, *see* coupling
 - graded-index multimode 261–263
 - loss, 263, 264
 - plastic, 263, 264, 278
 - silica, 263, 264, 278
 - single-mode 261–263
 - step-index multimode 261–263
- fineness of cavity, 182, 201
- flip-chip packaging, 149, 150
- forward voltage, 58, 59, 77–83
 - temperature dependence, 79–83, 174
- free-carrier absorption, *see* absorption
- free-space communication, *see* communication
- free spectral range, 181, 182
- Fresnel reflection, 95, 138, 184
- GaAs, 3–7, 118, 160–165, 167–169, 212, 282
- GaAsP, 7–14, 155–160, 167–169
- GaAsP:N, 155–160, 167–169
- GaInAs, 203–208, 212
- GaInN 16–18, 131–135, 166–169, 171–174, 250–253
- GaInPAs, 279–281
- gamut, *see* color gamut
- GaN, 14–18, 118

GaP, 5, 6, 9–14, 118, 123–128, 136–138, 155–160, 167–169
 GaP:N, 9–14, 155–160, 167–169
 gain, *see* optical gain
 general illumination, 245
 graded-index fiber, *see* fiber
 grading, *see* composition grading
 group refractive index, *see* refractive index

 Heisenberg uncertainty principle, *see* uncertainty principle
 hemispherical LED, 94
 heterojunction, 63–77
 history of LEDs, 1–19
 homojunction, 62, 63
 hue, 229
 human eye, *see* eye
 human vision, 219–240

 ideal isotropic emitter, *see* isotropic emitter
 ideality factor, 59, 60
 illuminance, 221
 illuminant, *see* CIE illuminant
 implantation, 209, 282
 impurity
 isoelectronic, 9–14, 155–160
 optically active, 9–14
 rare-earth, 248
 incandescence, 3
 infrared communication LEDs, 277–287
 injection laser, *see* laser
 InP, 118
 internal quantum efficiency, *see* quantum efficiency
 isoelectronic impurity, *see* impurity
 isotropic emitter, 274
 ideal 206, 207

 joint dispersion relation, 87
 joint density of states, 87, 88
 junction displacement, 103–105

 lambertian emission pattern, *see* emission pattern
 laser
 blue, 17
 injection, 4, 17, 178, 213, 214, 215, 264, 284, 286
 microdisk, 215
 vertical-cavity surface-emitting, 202, 203
 lattice mismatch, 109–112
 leakage, *see* carrier leakage

lens 271–273, 281
 lifetime, *see* reliability
 carrier 29–34, 45, 46, 187, 291, 292, 297–299
 light escape cone, 89–92, 119–121
 light extraction, 92–97
 light, trapped, 84, 119, 120
 light-versus-current characteristic, 172, 286
 linewidth, *see* emission linewidth
 loss, *see* carrier loss, *see also* fiber loss, *see also* absorption
 lumen, definition, 220
 luminance, 225
 luminescence decay, 33, 34
 luminescence killer, 35
 luminous efficacy, 222–226
 luminous efficiency, 167–169, 170, 222–226, 255–258
 luminous flux, 170, 220–226, 256
 luminous intensity, 220–226
 lux, definition, 221

 material dispersion, 266–269
 Matthews–Blakeslee law, *see* critical thickness
 maximum data rate, 266, 267, 275, 290–299
 mesa, 108, 109
 metal mirror, 184, 186
 metal–semiconductor diode, *see* Schottky diode
 metal–semiconductor emitters, *see* Schottky diode
 metal–insulator–semiconductor emitter, 15, 16
 microcavity, *see* resonant cavity
 minority carrier lifetime, *see* lifetime
 MIS, *see* metal–insulator–semiconductor
 misfit dislocation, 109–112
 modal dispersion, 265, 266, 268, 269
 modes, *see* optical modes
 momentum of carrier and photon, 87
 monochromatic nature of LEDs, 89
 monomolecular recombination, *see* recombination
 multimode fiber, *see* fiber
 multipath distortion, 274, 275
 multipath delay, *see* multipath distortion

 noise, 275, 276
 non-adiabatic injection, 77
 non-radiative recombination, *see* recombination
 numerical aperture, 269, 270

 optical communication, 199, 208, 213, 261–276, 277

- free space, 273–276, 277, window, 263, 264
- optical feedback, 284, 285
- optical gain, 203
- optical mode density, 187–191
- optical modes, 180–183, 286
- optical power, 220
- optically active impurity, *see* isoelectronic impurity, *see also* rare-earth impurity
- parabolic LED, 94
- parallelepiped, 119–121
- parallel resistance, *see* resistance
- penetration depth, 186, 187
- phase refractive index, *see* refractive index
- photometric units, 220, 221
- phonon emission, 77
- phosphor, 247, 248, 250–253
- phosphorescence, 250, 251
- photon recycling, 211–213, 253–259
- photon recycling semiconductor LED, 253–259
- photonic bandgap structures, 178, 216
- photopic vision, 220
- planar LED, 94
- plastic optical fiber, *see* fiber
- Planck black-body radiation, *see* black-body radiation
- planckian locus, 233, 234
- p-n junction, 55–68
- p-n junction displacement, *see* junction displacement
- polymer encapsulant, *see* encapsulant
- polymer light-emitting diode, 97, 199, 215
- point defect, *see* defect
- post-growth anneal, *see* anneal
- power efficiency, *see* efficiency
- power ratio of dichromatic source 246, 254, 255
- primary color, *see* color
- quantum dot, 171
- quantum efficiency
 - external, 85, 114, 226
 - internal, 46, 84, 114, 226
- quantum well structures, 32, 33, 161, 162, 203
- radiant flux, *see* optical power
- radiation pattern, *see* emission pattern
- radiative recombination, *see*, recombination
- radiometric units, 220
- rare-earth impurity, *see* impurity
- RCLED, *see* resonant-cavity light-emitting diode
- RC time constant, 288–299
- RC time circuit, 288–299
- recombination
 - bimolecular, 27, 32, 33, 51–53, 64
 - donor–acceptor pair, 11
 - lifetime, *see* lifetime
 - monomolecular, 29, 292
 - non-radiative, 26, 34–46, 97, 108
 - radiative, 26–34, 44–46, 48–54
 - spontaneous, 48–54
 - stimulated, 53, 54
 - surface, 40–44, 97, 108
- reduced mass, 87
- reference light source, 237
- reference object, *see* CIE reference object
- reflectors, 183–187, 200–203
- refractive index
 - effective, 185
 - group, 267, 268
 - phase, 266–268
- regrowth, 147, 148, 282
- reliability, 108, 109, 163
- resistance, 59–61, 64–68, 82, 204
- resonant-cavity light-emitting diode, 6, 7, 179, 198–216, 282
- resonant-cavity structure 178–194, 199–203
- resonant cavity, tunable, 215
- rise time, 288–297
- rod cells, 219, 220
- roughened surface, *see* surface texturing
- saturation current, 56–57
- saturation of color, *see* color purity
- secondary ion mass spectrometry, 104
- self-absorption, *see* absorption
- series resistance, *see* resistance
- Schottky diode, 2, 14–16
- scotopic vision, 220
- shaping of die, *see* die shaping
- Shockley equation, 56–59
- Shockley–Read recombination, 35–38
- shunt, *see* resistance
- SiC, 1–3
- Si₃N₄, 139
- single-mode fiber, *see* fiber
- SiO₂, *see* silica
- silica (SiO₂), 139
- silica fiber, *see* fiber
- silver mirror, *see* metal mirror
- single-mode LED, 179
- Snell's law, 92, 114, 119, 269
- Solar spectrum, 232

solubility of nitrogen, 159, 160
 spectral purity, 198, 206
 spontaneous emission enhancement, 178–194, 203
 spontaneous emission lifetime, *see* carrier lifetime
 spontaneous emission rate, 187–194
 spontaneous recombination, *see* recombination
 step-index fiber, *see* fiber
 stimulated emission, 53, 54, 284–287
 stop band, 142, 184, 185
 strain, coherent, 110–112
 stretched exponential decay, 33, 34
 stripe contact, *see* contact
 superlinear light-vs.-current characteristic, 286, 287
 superluminescent LED, 178, 263, 283–287
 surface-emitting communication LED, 204, 209, 273, 287–283, 286
 surface recombination, *see* recombination velocity, 42–44
 surface reconstruction, 41
 surface roughening, *see* surface texturing
 surface states, 40–44
 surface texturing, 112, 123
 surfactant, 103
 synchronous optical network, 297

temperature, characteristic, 97, 98
 temperature coefficient of bandgap energy, 80, 81
 textured surface, *see* surface texturing
 thickness, critical, 110–113
 3 dB frequency, 288–290, 297–299
 threshold voltage, 57, 58
 thresholdless laser, *see* zero-threshold laser
 TiO₂, *see* titania
 titania, 139
 total internal reflection, 184, 261
 transfer function, 290
 transparent substrate technology, 136–138, 162, 163
 trapped light, *see* light
 tristimulus value, 227–229
 tunable resonant cavity, *see* resonant cavity
 tunnel junction, 128

ultraviolet, 58, 59, 252
 uncertainty principle, 9, 10
 uniform chromaticity coordinates, *see* chromaticity coordinates
 Urbach energy, *see* Urbach tail
 Urbach tail, 114–116

van Roosbroeck–Shockley equation, 50
 van Roosbroeck–Shockley model, 48–53
 vertical-cavity surface-emitting laser, *see* laser
 vertical transition, 86, 87
 visible-spectrum LEDs, 155–174
 vision, *see* human vision

wall-plug efficiency, *see* efficiency
 waveguided modes, 202, 284
 waveguide, *see* fiber
 wavelength converter, 246–259
 white illuminant, 232–234
 white LED, 18, 245–259
 based on phosphor, 250–253
 Wiens’s law, 232, 233
 window layer, *see* current spreading

YAG phosphor, 248, 250
 yellow luminescence, 39

zero-threshold laser, 179, 180, 203, 213–215
 Zn–O, 10–12
 ZnS, 3, 139
 ZrO₂, *see* zirconia
 zirconia (ZrO₂), 203