Optoelectronic Components

Wolfgang Freude*
Institute of Photonics and Quantum Electronics (IPQ)
(Institut für Photonik und Quantenelektronik)
http://www.ipq.kit.edu
Karlsruhe Institute of Technology (KIT)
http://www.kit.edu

International Department
http://hector.idschools.kit.edu

Karlsruhe School of Optics & Photonics (KSOP)
http://ksop.idschools.kit.edu

SS 2015

April 13, 2015

mailto:W.Freude@kit.edu

Contents

Preface

Lightwave technology developed over the last 40 years has greatly influenced our needs for communication. Resources made accessible in the World Wide Web (WWW) have changed our attitude towards information acquisition, which is being regarded as an everyday's necessity, and even as a natural right for everybody. Today's undersea and underground optical cables provide large-capacity links carrying more than 90 % of the communication traffic.

This course concentrates — after a brief introduction to optical communication systems — on basic optical communication components. A more specialized view on advanced transmitters and receivers for modern optical transmission systems is presented in the course "Optical Transmitters and Receivers (OTR)" in the Winter term.

Emphasis is on physical understanding, exploiting results from electromagnetic field theory (light waveguides), solid-state physics (laser diodes, LED, and photodiodes), and communication theory (receivers, noise). A selection of topics will be presented, on which the questioning during the oral examination will be based. — Some minimal background is required: Calculus, differential equations, linear systems, Fourier transform, and pn-junction physics. For further reading, the following list provides some material. References on more specialized topics were cited in the text.

Texbooks: Grau, G.; Freude, W.: **Optische Nachrichtentechnik**, 3. Ed. Berlin: Springer-Verlag 1991. In German. Since 1997 out of print. Corrected reprint from University Karlsruhe 2005, available in electronic form via W. F. (`w.freude@kit.edu`). **The present course is based on this book and on material found in:** Agrawal, G. P.: **Fiber-optic communication systems**. Chichester: John Wiley & Sons 1997 — Agrawal, G. P.: **Lightwave technology. Vol. 1: Components and devices.** Vol. 2: Telecommunication systems. Hoboken: John Wiley & Sons 2004 — Ebeling, K. J.: Integrierte Optoelektronik, 2. Ed. Berlin: Springer-Verlag 1992. In German, English translation available. — Freude, W.: Vielmodenfasern (Multimode fibres). In: Voges, E.; Petermann, K. (Eds.), Optische Kommunikationstechnik. Handbuch für Wissenschaft und Industrie (Handbook of optical communications). Springer-Verlag, Berlin 2002, pp. 214–260. In German — Ghatak, A.; Thyagarajan, K.: **Introduction to fiber optics**. Cambridge: University Press 1998 — Hecht, E.: Optics, 2. Ed. Reading: Addison-Wesley 1974 — Hecht, J.: **Understanding fiber optics**, 4. Ed. Upper Saddle River: Prentice Hall 2002 — Iizuka, K.: **Elements of photonics**, Vol. I and II. New York: John Wiley & Sons 2002 — Jahns, J.: Photonik. Grundlagen, Komponenten und Systeme. München: Oldenburg 2001. In German — Liu, M. M.-K.: **Principles and applications of optical communications**. Chicago: McGraw-Hill 1996 — Singh, J.: **Physics of semiconductors and their heterostructures**. New York: McGraw-Hill 1993. — Snyder, A. W.; Love, J. D.: Optical waveguide theory. London: Chapman and Hall 1983 — Sze, S. M.: Physics of semiconductor devices. New York: John Wiley & Sons 1985 — Unger, H.-G.: Planar optical waveguides and fibres. Oxford: Clarendon Press 1977 — Unger, H.-G.: Optische Nachrichtentechnik, Part I and II, 2. Ed. Heidelberg: Dr. Alfred Hüthig 1990 and 1992. In German.

There are other courses on Optical Communications, which cover the material in more detail: "Optical Waveguides and Fibres (OWF / IPQ)", "Field Propagation and Coherence (FPC / IPQ)", and "Optical Transmitters and Receivers (OTR / IPQ)" as already mentioned. A course on "Nonlinear Optics (NLO / IPQ)" discusses nonlinear phenomena, which become increasingly important.

Many figures of this compuscript were taken from "Optische Nachrichtentechnik" (see above) carrying German lettering and decimal commas. If needed, appropriate translations are given in the figure captions.

Directions

At first sight, these lecture notes may be scaring: When you browse through the material from the beginning of Chapter 1 on Page 1 to the end of Chapter 7 on Page 132 you may perceive an overwhelming mass of formulas, graphs and text. How can all these details on optical communication components be taught and understood during a single compact course? I think it is possible, if a few rules are observed:

- Scan the text for physical explanations which could help you developing a pictorial view of the problem.
- Do try to understand the contents and the associated assumptions of important formulae in the given physical context. Best is to put the meaning into words.
- Study the graphical display of major findings. Begin with the axis labels, read the caption, and look carefully at the graph and its parameters.
- Do not start with deriving formulas, and do not learn them by heart (a few exceptions will be named during the lecture). Deriving a relation may be delayed until you have some basic understanding and become curious to learn more about the assumptions and implications. As long as you are fighting with the physical interpretation of a relation, the mathematical details should be of no concern for you, and you will not be questioned on them — it could be different if xou aim at an outstanding examination mark.
- Do not practice for the examination during the examination itself. The only proof of understanding physics is the ability to explain the problem to somebody else *before* you meet me in an examination. Best is you work in groups. Do your job at its full potential at home by exercising from the tutorial, from the summaries, problems and quizzes as listed below. Exercising is very important!

To help you understanding the physical background and the technical implications, you find at the end of the script for each major section the following material:[1]

1. A summary of what you have learned, beginning on Page 135
2. Some questions to think about, which take the form of problems or a quiz.
3. Solutions to the problems and keys for quizzes, beginning on Page 167 and ending on Page 173

During the lectures and especially during the tutorial, which to attend actively I do recommend strongly, there will be time to answer your questions which may arise from studying problems and quizzes (more than one correct answer is possible). If you are then able to solve the problems and to answer the quizzes, *and* if fellow students' faces do not wear a quizzical expression when you try explaining things, you are on firm grounds and are well prepared for the examination.

The presentation and the examination will concentrate on the aforementioned points, while the lecture notes (which will be made available to you during the examination) provide a more complete background for your reference. You can also download the lecture slides which serve as a reminder of the actually presented material. Many slides are hidden, so sometimes the slide numbers increment unevenly. Remember also that there are right-pointing arrow-shaped links (mostly in the upper-right corner), which when clicking on them carry you forward. The skipped pages were also omitted during the lectures, but are kept to satisfy your curiosity (in case you have time for such a thing). *Once more: The omitted material is not relevant for the examination.*

Studying the lecture notes on paper is fine, but because all cross-references are linked in the electronic *p*ortable *d*ocument *f*ormat, it may be helpful to read the pdf-version in parallel. A click on a link carries you immediately to the target,[2] and you can navigate at will. Acrobat Reader[3] or Foxit Reader[4] allow you to search the document for text. You can mark and comment certain lines with an electronic text marker, you can store your comments, and you can retrieve this information later on.

[1] Hecht, J.: Understanding fiber optics, 4. Ed. Upper Saddle River: Prentice Hall 2002. Some problems and quizzes were adapted from this book.

[2] On rare occasions, the target page for "floating" objects like figures or tables is wrong by one (e. g., you arrive on Page 14 instead on Page 15), however, the page number printed in the originating text is always correct.

[3] `http://www.adobe.de`

[4] `http://www.foxitsoftware.com/pdf/rd_intro.php` — A very lean application, no installation is required. Better refrain from establishing Foxit as the default reader if you consider using Acrobat in parallel.

Chapter 1

Introduction

The nature of light

According to Maxwell,[1] light propagates as a wave having a wavelength λ. In vacuum, the speed of light is $c = 2.997\,924\,58 \times 10^8\,\mathrm{m/s}$. However, Planck[2] found that the energy of light radiated from a hot black body is emitted in quanta, the energy of which is in proportion to the observed frequency $f = c/\lambda$, so that each quantum or "photon" has an energy $W = hf$; Planck's constant is $h = 6.626\,075\,5 \times 10^{-34}\,\mathrm{Ws}$. Further, it was shown by de Broglie[3] that each particle having momentum p may be associated with a wavelength $\lambda = h/p$. Obviously, the nature of light is ambiguous. Einstein[4] formulated:[5] "Light is like the French philosopher Voltaire.[6] Voltaire was born catholic, converted as a young man to Protestantism, and returned to Catholicism shortly before his death. Light is born as a particle, lives as a wave, and dies as a photon when being absorbed."

Communication with light

An optical communication system uses lightwaves in a vacuum wavelength range $0.6\,\mu\mathrm{m} \ldots 1.2\,\mu\mathrm{m} \leq \lambda \leq 1.6\,\mu\mathrm{m}$ corresponding to carrier frequencies $f = c/\lambda$ of $500\,\mathrm{THz} \ldots 250\,\mathrm{THz} \geq f \geq 190\,\mathrm{THz}$. A communication *system* is referred to as a point-to-point transmission link. When many transmission links are interconnected with multiplexing or switching functions, they are called a communication *network*. The principle of an optical transmission link is shown in Fig. 1.1. A semiconductor device (*l*aser *d*iode LD, *l*ight-*e*mitting *d*iode LED) emitting light near a wavelength λ is excited by an electric current, thereby converting the electrical signal information to light. The signal is transmitted as a digital modulation of the light power $P(t)$ (unit W) or intensity $I(t) = P(t)/F$ (unit $\mathrm{W/m^2}$), the classical power resulting from an average over a few optical cycles. Sometimes, also analogue modulation is used. The light is transported through a dielectric *l*ight *w*ave*g*uide (LWG), consisting of a low-refractive index cladding and a high-index core, which confines and guides the light in a cross-sectional area F. For long-distance communication, optical quartz glass fibres are

[1] James Clerk Maxwell, mathematician and physicist, ⋆ Edinburgh 13.6.1831, † Cambridge 5.11.1879. Professor in Cambridge, UK

[2] Max Planck, physicist, ⋆ Kiel 23.4.1858, † Göttingen 4.10.1947. Professor in Kiel and Berlin. Nobel prize in physics 1918

[3] Louis Victor, 7. Duke of Broglie (since 1960), named Louis de Broglie, physicist, ⋆ Dieppe 15.8.1892, † Louveciennes (Département Yvelines) 19.3.1987. Nobel prize in physics 1929 (together with O. W. Richardson)

[4] Albert Einstein, physicist, ⋆ Ulm 14.3.1879, † Princeton (NJ) 18.4.1955. "Technical expert 3rd class" at the patent office in Bern (1902–09). Professor at the University of Zurich and Prague (1911/12) and at the Swiss Federal Institute of Technology (ETH) in Zurich. Emigration to the USA in 1933. Professor at the Institute for Advanced Study in Princeton (NJ). American citizen since 1940. Formulated in 1905 (1914–16) the special (general) theory of relativity. Nobel prize in physics 1921

[5] Jahns, J.: Photonik. Grundlagen, Komponenten und Systeme. München: Oldenbourg-Verlag 2001. Page 9

[6] Pseudonym or pen-name of François Marie Arouet, philosopher and writer, ⋆ Paris 21.11.1694, † Paris 30.5.1778

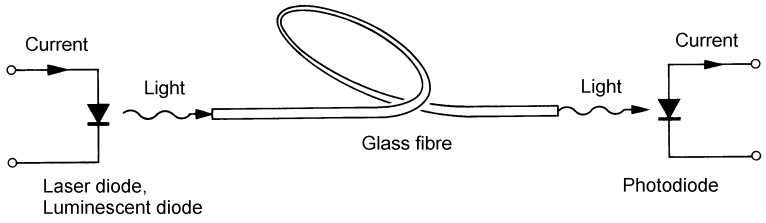
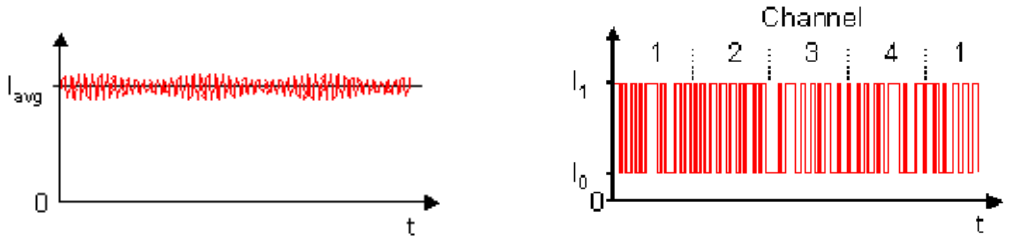


Fig. 1.1. Optical point-to-point transmission link with an intensity-modulated carrier centered at a wavelength λ and direct (incoherent) detection

used. Glass-based fibre waveguides are very thin, immune to electromagnetic interference, have low loss and guide the light over thousands of kilometers. At the receiver end, a *p*hoto*d*etector (PD) with a sensitivity S (unit A / W) reconverts the light to an electric photocurrent $i(t) = SFI(t)$ in proportion to the light intensity. The very straightforward type of reception in Fig. 1.1 is called "direct" or incoherent, as opposed to coherent receivers using a heterodyne or homodyne technique with a *l*ocal laser *o*scillator (LO).

Modulation For encoding the signal information, the transmitted light must be altered or modulated in some way. The physical quantity to be modulated could be the frequency (like in a *f*requency *m*odulation or FM radio receiver), the phase, the electric field amplitude (as in an *a*mplitude *m*odulation or AM radio receiver), or, as it is common in optical communications, the optical intensity I (*i*ntensity *m*odulation IM) as depicted in Fig. 1.1. There are two major methods of modulation — analogue and digital. Analogue modulation, Fig. 1.2(a), uses less bandwidth and is simpler than digital modulation, which however provides a better signal quality at the expense of larger bandwidth requirements and a more complicated circuitry. Present-day communications are controlled and initiated by digital computers, so it is natural to use a digital modulation format for transmission, too.

The most common digital modulation format is pulse code modulation (PCM). Here, the value of an analogue signal $v(t)$ is sampled, and the values are then converted into a binary form. If the signal has a maximum bandwidth of B (unit Hz) then sampling with the so-called Nyquist sampling rate $2B$ at equidistant time increments $1/(2B)$ apart allows an exact reconstruction of the analogue signal, if the sampling values are properly interpolated. This is known as Shannon's[7] sampling theorem.[8]



(a) Analogue intensity modulation

(b) Digital intensity modulation with TDM

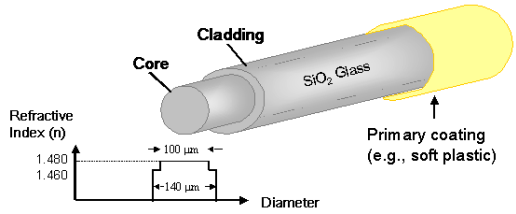
Fig. 1.2. Modulation formats (a) Analogue intensity modulation around an operating point I_{avg} (b) Digital intensity modulation between an off (I_0) and an on value (I_1). For a 4-channel *t*ime *d*ivision *m*ultiplexing scheme (TDM) individual transmission time slots $1 \ldots 4$ are assigned to each data source

[7] Claude Elwood Shannon, engineer and physicist, ⋆ Gaylord (Michigan) 30.4.1916, † Medford (Massachusetts) 24.2.2001. Seminal papers on information theory in 1948. Professor at Massachusetts Institute of Technology since 1956

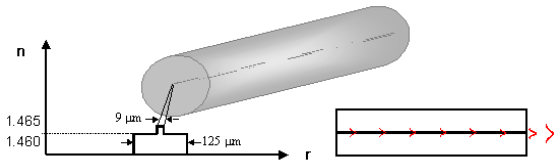
[8] Shannon, C. E.: A mathematical theory of communication. Bell Syst. Tech. J. 27 (1948) 379–423, 623–56

The sampled values (e. g., the numbers $1, 4, 2, 5, 9 \ldots$) are then converted into a form suitable for transmission. For a binary format only two states are physically discriminated, light "off" or I_0 in Fig. 1.2(b) corresponding to a logical "0", and light "on" or I_1 corresponding to a logical "1". The decimal numbers $1, 4, 2, 5, 9$ would first be converted into binary numbers 0001, 0100, 0010, 1001, and then transmitted as time sequences (I_1, I_0, I_0, I_0), (I_0, I_0, I_1, I_0), (I_0, I_1, I_0, I_0), (I_1, I_0, I_0, I_1) of low and high optical intensities by switching the control current of the laser diode in Fig. 1.1 on and off. After transmission, an optical receiver converts the light impulses back into an electrical signal of low (i_0) and high currents (i_1). A digital-to-analogue converter reconstructs the original signal $v(t)$.

Fibres There are two types of fibers, multimode fibres with core diameters of 50 μm, 65 μm, 100 μm, 200 μm, 1 000 μm and 3 000 μm, Fig. 1.3(a), and single-mode fibres with a core diameter of 9 μm or 10 μm, Fig. 1.3(b). For the multimode fibre, coupling light from the transmitter into the fiber core is easier than into the much smaller diameter single-mode fibre. The disadvantage is the stronger light impulse distortion of signals propagating in multimode fibres. A standard cladding diameter for single-mode communication fibres is 125 μm.



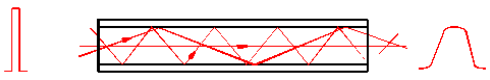
(a) Multimode fibre with step-index profile



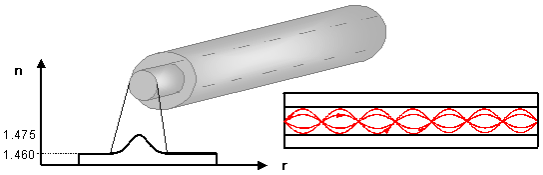
(b) Single-mode fibre with step-index profile

Fig. 1.3. Fibre types with step-shaped refractive index profile comprising a higher-index core and a lower-index cladding (a) Fat-core step-index multimode fibre with a relative refractive index difference $\Delta \approx 1.3\,\%$ (b) Long-haul step-index single-mode communication fibre with $\Delta \approx 0.33\,\%$

Intermodal dispersion Variation in propagation time among different modes creates intermodal dispersion, i. e., group delay differences, which are caused by optical path differences in a step-index multimode fibre. The effect on a light impulse entering the multimode fibre is shown in Fig. 1.4(a). The output impulse is broadened because it is composed of many smaller impulses arriving at different instances of time. If the group velocity in the outer-core regions could be increased, the group delay of these rays could be made the same as for rays propagating mostly in the core centre. This is achieved by gradually reducing the refractive index away from the fibre axis. Such a waveguide is dubbed a graded-index fibre.



(a) Multimode fibre with step-index profile

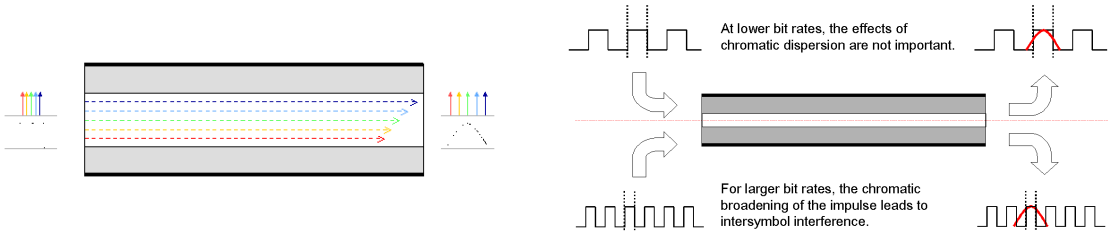


(b) Multimode fibre with graded-index profile

Fig. 1.4. Intermodal dispersion for multimode fibres. (a) Step-index profile with significant group delay differences (b) Graded-index profile, where geometrical path length differences are compensated by radial variations in the refractive index

Chromatic dispersion Dispersion in an optical fibre is not limited to intermodal dispersion. Even a single-mode fibre suffers from different group delays depending on the spectral content of the optical signal. This again leads to output impulse broadening or intramodal dispersion, Fig. 1.5(a). At higher bit rates, the broadened impulses spill into neighbouring time slots (intersymbol

interference), and it becomes increasingly difficult to decide between a logical "0" and a logical "1", Fig. 1.5(b). Therefore, the bit error probability (bit error *ratio* BER; frequently, but wrongly named "bit error *rate*") increases, such limiting the maximum transmission rate.



(a) Chromatic dispersion in a single-mode fibre (b) Group delay dispersion and intersymbol interference

Fig. 1.5. Group delay dispersion and bit error probability (bit error ratio, BER). (a) Different wavelengths ("colours", therefore "chromatic") inside the same mode propagate with different velocities, thereby increasing the output impulse width (b) Broadening of the transmitted impulse leads to bit detection errors

Wavelength division multiplexing The capacity of transmission links can be greatly extended by employing more than one optical carrier in a *w*avelength *d*ivision *m*ultiplexing scheme (WDM),[9,10] Figure 1.6. A number N of laser diodes are modulated in intensity and emit light at wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots, \lambda_N$ near $\lambda = 1.55\,\mu$m. An information channel when realized with a modulated optical carrier is termed an "optical channel" or a "wavelength channel", and referred to as "channel" for brevity. I addition to WDM, channel multiplexing can be done in the time domain, leading to (optical) time-division multiplexing (OTDM).

For a WDM scheme, the optical carriers are separated by 25 GHz, 50 GHz, 100 GHz ($\Delta\lambda = 0.78\,$nm @ $\lambda = 1.55\,\mu$m) and 200 GHz (e. g., 100 channels at $\lambda = 1.55\,\mu$m and 40 Gbit/s are commercially available, see Chapter 4). Thus, the total capacity of one single-mode fibre amounts to 4 Tbit/s if the channels are separated in frequency by, e. g., $\Delta f = 100\,$GHz. An optical *mu*ltiple*x*er (MUX) spatially concentrates these modulated carriers to propagate as wavelength channels in one single fibre. An *o*ptical *a*mplifier (OA) in combination with a *equal*izer (EQUAL) for equalizing the gain in all channels amplifies the signals, which are then transmitted through the single-mode transport fibre. Optical amplifiers overcome the power loss in very long communication links. They

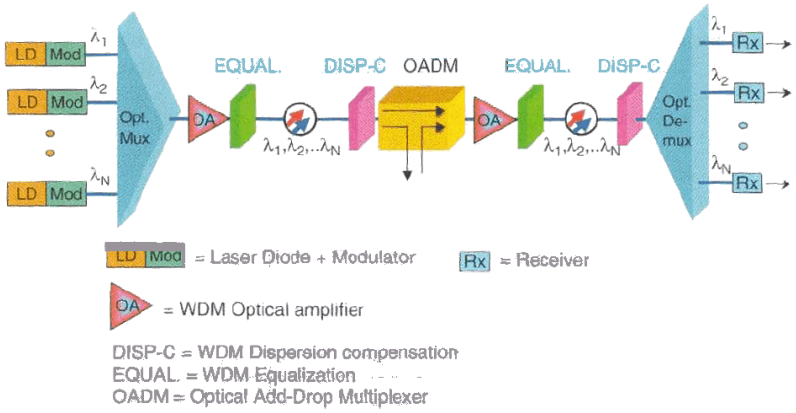


Fig. 1.6. Wavelength division multiplexing transmission scheme. The path from LD MOD(λ_i) to Rx(λ_i) corresponds to the simplified point-to-point transmission depicted in Fig. 1.1. [after Fig. iii on Page xxiv in reference Footnote 10 on Page 4]

[9] Agrawal, G. P.: Fiber-optic communication system. Chichester: John Wiley & Sons 1997. Chapter 7 Page 284
[10] Kartalopoulos, S. V.: DWDM — Networks, devices, and technology. John Wiley & Sons 2003

have bandwidths in the order of $\Delta f = 5 \ldots 10\,\mathrm{THz}$ centred at $\lambda = 1.3\,\mu\mathrm{m}$ and $\lambda = 1.55\,\mu\mathrm{m}$, and remove the speed bottleneck from electronics by optics implementation. There are two primary types of OA, *s*emiconductor *o*ptical laser *a*mplifiers (SOA) and *d*oped *f*ibre *a*mplifiers (DFA). Among all DFA, Er^{3+}-*d*oped *f*ibre *a*mplifiers (EDFA) that amplify light around $\lambda = 1.55\,\mu\mathrm{m}$ are the most mature.

A *disp*ersion *c*ompensator (DISP-C) compensates the wavelength-dependent delay times inside each channel. At the end of the first span, an *o*ptical *a*dd-*d*rop *m*ultiplexer (OADM) adds or drops selectively certain wavelength channels. Then, after a possibly repeated sequence of such spans, an optical *demu*ltiple*x*er (DEMUX) finally separates all optical channels spatially, and optical receivers (Rx) attached to the DEMUX outputs detect the modulated optical signal directly. The path from LD MOD(λ_i) to Rx(λ_i) corresponds to the simplified point-to-point transmission depicted in Fig. 1.1.

Transmission bands Various transmission bands are designated with the following letters: Extended *s*hort-wavelength band "S+", *s*hort-wavelength band "S", *c*onventional or *c*entral band "C", *l*ong-wavelength band "L", extended *l*ong-wavelength band "L+". The attributed wavelengths (unit μm) are seen from Table 1.1. A typical DWDM ITU-T channel grid[11] for the C band is specified in Table 1.2.

Designation of 40-nm bands ($\lambda/\,\mu\mathrm{m}$) at $\lambda = 1.550\,\mu\mathrm{m}$										
S+		**S**		**C**		**L**		**L+**		
1.450	**1.470**	1.490	**1.510**	1.530	**1.550**	1.570	**1.590**	1.610	**1.630**	1.650

Table 1.1. Designation of bands at $\lambda = 1.550\,\mu\mathrm{m}$

Wavelength table for the C band (DWDM ITU-T grid)					
$\lambda_{\mathrm{ITU}}/\,\mathrm{nm}$	$\lambda_{\mathrm{ITU}}/\,\mathrm{nm}$	$\lambda_{\mathrm{ITU}}/\,\mathrm{nm}$	$\lambda_{\mathrm{ITU}}/\,\mathrm{nm}$	$\lambda_{\mathrm{ITU}}/\,\mathrm{nm}$	$\lambda_{\mathrm{ITU}}/\,\mathrm{nm}$
1 527.99	1 534.25	1 540.56	1 546.92	1 553.33	1 559.78
1 528.77	1 535.04	1 541.35	1 547.72	1 554.13	1 560.61
1 529.55	1 535.82	1 542.14	1 548.51	1 554.94	1 561.42
1 530.33	1 536.61	1 542.94	1 549.32	1 555.75	1 562.23
1 531.12	1 537.40	1 543.73	1 550.12	1 556.55	1 563.05
1 531.90	1 538.19	1 544.53	1 550.92	1 557.36	$\Delta = 0.79$
1 532.68	1 538.98	1 545.32	1 551.72	1 558.17	**all:**
1 533.47	1 539.77	1 546.12	1 552.52	1 558.98	$\pm$**0.1**

Table 1.2. DWDM ITU-T grid at $\lambda = 1.550\,\mu\mathrm{m}$. Channel spacing corresponds to frequency grid $\Delta f = 100\,\mathrm{GHz}$

Advantages of optical communications Obviously, optical communication systems can replace conventional electrical systems only, if there is some advantage to be gained, which justifies the additional expenses of a twofold conversion current-light and light-current. Some important advantages of optical signal transport are:

- Large transmission capacity because of the large fibre bandwidth in the order of $(250 - 190)\,\mathrm{THz} = 60\,\mathrm{THz}$
- Low fibre loss, about $2.2, 0.35, 0.15\,\mathrm{dB}\,/\,\mathrm{km}$ at $\lambda = 0.85, 1.3, 1.55\,\mu\mathrm{m}$, i. e., down to 3 dB loss for a fibre length of $L = 20\,\mathrm{km}$ corresponding to a power attenuation by a factor of only 2

[11] Fujitsu: Lightwave Components & Modules Databook. (1998) p. 38. DFB lasers are commercially available with these wavelength gradings.

- Immunity to interference because of the high carrier frequency, and because of the strong confinement of the light inside the fibre

Three milestones of lightwave technology are especially noteworthy. Following an earlier suggestion,[12,13] the first *low-loss fibres* were produced[14] in 1970 reducing the loss from 1 000 dB / km to below 20 dB / km. Further progress[15] resulted by 1979 in a loss of only 0.2 dB / km near $\lambda = 1.55\,\mu$m. The ultimate low loss[16] of 0.154 dB / km for fibres with a silica (SiO_2) core and a F-doped cladding is limited only by the amorphous structure of silica (Rayleigh scattering) and was reached in 1986. Although *semiconductor lasers* were first made[17] in 1962, their use became practical only after 1970 when GaAs lasers operating continuously at room temperature were available.[18] Finally, it was only after the invention and perfection of the *Er-doped fibre amplifier*[19] in 1986 that optical communication became so powerful as it is today.[20,21]

In the following, the most basic point-to-point optical communication blocks are discussed: optical transmitters, waveguides, amplifiers and receivers.

Optical waveguiding is important for transmitters, receivers, and, naturally, for the channel itself. Because the mathematics of dielectric fibre waveguides is more involved, we start in Chapter 2 with an important, yet simply to be described device, namely the slab waveguide, which also represents a useful model for the waveguiding regions of optical sources, detectors and integrated passive components. Having understood the concept of waveguiding, the discussion of wave propagation in optical fibres may be kept simple.

The main component of an optical transmitter is its light source for *e*lectro-*o*ptic (EO) conversion. In Chapter 3, the principles of laser diodes and LED are explained. The characteristics of two types of optical amplifiers are discussed very briefly in Chapter 4, namely semiconductor optical amplifiers, and doped fibre amplifiers.

Chapter 5 explicates the properties of pin photodiodes. Chapter 6 reviews some important noise mechanisms. Finally, Chapter 7 explains their influence on optical reception.

---

[12] Kao, K. C.; Hockham, G. A.: Proc. IEE 113 (1966) 1151

[13] Werts, A.: Onde Electr. 45 (1966) 967

[14] Kapron, F. P.; Keck, D. .B; Maurer, A. D.: Appl. Phys. Lett. 17 (1970) 423

[15] Miya, T.; Terunuma, Y.; Hosaka, T.; Miyashita, T.: Ultimate low-loss single-mode fibre at 1.55 μm. Electron. Lett. 15 (1979) 106–108

[16] Kanamori, H.; Yokota, H.; Tanaka, G.; Watanabe, M.; Ishiguro, Y.; Yoshida, I.; Kakii, T.; Itoh, S.; Asano, Y.; Tanaka, S.: Transmission characteristics and reliability of pure-silica-core single-mode fibers. IEEE J. Lightwave Technol. LT-4 (1986) 1144–1149

[17] Nasledov, D. N.; Rogachev, A. A.; Ryvkin, S. M.; Tsarenkov, B. V.: Fiz. Tverd. Tela. 4 (1962) 1062 (Soviet Phys. Solid State 4 (1962) 782

[18] Alferov, Z.: IEEE Sel. Topics Quantum Electron. 6 (2000) 832

[19] Poole, S. B.; Payne, D. N.; Mears, R. J.; Fermann, M. E.; Laming, R. E.: J. Lightwave Technol. 4 (1986) 870

[20] *Why need amplifiers be distributed along a transmission distance?* Due to attenuation in the transmitting fibre the optical signal decays exponentially with the transmission span. Practical spans without amplification are about 70 km. Why are the spans so short (see Footnote 21 on Page 6)?

A transatlantic transmission from New York to London experiences an attenuation of about 1 400 dB (7 000 km @ 0.2 dB / km). Thus, for receiving one photon in London we have to inject 10^{140} photons into the optical fibre end in New York. If all the mass of our sun ($m_{\text{sun}} = 3\times10^{33}$ g) having an energy equivalent of $W_{\text{sun}} = mc^2 = 1.8\times10^{47}$ Ws could be converted into photons with a photon energy $hf = 6\times10^{-34}\,\text{Ws}^2 \times 200\,\text{THz} = 1.2\times10^{-19}$ Ws, we had generated 1.5×10^{66} photons at a wavelength of 1.55 μm ($f \approx 200$ THz), and could bridge a span with 660 dB loss, corresponding to a transmission distance of 3 300 km only. For a direct transmission New York – London we thus had to evaporate $10^{140}/10^{66} = 10^{74}$ suns.

This is quite a bit. The (observable) universe is estimated to have an extension of 1.4×10^{10} light years. Its mean density is supposed to be 3×10^{-30} g / cm^3 (`http://curious.astro.cornell.edu/question.php?number=342`). Therefore, the universe's mass (comprising not only suns) is $m_{\text{univ}} = 7\times10^{54}$ g, and its energy equivalent is $W_{\text{univ}} = m_{\text{univ}}c^2 = 6\times10^{68}$ Ws corresponding to 4.7×10^{87} photons at a wavelength of 1.55 μm. If we are able to receive one photon then the maximum span will be 877 dB /(0.2 dB / km) = 4 385 km. However, for bridging the distance New York – London in one go we had to burn $10^{140}/10^{87} = 10^{53}$ universes!

[21] Calculations stimulated by an oral presentation of N. J. Doran (S. K. Turitsyn, M. P. Fedoruk, N. J. Doran and W. Forysiak: Optical soliton transmission in fiber lines with short-scale dispersion management. 25th European Conference on Optical Communication (ECOC'99), Nice, France, September 26–30, 1999). — Universe's mass calculations and web address contributed by Dipl.-Phys. Jan Brückner, DFG Research Training Group 786 "Mixed Fields and Nonlinear Interactions" (`http://www.gkmf.uni-karlsruhe.de`), Karlsruhe, Germany, June 23, 2005

Chapter 2

Light waveguides

2.1 Fundamentals of wave propagation

Electromagnetic waves are described by magnetic and electric field vectors $\vec{H}$ and $\vec{E}$, by the electric displacement $\vec{D}$, the material polarization $\vec{P}$, and by the magnetic induction $\vec{B}$. These quantities are solutions of Maxwell's equations as functions of time and space. We assume that neither currents nor electric space charges are involved. At frequencies of interest the medium is isotropic, linear and non-magnetic, so that the medium properties are given by scalar, amplitude-independent quantities; the relative magnetic permeability is $\mu_r = 1$. Dielectric constant and magnetic permeability as well as velocity of light, wavelength for frequency f, angular frequency $\omega = 2\pi f$ in vacuum are ϵ_0, μ_0, $c = 1/\sqrt{\epsilon_0\mu_0}$, $\lambda = c/f$. The wave impedance of vacuum is $Z_0 = \sqrt{\mu_0/\epsilon_0}$. With these notations, Maxwell's equations are:

$$\begin{aligned} \operatorname{curl}\vec{H} &= \frac{\partial\vec{D}}{\partial t}\,, & \operatorname{curl}\vec{E} &= -\frac{\partial\vec{B}}{\partial t}\,, \\ \operatorname{div}\vec{D} &= 0\,, & \operatorname{div}\vec{B} &= 0\,, \\ \vec{D} &= \epsilon_0\vec{E} + \vec{P}\,, & \vec{B} &= \mu_0\vec{H}\,. \end{aligned} \tag{2.1}$$

All vector quantities are functions of time t and position vector $\vec{r}$. The time-frequency Fourier transform relation (FT $= \breve{\Psi}(f) = \mathcal{F}\{\Psi(t)\}$) and its inverse (IFT $= \Psi(t) = \mathcal{F}^{-1}\{\breve{\Psi}(f)\}$) are:

$$\Psi(t) = \int_{-\infty}^{+\infty} \breve{\Psi}(f)\,\mathrm{e}^{+\mathrm{j}\,2\pi f t}\,\mathrm{d}f, \qquad \breve{\Psi}(f) = \int_{-\infty}^{+\infty} \Psi(t)\,\mathrm{e}^{-\mathrm{j}\,2\pi f t}\,\mathrm{d}t\,. \tag{2.2}$$

Functions are often discriminated only by their argument, so that $\Psi(t) \neq \Psi(f = t)$, but $\Psi(f) \equiv \breve{\Psi}(f)$ holds.

2.1.1 Medium properties

A medium consists of atoms made up of positive and negative charges (e. g., ions and electrons). When an electric field is present, it separates the charges of opposite polarity (periodically in the case of a time periodic field).[1] This charge separation results in an additional electric field, called (induced) polarization. The polarization vector $\vec{P}$ in Eq. (2.1) follows $\vec{E}$ in real media with a time-delay which described by a causal influence function $\chi(t)$ where $\chi(t < 0) = 0$, describes relaxation processes of free and bound charges. This represents a "memory" present in each real-

[1] Hecht, E.: Optics. 2nd Ed. Reading: Addison Wesley 1987. See Chapter 3 Sect. 3.5.1 on Page 60

world medium, and we find:

$$\begin{aligned}
\vec{P}(t,\vec{r}) &= \epsilon_0 \int_0^\infty \chi(\tau,\vec{r})\,\vec{E}(t-\tau,\vec{r})\,\mathrm{d}\tau\,, \\
\vec{P}(f) &= \epsilon_0 \underline{\chi}(f)\vec{E}(f), \qquad \underline{\chi}(f) = \int_0^\infty \chi(t)\,\mathrm{e}^{-\mathrm{j}\,2\pi f t}\,\mathrm{d}t\,, \\
\underline{\chi}(f) &= \chi(f) + \mathrm{j}\,\chi_i(f) = \epsilon_r(f) - 1 - \mathrm{j}\,\epsilon_{ri}(f), \quad \underline{\chi}(f) = \underline{\chi}^*(-f).
\end{aligned} \tag{2.3}$$

The frequency-dependent proportionality constant between the spectra $\vec{E}(f)$ and $\vec{P}(f)$ is an analytic function[2] called "electric susceptibility" $\underline{\chi}(f)$ (real part $\chi(f)$, imaginary part $\chi_i(f)$); it defines a complex relative dielectric constant $\bar{\epsilon}_r$ (real part $\epsilon_r(f)$, imaginary part $-\epsilon_{ri}(f)$). A complex refractive index $\bar{n}$ (real part $n(f)$, imaginary part $-n_i(f)$) is determined from the relation $\bar{\epsilon}_r = \bar{n}^2$:

$$\begin{aligned}
&\bar{n} = n - \mathrm{j}\,n_i, && \bar{\epsilon}_r = \epsilon_r - \mathrm{j}\,\epsilon_{ri}, \\
&\epsilon_r = n^2 - n_i^2, && \epsilon_{ri} = 2nn_i, \\
&n^2 = \tfrac{1}{2}\epsilon_r\left(1 + \sqrt{1 + \epsilon_{ri}^2/\epsilon_r^2}\right), && n_i = \epsilon_{ri}/(2n), \\
&n \approx \sqrt{\epsilon_r} \quad (\text{for } |\epsilon_{ri}| \ll \epsilon_r) && n_i \approx \epsilon_{ri}/(2\sqrt{\epsilon_r}), \\
&n \approx \sqrt{|\epsilon_{ri}|/2} \quad (\text{for } |\epsilon_{ri}| \gg \epsilon_r) && n_i \approx \mathrm{sgn}(\epsilon_{ri})\sqrt{|\epsilon_{ri}|/2}\,.
\end{aligned} \tag{2.6}$$

For $\epsilon_{ri}, n_i > 0$ the medium has losses, if $\epsilon_{ri}, n_i < 0$ it amplifies, and with $\epsilon_{ri}, n_i = 0$ it is said to be transparent. Because of causality, the susceptibility $\underline{\chi}(f)$ is an analytic spectral function, and the so-called Kramers-Kronig relation follows,[3]

$$-\epsilon_{ri}(f) = \mathcal{H}_F\{\epsilon_r(f) - 1\} = \frac{1}{\pi}\,\mathcal{P}\int_{-\infty}^{+\infty} \frac{\epsilon_r(f\,') - 1}{f\,' - f}\,\mathrm{d}f\,' \tag{2.7}$$

For $\chi(f) = \epsilon_r(f) - 1 = \mathrm{const}_f$, we have from Eq. (2.7) $\epsilon_{ri}(f) = 0$ and therefore $\chi(t) = (\epsilon_r - 1)\,\delta(t)$, i. e., there is no memory, and therefore neither a medium nor a polarization. Real media always have a memory, so with Eq. (2.3) a frequency dependence of χ, χ_i is unavoidable. However, in certain frequency ranges χ may be a constant and $\chi_i = 0$. For this regime only the medium may be described by a real and constant refractive index. That this holds true for all frequencies (as we have seen this is valid only for vacuum) is implicitly assumed in the usual ansatz for $\vec{D}(t)$,

$$\vec{D} = \epsilon_0\epsilon_r\vec{E}, \qquad n = \sqrt{\epsilon_r}\,. \tag{2.8}$$

Typical frequency dependencies of n and n_i for freely movable and for bound charges are depicted in Fig. 2.1.

[2]Definition of analytic (or regular or holomorphic) functions: Consider a causal time function $\Psi(t < 0) = 0$. Then, the Fourier spectrum of $\Psi(t)$ is a complex analytic function $\breve{\underline{\Psi}}(f)$ with real part $\breve{\Psi}(f)$ and imaginary part $\breve{\Psi}_i(f)$,

$$\breve{\Psi}_i(f_0) = \frac{1}{\pi}\,\mathcal{P}\int_{-\infty}^{+\infty} \frac{\breve{\Psi}(f)}{f - f_0}\,\mathrm{d}f, \qquad \breve{\Psi}(f_0) = -\frac{1}{\pi}\,\mathcal{P}\int_{-\infty}^{+\infty} \frac{\breve{\Psi}_i(f)}{f - f_0}\,\mathrm{d}f. \tag{2.4}$$

Real and imaginary parts are interconnected (because of causality) by the so-called Hilbert transform and its inverse, $\breve{\Psi} = \mathcal{H}_F\{\breve{\Psi}_i\}$ and $\breve{\Psi}_i = \mathcal{H}_F^{-1}\{\breve{\Psi}\}$ in Eq. (2.4). For a general function $f(x)$, Cauchy's principle value integral ($\mathcal{P}$ means *valor principalis*, Latin for principle value) is defined by

$$\mathcal{P}\int_{-\infty}^{+\infty} \frac{f(x)}{x - x_0}\,\mathrm{d}x = \lim_{\varepsilon\to 0}\left(\int_{-\infty}^{x_0-\varepsilon} \frac{f(x)}{x - x_0}\,\mathrm{d}x + \int_{x_0+\varepsilon}^{+\infty} \frac{f(x)}{x - x_0}\,\mathrm{d}x\right). \tag{2.5}$$

[3]See Eq. (2.4) Footnote 2 on Page 8

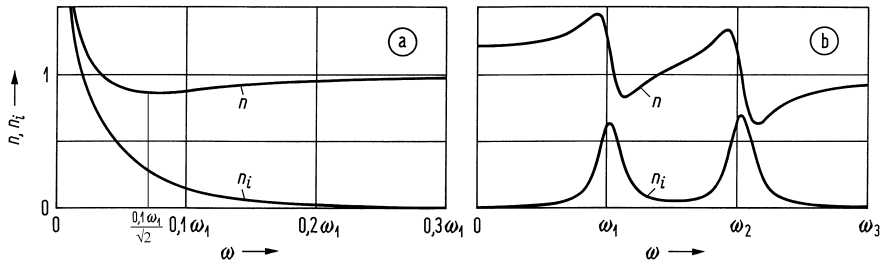


Fig. 2.1. Real part n und negative imaginary part (n_i) of complex refractive index $\bar{n} = n - \mathrm{j}\, n_i$: Frequency dependence (a) Free carriers only (b) Two collectives of bound charges with high mass (ions, low angular resonance frequency ω_1) and low mass (electrons, high angular resonance frequency ω_2)

For a model of a conductor with freely movable charge carriers, Fig. 2.1(a) shows the typical dependencies of n and n_i. Large frequencies lead to real n and small losses (or gain) $|n_i| \ll 1$, so the conductor is transparent. The larger the carrier concentration n_T is, the deeper the dent in the n-curve becomes. For a suitable carrier concentration, the dielectric constant can even vanish at the so-called plasma frequency[4] $f_P \sim \sqrt{n_T}$, $\epsilon_r(f_P) = 0$, and be negative below, $\epsilon_r(f < f_P) < 0$. In this case a penetrating wave drops off exponentially.

For a model of a dielectric medium, only two resonant charge collectives are regarded, Fig. 2.1(b). There is a low-loss frequency range with *normal*[5] *dispersion* $\mathrm{d}n/\,\mathrm{d}\omega > 0$, $\mathrm{d}n/\,\mathrm{d}\lambda < 0$. Absorption occurs in the frequency ranges near the resonances $\omega_{1,2}$ with *anomalous*[6] *dispersion* $\mathrm{d}n/\,\mathrm{d}\omega < 0$, $\mathrm{d}n/\,\mathrm{d}\lambda > 0$,

$$\begin{aligned} &\frac{\mathrm{d}n}{\mathrm{d}\omega} > 0\,, \quad \frac{\mathrm{d}n}{\mathrm{d}\lambda} < 0 \qquad \text{normal dispersion} \\ &\frac{\mathrm{d}n}{\mathrm{d}\omega} < 0\,, \quad \frac{\mathrm{d}n}{\mathrm{d}\lambda} > 0 \qquad \text{anomalous dispersion} \end{aligned} \tag{2.9}$$

In the region of normal dispersion, the deviation angle of a dispersing prism decreases continuously with λ, while near the infrared (ω_1) and ultraviolet (ω_2) absorption lines in the region of anomalous dispersion unusual dependencies of the deviation angle are to be observed.

2.1.2 Wave equation

Reshaping of Eqs. (2.1), (2.8) leads to the exact vector wave equations for the real instantaneous quantities $\vec{E}$, $\vec{H}$,

$$\begin{aligned} \nabla^2\vec{E} + \operatorname{grad}\left((\operatorname{grad}\ln n^2)\cdot\vec{E}\right) &= \frac{n^2}{c^2}\frac{\partial^2\vec{E}}{\partial t^2}\,, \\ \nabla^2\vec{H} + (\operatorname{grad}\ln n^2)\times\operatorname{curl}\vec{H} \quad &= \frac{n^2}{c^2}\frac{\partial^2\vec{H}}{\partial t^2}\,. \end{aligned} \tag{2.10}$$

The differential operator[7] ∇^2 (del square) operates on a vector $\vec{E}$, and is not identical with the Laplace operator $\triangle$ as applied to a scalar function Ψ,

$$\begin{aligned} \nabla^2\vec{E} &= \operatorname{grad}\operatorname{div}\vec{E} - \operatorname{curl}\operatorname{curl}\vec{E}, \\ \triangle\Psi &= \operatorname{div}\operatorname{grad}\Psi. \end{aligned}$$

[4] See Chapter 4 Sect. 4.3.5 Eq. (4.81) on Page 112 in reference Footnote 1 on Page 7

[5] Latin *norma*, carpenter's square, i. e., an L-shaped or T-shaped instrument used for obtaining or testing right angles; conforming to a standard

[6] Greek ἀν-ώμαλος, un-even, deviant, different

[7] Morse, P. M.; Feshbach, H.: Methods of theoretical physics, Volume 1 and 2. New York: McGraw-Hill 1953. End of Chapter 1 on Page 116

Vector components are subscripted with their corresponding coordinates. In Cartesian coordinates x, y, z we have the unit vectors $\vec{e}_x, \vec{e}_y, \vec{e}_z$. Only in Cartesian coordinates the following relations hold:

$$\begin{aligned}\nabla^2\vec{E} &= \operatorname{grad}\operatorname{div}\vec{E} - \operatorname{curl}\operatorname{curl}\vec{E} = \\ &= \vec{e}_x \operatorname{div}\operatorname{grad} E_x + \vec{e}_y \operatorname{div}\operatorname{grad} E_y + \vec{e}_z \operatorname{div}\operatorname{grad} E_z = \\ &= \vec{e}_x\nabla^2 E_x + \vec{e}_y\nabla^2 E_y + \vec{e}_z\nabla^2 E_z = \vec{e}_x\triangle E_x + \vec{e}_y\triangle E_y + \vec{e}_z\triangle E_z.\end{aligned} \tag{2.11}$$

2.1.3 Homogeneous medium

The differential equations for components of $\vec{E}$ and $\vec{H}$ in inhomogeneous media are coupled. Now we assume a lossless homogeneous medium $n = \text{const}$ leading to $\operatorname{grad}\ln n^2 = 0$ in Eq. (2.10). In Cartesian coordinates, we find 2×3 decoupled wave equations for three Cartesian components each for both field vectors $\vec{E}$ and $\vec{H}$. In cylindrical coordinates r, φ, z, only the differential equation for the z-components becomes decoupled from the other ones. Scalar wave equations with identical structure result for these decoupled components,

$$\begin{aligned}\Psi(t,x,y,z) &= E_q(t,x,y,z), H_q(t,x,y,z), \qquad q = x, y, z\,, \\ \Psi(t,r,\varphi,z) &= E_z(t,r,\varphi,z), H_z(t,r,\varphi,z), \\ \nabla^2\Psi(t,\vec{r}) &= \frac{n^2}{c^2}\frac{\partial^2}{\partial t^2}\Psi(t,\vec{r}).\end{aligned} \tag{2.12}$$

The other field components may be derived via Maxwell's equations from the z-components. Further interdependencies of the vector components are given by initial and boundary conditions which define the problem. For the following we assume Cartesian coordinates x, y, z.

Monochromatic waves

Consider the monochromatic wave (A, φ, φ_i real)

$$\Psi(t,\vec{r}) = A(\vec{r})\,\mathrm{e}^{\mathrm{j}[\omega t - \varphi(\vec{r})]} = \mathrm{e}^{\mathrm{j}\,\omega t}\,\mathrm{e}^{-\mathrm{j}\,\bar{\varphi}(\vec{r})}, \quad \bar{\varphi}(\vec{r}) = \varphi(\vec{r}) + \mathrm{j}\,\varphi_i(\vec{r}) \tag{2.13}$$

$A(\vec{r})$ and $\varphi(\vec{r})$ are amplitude and phase of the wave. Surfaces $A(\vec{r}) = \text{const}$ (or $\varphi_i(\vec{r}) = \text{const}$) are called amplitude surfaces. Surfaces $\varphi(\vec{r}) = \text{const}$ are denoted as phase surfaces. The amplitude decreases fastest in the direction of the amplitude vector $\vec{a} = -\operatorname{grad}\varphi_i$, the phase increases fastest in direction of the phase vector $\vec{b} = \operatorname{grad}\varphi$. The propagation vector $\bar{\varphi}$ is defined as (real $\vec{a}$, $\vec{b}$)

$$\operatorname{grad}\bar{\varphi} = \operatorname{grad}\varphi + \mathrm{j}\operatorname{grad}\varphi_i = \vec{b} - \mathrm{j}\,\vec{a}\,. \tag{2.14}$$

The wave does not propagates in the direction of the propagation vector $\operatorname{grad}\bar{\varphi}$, but in the direction of the phase vector $\vec{b}$ normal to surfaces of constant phase (unit vector $\vec{e}_b = \vec{b}/|\vec{b}| = \operatorname{grad}\varphi/|\operatorname{grad}\varphi|$). The velocity of the phase surface (phase front) in propagation direction is the phase velocity v of wave. From $\omega\,\mathrm{d}t - \operatorname{grad}\varphi\cdot\mathrm{d}\vec{r} = 0$, $\mathrm{d}\vec{r} = \vec{e}_b\,\mathrm{d}s$ we find

$$\omega\,\mathrm{d}t = \frac{(\operatorname{grad}\varphi)\cdot(\operatorname{grad}\varphi)}{|\operatorname{grad}\varphi|}\,\mathrm{d}s = \frac{|\operatorname{grad}\varphi|^2}{|\operatorname{grad}\varphi|}\,\mathrm{d}s \;\rightsquigarrow\; v = \frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\omega}{|\operatorname{grad}\varphi|} = \frac{\omega}{|\vec{b}|}\,. \tag{2.15}$$

Waves are classified according to the shape of their phase surfaces. Plane phase surfaces belong to plane waves. We call plane waves "homogeneous", if phase and amplitude surfaces coincide.

Plane waves We solve the wave equation (2.12) in a homogeneous medium with the separation ansatz

$$\Psi(t,\vec{r}) = \exp(\mathrm{j}\,\omega t)\exp\bigl[-\mathrm{j}\,\vec{k}\cdot\vec{r}\bigr] = \exp(\mathrm{j}\,\omega t)\exp\bigl[-\mathrm{j}(k_x x + k_y y + k_z z)\bigr], \tag{2.16}$$

which is valid for

$$\begin{aligned}(-k_x^2 - k_y^2 - k_z^2)\,\Psi(t,\vec{r}) &= -n^2\left(\frac{\omega}{c}\right)^2\Psi(t,\vec{r}),\\ k_x^2 + k_y^2 + k_z^2 &\stackrel{!}{=} n^2\left(\frac{\omega}{c}\right)^2 = n^2k_0^2,\end{aligned} \tag{2.17}$$

The separation condition Eq. (2.17) leads to a real separation constant k_0, thus

$$\bar{\varphi}(\vec{r}) = \vec{k}\cdot\vec{r}, \qquad \vec{k}^{\,2} = \vec{k}\cdot\vec{k} = n^2k_0^2 \quad \text{(real!)}, \qquad k_0 = \frac{\omega}{c}\,. \tag{2.18}$$

With the definition of the propagation vector Eq. (2.14) we see that

$$\operatorname{grad}\bar{\varphi} = \vec{k}, \quad \vec{k} = \vec{b} - \mathrm{j}\,\vec{a}\,, \quad \vec{k}^{\,2} = \vec{b}^{\,2} - \vec{a}^{\,2} - \mathrm{j}\,2\,\vec{a}\cdot\vec{b} \text{ (real!)} \;\rightsquigarrow\; \vec{a}\cdot\vec{b} = 0\,. \tag{2.19}$$

Because $\vec{k}^{\,2}$ is real, either $|\vec{a}| = 0$ (homogeneous plane wave) holds, or, more generally, the phase vector points perpendicularly to the amplitude vector, $\vec{b} \perp \vec{a}$. The amplitude decreases fastest normal to the propagation direction; such a "disappearing" wave is called evanescent in direction of $\vec{a}$.

Because $\vec{E}$ and $\vec{H}$ are source free, Eq. (2.1), we find

$$\vec{E}\cdot\vec{k} = 0\,, \quad \vec{H}\cdot\vec{k} = 0\,.$$

For example: From Eq. (2.13), (2.19) and $\vec{E}(\vec{r}) = \vec{E}\,\mathrm{e}^{-\mathrm{j}\,k\cdot\vec{r}}$ the relation $\operatorname{div}\vec{E} = -\mathrm{j}\,k_xE_x - \mathrm{j}\,k_yE_y - \mathrm{j}\,k_zE_z \mathrel{\widehat{=}} -\mathrm{j}\,\vec{E}\cdot\vec{k} \stackrel{!}{=} 0$ follows. From Eq. (2.1) we see that $\vec{E}$, $\vec{H}$ and $\vec{k}$ are orthogonal and right handed in this sequence (e. g., $\operatorname{curl}\vec{H} \perp \vec{H}$, $\operatorname{curl}\vec{H} \parallel \vec{D}$, $\vec{E} \perp \vec{H}$). Generally, $\vec{E}$ and $\vec{H}$ possess components in the direction of the phase vector $\vec{b}$, i. e., in the direction of propagation.

Homogeneous plane waves For homogeneous plane waves, propagation vector $\operatorname{grad}\bar{\varphi}$ and phase vector $\vec{b}$ are identical and real, so that $\vec{b}\cdot\vec{E} = 0$, $\vec{b}\cdot\vec{H} = 0$. The vectors $\vec{E}$, $\vec{H}$, $\vec{b}$ form a right handed orthogonal system in this sequence. Because *e*lectric and *m*agnetic field are *t*ransverse ($\vec{E} \perp \vec{b}$, $\vec{H} \perp \vec{b}$, normal to the propagation direction $\vec{b}$) we call it a TEM wave. $\vec{E}$ and $\vec{H}$ are in phase, and the wave impedance Z_0/n of the medium (ratio of two orthogonal components of $\vec{E}$ and $\vec{H}$) is real. The Poynting[8] vector $\vec{\mathcal{S}}$ is generally complex, but for homogeneous plane waves it becomes real,

$$\vec{\mathcal{S}} = \frac{1}{2}\,\vec{E}\times\vec{H} = \frac{1}{2}\,\frac{n}{Z_0}\left|\vec{E}\right|^2\vec{e}_b\,. \tag{2.20}$$

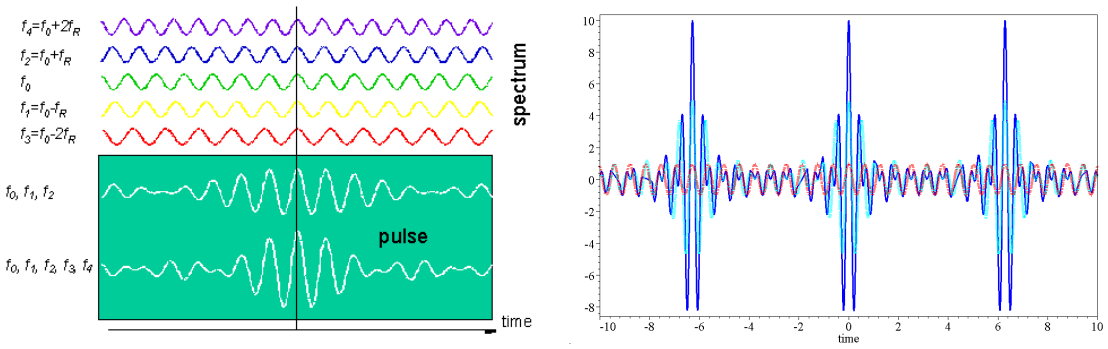
Phase and group velocity

For homogeneous plane waves the phase velocity was defined in Eq. (2.15). Further quantities of interest which will be discussed later in more detail are the group velocity v_g, and the group delay time t_g with respect to the geometrical propagation length L,

$$\begin{aligned}v = \frac{\omega}{k} = \frac{c}{n}\,, \qquad v_g = \frac{\mathrm{d}\omega}{\mathrm{d}k} = \frac{c}{n_g}\,, \qquad n_g = n + f\frac{\mathrm{d}n}{\mathrm{d}f} = n - \lambda\frac{\mathrm{d}n}{\mathrm{d}\lambda}\,,\\ \frac{t_g}{L} = \frac{\mathrm{d}k}{\mathrm{d}\omega} = \frac{n_g}{c}\,, \qquad \frac{\mathrm{d}n_g}{\mathrm{d}\lambda} = -\lambda\frac{\mathrm{d}^2n}{\mathrm{d}\lambda^2}\,.\end{aligned} \tag{2.21}$$

The group refractive index n_g represents an effective refractive index for the propagation of a wave group in a dispersive medium, its derivative measures the group delay time difference.

[8] John Henry Poynting, physicist, ⋆ Monton (near Manchester) 9.9.1852, † Birmingham 30.3.1914



(a) Superposition of three and five sinusoidals

(b) Superposition of one, five and ten sinusoidals

Fig. 2.2. Signal localized in time, superposition of n sinusoidals with equidistant frequencies. (a) $N = 3, 5$ for a limited time span (b) $N = 1, 5, 10$ for a more extended time span. If the frequency increment $f_R \to 0$ decreases and the number of sinusoidal $N \to \infty$ increases, the peaks become narrower and their time distance in (b) approaches infinity, so that a single Dirac impulse $\delta(t)$ remains.

Physical interpretation of group velocity How do we transport information? The answer is, by signals localized in time. Why so? Because otherwise there is no surprise: At one point of time there is some level of light, at another one we find another level, and this carries the information. Information theory[9] as developed by Shannon[10] says that the less probable (or the less predictable) an event is the more information it carries. In this sense a true sinusoidal (or any periodic signal) extending indefinitely and invariably over all time $-\infty < t < +\infty$ does not transport any information at all. However, the superposition of sinusoidals with different frequencies *is* localized in time as may be seen from Fig. 2.2. Analytically, this may be deduced from the sum formula[11]

$$\sum_{i=0}^{N-1} \cos\left(\omega_0 + i\omega_R\right) t = \frac{\sin\left(N\omega_R t/2\right)}{\sin\left(\omega_R t/2\right)} \cos\left(\omega_0 + \tfrac{N-1}{2}\,\omega_R\right) t\,. \tag{2.22a}$$

We recognize a carrier with mean angular frequency $2\pi f = \omega = \omega_0 + \frac{N-1}{2}\,\omega_R$ and envelope $\sin\left(N\omega_R t/2\right) / \sin\left(\omega_R t/2\right)$. For $\omega_0 = 0$, $\omega_R \to 0$, $N \to \infty$, $N\omega_R = \text{const}$, with $\cos(x-y) = \cos x \cos y + \sin x \sin y$ and $\sin x \cos x = \frac{1}{2}\sin(2x)$, and because $\int_{-Nf_R}^{+Nf_R} \sin\left(2\pi f t\right) \mathrm{d}f = 0$ due to the unsymmetrical integrand, we may write with the help of $\cos\left(2\pi f t\right) = \exp\left(-\mathrm{j}\,2\pi f t\right) + \mathrm{j}\sin\left(2\pi f t\right)$:

$$\lim_{\substack{\omega_R \to 0 \\ N \to \infty \\ N\omega_R = \text{ const}}} \frac{\omega_R}{\pi} \sum_{i=0}^{N-1} \cos\left(i\omega_R t\right) = \frac{N\omega_R}{\pi} \frac{\sin(N\omega_R t)}{N\omega_R t} \underset{N\omega_R \to \infty}{=} \lim_{Nf_R \to \infty} \int_{-Nf_R}^{+Nf_R} \mathrm{e}^{-\mathrm{j}\,2\pi f t}\,\mathrm{d}f = \delta(t). \tag{2.22b}$$

Equation (2.22b) expresses the well-known fact that the inverse Fourier transform Eq. (2.2) of a constant spectrum with amplitude 1 is represented by the symbolic function[12] (Dirac[13] impulse) $\delta(t)$.

[9] See reference Footnote 8 on Page 2

[10] See Footnote 7 on Page 2.. — The amount of information received from the environment by human receptors is in the order of 10 Gbit/s. However, the processing capacity of the brain is only about 25 bit/s, so during the recognition process the amount of information regarded as unecessary for the present purpose (the "redundancy") is significantly reduced.

[11] Gradstein, I.; Ryshik, I.: Summen-, Produkt- und Integral-Tafeln, 5th Russian Ed., 1st German-English Ed. Volume 1 and 2. Thun: Harri Deutsch 1981. Formula 1.341.3: $\sum_{k=0}^{n-1} \cos\left(x + ky\right) = \cos\left(x + (n-1)\,y/2\right) \sin\left(ny/2\right) / \sin\left(y/2\right)$

[12] Lighthill, M. J.: Einführung in die Theorie der Fourier-Analysis und der verallgemeinerten Funktionen. Mannheim: Bibliographisches Institut 1966. Kapitel 2

[13] Paul Adrien Maurice Dirac, physicist, ⋆ Bristol 8.8.1902, † Tallahassee 20.10.1984 (Florida). Nobel prize 1933 (together with E. Schrödinger)

Consider the superposition of $N = 2$ homogeneous plane waves, which propagate along the z-direction according to Eq. (2.13) ($\cos(x+y) = 2\cos(x-y)\cos(x+y)$),

$$\begin{aligned}\psi(t,z) &= \cos(\omega_1 t - k_1 z) + \cos(\omega_2 t - k_2 z) \\ &= 2\cos\left(\frac{\omega_1-\omega_2}{2}\,t - \frac{k_1-k_2}{2}\,z\right)\cos\left(\frac{\omega_1+\omega_2}{2}\,t - \frac{k_1+k_2}{2}\,z\right).\end{aligned} \tag{2.22c}$$

As in Eq. (2.22a) we see a carrier with mean angular frequency $\omega = (\omega_1+\omega_2)/2$ having a mean propagation constant $k = (k_1+k_2)/2$, and a signal envelope with differential frequency $\Delta\omega = (\omega_1-\omega_2)/2$ and differential propagating constant $\Delta k = (k_1-k_2)/2$. Therefore, according to Eq. (2.21), the carrier propagates with a phase velocity v, and the envelope with a generally different group velocity v_g,

$$v = \frac{\omega}{k} = \frac{c}{n}\,, \qquad v_g = \frac{\Delta\omega}{\Delta k} = \frac{c}{n_g}\,. \tag{2.22d}$$

If in a medium waves at different frequencies propagate with different phase velocities, then the constituting waves of an impulse as in Fig. 2.2 do not necessarily maintain the same relative position along z, so that the shape of the envelope may be distorted.

2.1.4 Properties of silica glass

Fused silica glass (SiO_2) is named quartz. It is today's material for high quality light waveguides exhibiting low attenuation and high transmission bandwidth. A model for the basic frequency dependence of the refractive index of glass is seen in Fig. 2.1(b) and will be further discussed below. By intentionally introducing impurities the refractive index may be changed. Doping materials are F for decreasing the refractive index, and GeO_2 for an increase.

Attenuation

Losses arise through scattering and absorption. Therefore, the power of a homogeneous plane wave decreases in z-direction from its initial value P_0 according to

$$P(z) = P_0\,\mathrm{e}^{-\alpha z}, \qquad a = 10\lg\frac{P_0}{P(z)} = \alpha z\,10\lg\mathrm{e} = 4.34\,\alpha z. \tag{2.23}$$

The power attenuation constant α (unit km^{-1}) is usually expressed by specifying the attenuation a (unit dB).

Scattering In a glass melt there is a random movement of its constituents. At the transformation temperature T_g the melt becomes glass, and actual irregularities "freeze". The higher T_g, the more irregularities are to be expected (quartz: $T_g \approx 1\,180\,^\circ\mathrm{C}$; multi-component glasses, e. g., sodium-silica glass: $T_g \approx 500\,^\circ\mathrm{C}$). Irregularities in the amorphous structure of the glass which are much smaller than λ give rise to so-called Rayleigh[14] scattering: Localized density fluctuations behave as Hertzian dipoles for nearly isotropic secondary radiation, thereby removing power from the signal path with a strongly wavelength-dependent power attenuation $\alpha_S \sim \lambda^{-4}$. Rayleigh scattering by an amorphous glass matrix is not avoidable. Doping with GeO_2 increases the refractive index $n_1 = n_2/\sqrt{1-2\Delta} \approx n_2(1+\Delta)$ by the relative refractive index difference Δ, but it also increases losses by Rayleigh scattering because of additional inhomogeneous regions of size $\ll \lambda/n$. With Ge-doping the power attenuation a_S is

$$\frac{a_S/z}{\mathrm{dB}\,/\,\mathrm{km}} = \frac{A(1+B\Delta)}{(\lambda/\,\mu\mathrm{m})^4}\,, \quad A = 0.63\,, \quad B = 180 \pm 35\,, \quad \Delta < 0.4\,\%\,. \tag{2.24}$$

[14] John William Strutt, 3. Baron Rayleigh (since 1873), physicist, ⋆ Langford Grove (County Essex) 12.11.1842, † Witham (County Essex) 30.6.1919. Professor in Cambridge and London. Nobel prize in 1894 (together with W. Ramsay)

For undoped quartz glass at $\lambda = 1\,\mu$m the attenuation then is $a_S/z = 0.63\,\mathrm{dB}\,/\,\mathrm{km}$. For a wider range $\Delta < 2\,\%$ and Ge-doped quartz glass fibres the constants in Eq. (2.24) were reported to be $A = 0.8$ and $B = 100$ or $A = 0.73 \ldots 0.78$ for undoped quartz glass. Using these parameters for a doping leading to $0.2\,\% < \Delta < 0.4\%$ the attenuation is estimated to be larger by $0.05\,\mathrm{dB}\,/\,\mathrm{km}$ compared to Eq. (2.24), but for $\Delta = 0$ the parameter set Eq. (2.24) coincidences with bulk material measurements and is therefore to be prefered.

Scattering at inhomogeneities of size larger than λ is called Mie[15] scattering. It is caused by gas bubbles, micro corrosion cracks and other irregularities, but may be avoided in principle.

Rayleigh and Mie scattering are linear processes, where a partial power of a propagating wave is transferred to a different direction due to an inhomogeneity of the refractive index (Rayleigh) or of any surface involved (Mie).

Nonlinear scattering Brillouin and Raman scattering are nonlinear phenomena, where a partial power of a propagating field is transferred to a field of a different frequency.

Brillouin scattering can be considered as a light carrier modulation by thermal molecular vibrations. As a result, the frequency of the modulated light is shifted down from the incident light carrier frequency having a linewidth smaller than 38 MHz. At a carrier wavelength of $\lambda = 1\,\mu$m the downshift is 15 GHz.

Raman scattering is similar to Brillouin except that the amount of the frequency shift is in the optical frequency range. For a carrier linewidth of less than 15 THz, the downshift is 13.5 THz.

For Brillouin and Raman nonlinear interactions the product between the local light intensity $I(f,z) \sim |E(f,z)|^2$ and length increments $\mathrm{d}z$ of the fibre is important. The decreasing intensity $I = I_0 \exp(-\alpha z)$ in a light waveguide has to be taken into account. For nonlinear interaction, the effective length L_{eff} is defined by

$$L_{\mathrm{eff}} = \int_0^L \frac{I\,\mathrm{d}z}{I_0} = \left(1 - \mathrm{e}^{-\alpha L}\right)/\,\alpha, \quad L_{\mathrm{eff}} = 1/\alpha \quad \text{for } L \to \infty. \tag{2.25}$$

For long fibres ($L \to \infty$) with an attenuation $a/L = 1\,\mathrm{dB}\,/\,\mathrm{km}$, the effective length is $L_{\mathrm{eff}} = 4.34\,\mathrm{km}$.

Absorption In non-conducting material like quartz glass, absorption results from exciting bound charges in atoms and molecules to vibrate (*intrinsic loss*), see Fig. 2.1(b). Bound electrons have a very high excitation energy $\hbar\omega_2 = 8.9\,\mathrm{eV}$ (band gap of amorphous SiO_2), and therefore the electronic absorption maximum $\lambda_2 = 140\,\mathrm{nm}$ ($f_2 = 2\,140\,\mathrm{THz}$) is far in the UV region. Molecule resonances of SiO_2 are found in the IR and have an absorption maximum at $8\,\mu$m (37.5 THz). Harmonics and combination wavelength up to $\lambda_1 = 3\,\mu$m (100 THz) set an upper limit to the useful λ-range, see Fig. 2.1(b).

The most important *extrinsic losses* are caused by atomic resonances of particles, which do not belong to the basic glass matrix. Impurities caused by ions of transition metals (Cu^{2+}, Fe^{2+}, Ni^{2+}, V^{3+}, Cr^{3+}, Mn^{3+}) give rise to further absorption resonances. Today's technology allows relative impurity weight levels below 1 ppb ($1\,\mathrm{ppb} \mathrel{\widehat{=}} 10^{-9}$; the large evaporation rate for high-melting point glasses helps), so these absorptions are not important any more. Multi-component glasses are more difficult to purify and have higher attenuation.

Most important is the absorption loss due to water or the (OH^-)-bond, whose fundamental resonance wavelength is at $\lambda_{\mathrm{OH}} = 2.72\,\mu$m (110 THz). This produces harmonics and combination wavelengths at $\lambda_{\mathrm{OH}} = 2.22,\ 1.90,\ 1.38,\ 1.24,\ 1.13,\ 0.945,\ 0.88\,\mu$m. With MCVD technology (modified chemical vapour deposition), relative (OH^-)-weights of 0.1 ppm ($1\,\mathrm{ppm} \mathrel{\widehat{=}} 10^{-6}$) are possible. VAD technology (vapour axial deposition) with dehydrated preform sets a lower limit of 1 ppb. Typical loss peaks at $\lambda_{\mathrm{OH}} = 1.38,\ 1.24,\ 0.945\,\mu$m are $a/z = 5.4,\ 0.23,\ 0.083\,\mathrm{dB}\,/\,\mathrm{km}$ (MCVD) and $a/z = 0.054,\ 0.0023,\ 0.00083\,\mathrm{dB}\,/\,\mathrm{km}$ (VAD).

[15] Gustav Mie, physicist, ⋆1868, † 1957

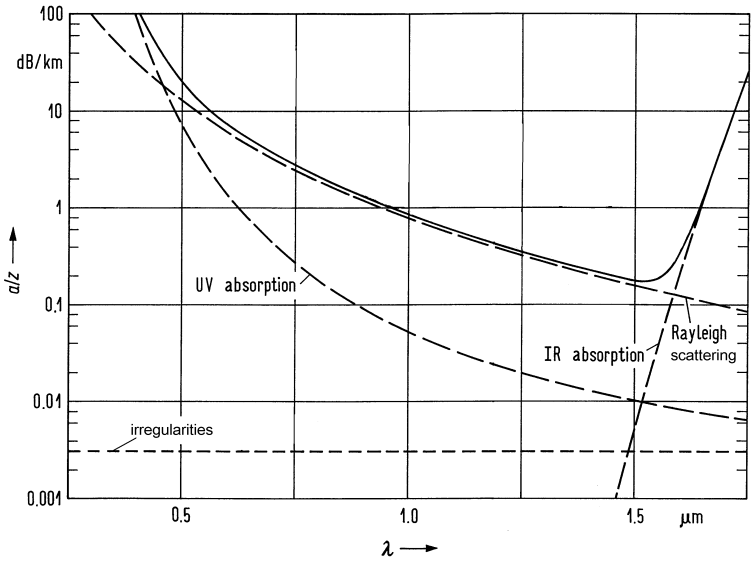


Fig. 2.3. Attenuation of Ge-doped ($\Delta = 0.25\,\%$) singlemode fibre. VAD: produced with VAD-technology MCVD: produced with MCVD-technology

Figure 2.3 shows measured values a/z (GeO_2-doped single-mode quartz glass fibres) broken down to the theoretical types of losses, compare Fig. 2.1(b). At $\lambda = 0.85,\ 1.3,\ 1.55\,\mu$m an attenuation of $a/z = 2.2,\ 0.35,\ 0.2\,\mathrm{dB\,/\,km}$ is feasible. Top results (undoped SiO_2-core with minimum Rayleigh scattering, matched F-SiO_2-cladding, $\Delta = 0.3\,\%$) for $\lambda = 1.3,\ 1.55\,\mu$m are $a/z = 0.291,\ 0.154\,\mathrm{dB\,/\,km}$. The basic *attenuation limit*[16] of quartz glass can be reached routinely at the wavelength λ_α,

$$\lambda_\alpha = 1.55\,\mu\mathrm{m} \qquad \text{with} \qquad a/z|_{\lambda_\alpha} = 0.154\,\mathrm{dB\,/\,km}\,.$$

Dispersion

According to Sect. 2.1.1 a real medium has losses (or possibly gain), which leads to $n(f)$, $v(f)$, and $v(f) \neq v_g(f)$. This frequency dependence is called "dispersion".[17] For different frequency components of a signal we have different group delay times t_g or group velocities v_g (Eq. (2.21), Fig. 2.2), and the signal becomes distorted after some propagation length L. Because of waveguiding properties, there are further influences on the group delay which spread the signal delay times. *In optical communications, the notion "dispersion" embraces all effects, which lead to a group delay time (or group velocity) spread.* If changes of the group delay with varying f are regarded, it is called "chromatic dispersion", sometimes also *g*roup *v*elocity *d*ispersion (GVD).

The group delay difference of two plane waves at optical carriers differing in wavelength λ by $\Delta\lambda = \lambda - \lambda_1$ are

$$\begin{aligned} t_g(\lambda) &= t_g(\lambda_1) + \frac{\mathrm{d}t_g}{\mathrm{d}\lambda}(\lambda - \lambda_1) + \frac{1}{2!}\frac{\mathrm{d}^2 t_g}{\mathrm{d}\lambda^2}(\lambda - \lambda_1)^2 + \dots\,, \\ \Delta t_g &= t_g(\lambda) - t_g(\lambda_1) = \frac{\mathrm{d}t_g}{\mathrm{d}\lambda}\Delta\lambda + \frac{1}{2!}\frac{\mathrm{d}^2 t_g}{\mathrm{d}\lambda^2}(\Delta\lambda)^2 + \dots\,. \end{aligned} \tag{2.26}$$

This is called "material dispersion". Because of Eq. (2.21), the length related group delay difference

[16] See reference Footnote 16 on Page 6
[17] Latin *dispersio*

results,

$$\Delta t_g/L = M\,\Delta\lambda + N\,(\Delta\lambda)^2 + \ldots, \qquad M = \frac{1}{c}\frac{\mathrm{d}n_g}{\mathrm{d}\lambda}, \quad N = \frac{1}{2c}\frac{\mathrm{d}^2 n_g}{\mathrm{d}\lambda^2}\,. \tag{2.27}$$

The first-order material dispersion coefficient M dominates, as long as $M(\lambda_0) \neq 0$ holds. The operating point λ_0 is (inaccurately) called "zero material dispersion wavelength". However, the material dispersion does *not* disappear, because higher order material dispersion comes into play, and the (second order) material dispersion coefficient N takes over.

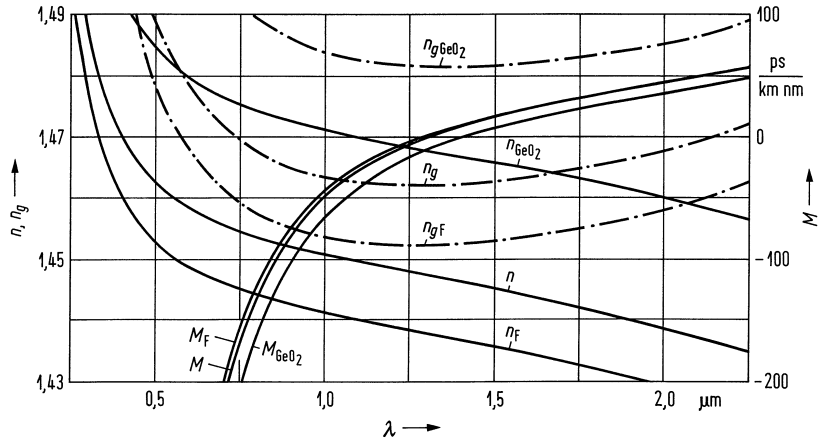


Fig. 2.4. Refractive index n, group index n_g and material dispersion coefficient M, computed using the so-called Sellmeier coefficients for pure SiO_2 and (appropriately subscripted) for SiO_2 with dopings of 2 mol % F, 13.3 mol % GeO_2. Zeros of M are at $\lambda_{0\mathrm{F}} = 1.2649\,\mu$m, $\lambda_0 = 1.2758\,\mu$m, $\lambda_{0\mathrm{GeO_2}} = 1.3722\,\mu$m

In Fig. 2.1, a basic refractive index model of glass with two resonance absorption frequencies $\omega_{1,2}$ was discussed. Figure 2.4 depicts the measured quantities n, n_g, M for pure fused silica and for F- and GeO_2-doped SiO_2 (appropriately subscripted) as a function of λ. The position of the zero of M at λ_0 depends on the doping.

The second order dispersion coefficient N_0 may be read from Fig. 2.4. It does not depend strongly on the doping. For undoped quartz glass we find:

$$N_0 \approx 5.3{\times}10^{-2}\,\mathrm{ps}\,/(\mathrm{km\,nm^2})\,, \qquad 2N_0\lambda_0 \approx 135\,\mathrm{ps}\,/(\mathrm{km\,nm}) \tag{2.28}$$

2.1.5 Plane boundary

Consider two homogeneous, isotropic regions with real refractive indices n_1, n_2 separated by a boundary plane at $x = 0$, Fig. 2.5(a). We look for stationary solutions of Maxwell's equations (2.1), (2.12) for a plane wave incident from the denser medium ($n_1 > n_2$). We try a solution with three superimposed monochromatic plane waves Eq. (2.16),

$$\vec{\Psi}(t,\vec{r}) = \vec{\Psi}\exp(\mathrm{j}\,\omega t - \mathrm{j}\,\vec{k}_s\cdot\vec{r}), \quad \vec{\Psi}_s = \vec{E}_s, \vec{H}_s\,, \quad s = 1,2,3\,.$$

The subscripts 1, 2 and 3 designate the incident, transmitted and reflected waves, respectively. Only the total of *three* waves can solve the problem in general. Vector components are subscripted with their respective coordinates $q = x, y, z$. For the incident wave we assume $k_{1y} = 0$. Fields have either E-polarization ($\vec{E}_s = E_s\vec{e}_y \parallel$ boundary plane, $E_{sz} = 0$, TE- or H-wave), or H-polarization ($\vec{H}_s = H_s\vec{e}_y \parallel$ boundary plane, $H_{sz} = 0$, TM- or E-wave). In Fig. 2.5(a), the $\vec{E}$- and $\vec{H}$-vectors (in parentheses $\vec{H}$- and $\vec{E}$-vectors) are depicted for E-polarization (H-polarization).

We have to impose the boundary condition that the *tangential* components of $\vec{E}_s$ and $\vec{H}_s$ be continuous in $x = 0$. A similar continuity condition applies for the *normal* components of $\vec{D}$ and $\vec{B}$. Therefore, in $x = 0$ the spatial phase dependencies of the three waves are identical, and the amplitudes for E- and H-polarization are continuous,

$$
\begin{aligned}
& k_{1y} = k_{2y} = k_{3y} = 0, && k_{1z} = k_{2z} = k_{3z}\,; \\
& E\text{-polarization:} && H\text{-polarization:} \\
& E_{2y} = E_{1y} + E_{3y}, && H_{2y} = H_{1y} + H_{3y}, \\
& H_{2z} = H_{1z} + H_{3z}, && E_{2z} = E_{1z} + E_{3z}, \\
& |H_{sz}| = H_s \sin\vartheta_s, \quad (H_{3z} < 0), && |E_{sz}| = E_s \sin\vartheta_s, \quad (E_{3z} > 0), \\
& Z_s = E_{sy}/H_{sz} = \tfrac{Z_0}{n_s} k_s/k_{sx}, && Y_s = -H_{sy}/E_{sz} = \tfrac{n_s}{Z_0} k_s/k_{sx}.
\end{aligned}
\tag{2.29}
$$

The z-components of $\vec{k}_s$ and the amplitudes at both sides of $x = 0$ must be identical, so that the fields propagate in synchronism along the z-axis. *Snell's law* and the limiting *angle ϑ_{1T} of total internal reflection* (TIR) follow,

$$
\vartheta_1 = \vartheta_3, \qquad n_1 \cos\vartheta_1 = n_2 \cos\vartheta_2, \qquad \cos\vartheta_{1T} = \frac{n_2}{n_1}. \tag{2.30}
$$

For $n_2 < n_1$ the incident angle is always larger than the angle of the transmitted wave, $\vartheta_1 > \vartheta_2$. At $\vartheta_1 = \vartheta_{1T}$ the transmitted wave propagates along the z-axis and $\vartheta_2 = 0$. From Eq. (2.29) we find *Fresnel's formulae*,

$$
\begin{aligned}
r_E &= \frac{E_3}{E_1} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{k_{1x} - k_{2x}}{k_{1x} + k_{2x}} = \frac{n_1 \sin\vartheta_1 - n_2 \sin\vartheta_2}{n_1 \sin\vartheta_1 + n_2 \sin\vartheta_2}\,, \\
r_H &= \frac{H_3}{H_1} = \frac{Y_2 - Y_1}{Y_2 + Y_1} = \frac{n_2^2 k_{1x} - n_1^2 k_{2x}}{n_2^2 k_{1x} + n_1^2 k_{2x}} = \frac{n_2 \sin\vartheta_1 - n_1 \sin\vartheta_2}{n_2 \sin\vartheta_1 + n_1 \sin\vartheta_2}\,, \\
t_E &= \frac{E_2}{E_1} = 1 + \frac{E_3}{E_1} = 1 + r_E\,, \\
t_H &= \frac{H_2}{H_1} = 1 + \frac{H_3}{H_1} = 1 + r_H\,, \\
t_{H,E} &= \frac{E_2}{E_1} = \frac{n_1}{n_2}\frac{H_2}{H_1} = \frac{n_1}{n_2}\, t_H\,.
\end{aligned}
\tag{2.31}
$$

The quantities Z_1, Z_2 and Y_1, Y_2 are the wave impedances and admittances of the field components in the boundary plane $x = 0$. In Fig. 2.6, the amplitude reflection coefficients r_E, r_H and the amplitude transmission coefficients t_E, t_H with respect to transverse components for E, H-polarization

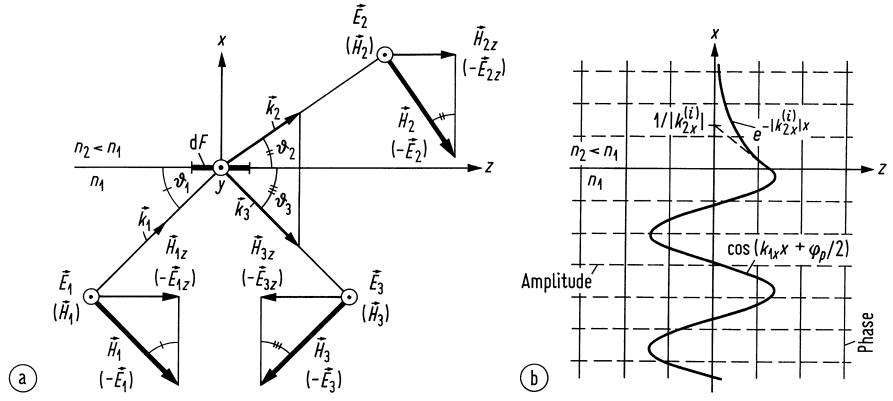


Fig. 2.5. Reflection and refraction of E-polarized (H-polarized) waves at a plane boundary; the subscripts 1, 2 and 3 designate the incident, transmitted and reflected waves, respectively. ⊙-vectors point to the observer (perpendicularly out of the drawing area). (a) Reflection und refraction at the transition to the denser medium (b) Total internal reflection

($k_{3x} = -k_{1x}$, $Z_3 = -Z_1$) are depicted. Further, the electric field amplitude transmission coefficient $t_{H,E}$ (sometimes denoted as $t_\parallel$) for H-polarization is graphed. Other notations sometimes found in books are $r_\perp \mathrel{\widehat{=}} r_E$, $r_\parallel \mathrel{\widehat{=}} r_{H,E} = r_H$ (because $n_1 = n_3$). The reflection coefficient for H-polarization is always smaller than the one for E-polarization, $|r_H| \le |r_E|$, because r_E does not intersect the abscissa, but $r_H = 0$ for the so-called Brewster angle

$$\tan\vartheta_{1B} = n_1/n_2\,, \qquad \tan\vartheta_{2B} = n_2/n_1\,, \qquad \vartheta_{1B} + \vartheta_{2B} = \pi/2\,. \tag{2.32}$$

In the region of total internal reflection $\vartheta_1 \le \vartheta_{1T}$, the propagation constant k_{2z} in z-direction

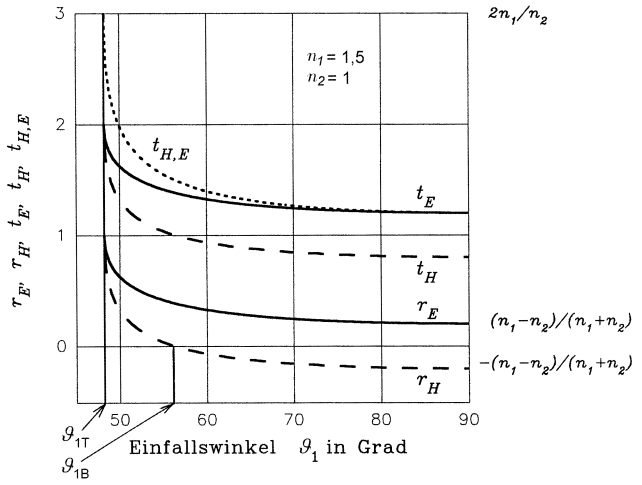


Fig. 2.6. Amplitude reflection and transmission coefficients as a function of incident angle ϑ_1 for $n_1 = 1.5$, $n_2 = 1$ (glass-air interface). Einfallswinkel ϑ_1 in Grad = incident angle ϑ_1 in degree

is larger than k_2 for wave propagation in a medium with index n_2. The propagation constant k_{2x} becomes negative imaginary, and this leads to a spatial contraction of the field components $\Psi_{2q}(t,\vec{r}) = E_{2q}, H_{2q}$,

$$\begin{aligned} k_{2x} &= -\sqrt{k_2^2 - k_{2z}^2} = -\,\mathrm{j}\,k_0\sqrt{n_1^2\cos^2\vartheta_1 - n_2^2} = -\,\mathrm{j}\,|k_{2x}^{(i)}|\,, \\ \Psi_{2q}(t,\vec{r}) &= \Psi_{2q}\,\mathrm{e}^{\,\mathrm{j}\,\omega t}\,\mathrm{e}^{-\,\mathrm{j}\,k_{2x}x}\,\mathrm{e}^{-\,\mathrm{j}\,k_{2z}z} = \Psi_{2q}\,\mathrm{e}^{-|k_{2x}^{(i)}|x}\,\mathrm{e}^{\,\mathrm{j}(\omega t - k_{2z}z)}\,. \end{aligned} \tag{2.33}$$

The field is called evanescent in $+x$-direction. Real and imaginary part of the complex propagation vector $\vec{k}_2 = -\,\mathrm{j}\,|k_{2x}^{(i)}|\,\vec{e}_x + k_{1z}\,\vec{e}_z$ as well as phase and amplitude surfaces are perpendicular to each other, Eq. (2.19) and Fig. 2.5(b). The real part of the Poynting vector Eq. (2.20) points along the z-axis, so there is no energy transport in x-direction, and all power is reflected at the boundary. The reflection coefficients $r_{E,H}$ may be calculated from Eq. (2.31). For E-polarization we find

$$r_E = \frac{k_{1x} - k_{2x}}{k_{1x} + k_{2x}} = \frac{\overbrace{k_{1x} + \mathrm{j}\,|k_{2x}^{(i)}|}^{A\exp\left[\mathrm{j}\arctan\left(|k_{2x}^{(i)}|/k_{1x}\right)\right]}}{\underbrace{k_{1x} - \mathrm{j}\,|k_{2x}^{(i)}|}_{A\exp\left[\mathrm{j}\arctan\left(-|k_{2x}^{(i)}|/k_{1x}\right)\right]}} = \frac{A}{A}\exp\left[\mathrm{j}\,2\arctan\left(|k_{2x}^{(i)}|/k_{1x}\right)\right].$$

The reflection factors $r_p = \exp(\mathrm{j}\,\varphi_p)$, $p = E, H$ have a modulus of one, but the reflected wave experiences a phase shift,

$$\varphi_E = 2\arctan\frac{|k_{2x}^{(i)}|}{k_{1x}}\,, \qquad \varphi_H = 2\arctan\frac{n_1^2|k_{2x}^{(i)}|}{n_2^2 k_{1x}}\,. \tag{2.34}$$

Total internal reflection summary Incident and reflected waves have equal amplitudes, but opposite propagation directions with respect to the x-axis. Both waves superimpose to form a standing wave along the x-axis, Fig. 2.5(b) on Page 17,

$$\Psi_{1q}(t,\vec{r}) = \Psi_{1q}\,\mathrm{e}^{\,\mathrm{j}\,\omega t}\left(\mathrm{e}^{-\,\mathrm{j}\,k_{1x}x} + \mathrm{e}^{+\,\mathrm{j}\,k_{1x}x}\right)\mathrm{e}^{-\,\mathrm{j}\,k_{1z}z} = 2\Psi_{1q}\cos(k_{1x}x)\,\mathrm{e}^{\,\mathrm{j}(\omega t - k_{1z}z)}\,.$$

In the cladding medium n_2 this wave decays as $\exp(-|k_{2x}^{(i)}|x)$, Eq. (2.33) and Fig. 2.5(b). The reciprocal cladding penetration depth is $|k_{2x}^{(i)}| = k_0\sqrt{n_1^2\cos^2\vartheta_1 - n_2^2}$. The larger the refractive index difference becomes, the faster the field decays. The phase shift for $\vartheta_1 \to 0$ approaches π. For small refractive index differences, there is no discrimination between E- and H-polarization, Eq. (2.31), (2.34). This is the scenario of scalar optics.

The power of the incident wave in a differential area $\mathrm{d}\,F$ (see Fig. 2.5(a)) is transferred to the reflected and transmitted waves,

$$\tfrac{1}{2}\,n_2|E_2|^2\;\mathrm{d}\,F\sin\vartheta_2 = \tfrac{1}{2}\,n_1|E_1|^2\;\mathrm{d}\,F\sin\vartheta_1 - \tfrac{1}{2}\,n_1|E_3|^2\;\mathrm{d}\,F\sin\vartheta_1\,.$$

Therefore, the power reflection and transmission factors (in short: reflection and transmission factors) are

$$T_p = 1 - R_p, \qquad R_p = |r_p|^2, \qquad p = E, H. \tag{2.35}$$

With perpendicular incidence for a glass-air interface and $n_1 = 1$, $n_2 = 1,5$ (or vice versa), the reflection factors are $R_E = R_H = 4\,\%$; for a GaAs-air interface with $n_2 = 3.6$, the reflection factors amount to $R_E = R_H = 32\,\%$.

Generalized Fresnel formulae exist for reflection and refraction at arbitrary curved boundary surfaces.[18,19]

2.2 Principles of waveguiding

Dielectric passive light waveguides transport the energy of an optical field in all points of their cross-section area only along the waveguide axis z, and concentrate the field in the vicinity of the axis.

Geometrical optics tells that the refractive index n should be maximum on the waveguide axis, because light rays become bent in the direction of increasing n. A light ray may be represented by a locally plane wave, i. e., a wave having a plane wavefront perpendicularly to the local ray direction in the immediate vicinity of the ray path, and being not defined elsewhere. Waveguide fields may be represented by the superposition of several so-called ray congruences.

A *congruence* is a system of curves which fills a portion of space in such a way that in general a single curve passes through each point of the region. If there exists a family of surfaces which cut each of the curves orthogonally the congruence is said to be normal.

The transverse components of all locally plane waves associated with each congruence compensate each other. The z-components of the associated propagation constants are identical in all points of the cross-section area.

Wave optics requires a guided field to have plane phase fronts perpendicularly to the waveguide axis z. No energy is streaming away from the axis, and no power enters the waveguide. In a homogeneous medium $n = \mathrm{const}$, a boundary field in plane $z = 0$ has the tendency to diverge, e. g., a point source emits spherical waves. However, an appropriate inhomogeneous medium flattens the sphere-like phase fronts by delaying the wave in the vicinity of the waveguide axis where the refractive index is largest.

[18] Snyder, A. W.; Love, J. D.: Reflection at a curved dielectric interface — electromagnetic tunnelling. IEEE Trans. Microwave Theory & Tech. MTT-23 (1975) 134–141

[19] Snyder, A. W.; Love, J. D.: Optical waveguide theory. London: Chapman and Hall 1983. Sect. 7-14

Waveguiding is achieved by surrounding a waveguide core along the z-axis with a lower density cladding medium, i. e., the refractive index decreases away from the axis to form a core-cladding or a graded-index waveguide.

Light waveguides are made from transparent dielectric materials like glass or semiconductors based on Si, GaAs or InP. Symmetric slab waveguides and strip waveguides are particularly important, serving as simple models for the waveguiding regions of *o*pto-*e*lectronic *i*ntegrated *c*ircuits (OEIC) or of semiconductor lasers (laser diodes). With a slab waveguide, all essential light waveguide properties may be studied at little mathematical cost. The physical insight gained from these studies may be transferred to more complex waveguiding structures. — With laser diodes, an anti-waveguiding effect can be important, when the real part n at the z-axis is reduced by increasing the carrier density, see Fig. 2.1 and the discussion on Page 2, thereby pushing away the field from the waveguide axis.

2.3 Slab waveguide

Slab waveguides represent the simplest light waveguide structures. They consist of three homogeneous, isotropic layers, which are infinitely extended along the y-axis, Fig. 2.7. The core area with height h and refractive index n_1 is surrounded by the infinitely extended cladding medium with refractive index $n_2 < n_1$. At $x = z = 0$ we assume a monochromatic line source Q, infinitely extended along the y-axis. Source Q emits both evanescent and homogeneous plane waves. We look for fields, which *propagate* in the semi-space $z > 0$. These waves will zigzag between the reflecting boundary surfaces $x = \pm h/2$, and they will be partially transmitted into the cladding medium (radiation modes). In Fig. 2.7, some typical curves of four normal congruences (see Page 19) are depicted representing the propagation directions of plane waves. There is one congruence each for the waves running into the direction of the boundaries $x = \pm h/2$, and one congruence each for the waves transmitted at $x = \pm h/2$ into the cladding. The small dashes indicate the phase planes.

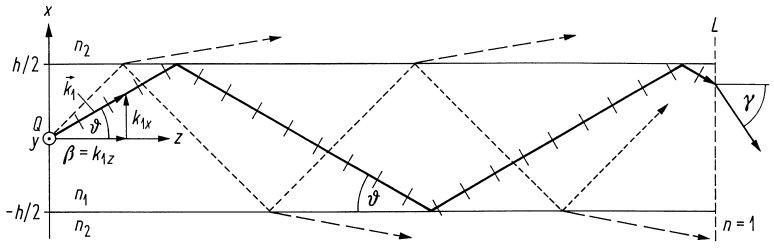


Fig. 2.7. Slab waveguide. Guided (——) and radiated waves (– – –)

If the angle ϑ between the propagation vector $\vec{k}_1$ of the incident plane wave and the z-axis remains smaller than the angle ϑ_T of total internal reflection, i. e., if $\vartheta < \vartheta_T = \arccos(n_2/n_1)$ according to Eq. (2.30), the wave is confined to the core area and does not lose energy to the cladding medium. An initial plane wave at $z = 0$ propagating into the direction of the upper boundary becomes then totally reflected at $x = +h/2$; it acquires a phase shift $\varphi_{E,H}$ according to Eq. (2.34), which depends on the type of polarization. The reflected, downwards propagating wave is totally reflected at the lower boundary $x = -h/2$, where again a phase $\varphi_{E,H}$ is added. After two reflection, the then upward propagating wave superimposes the initial wave. If there is a phase difference other than integer multiples m of 2π, the initial wave and the reflected waves do not interfere constructively, and therefore they will die out after a certain propagation length in the order of a few λ. There will be no energy loss into the cladding medium, but the wave does not propagate along the z-direction and is called *evanescent*. Therefore it will be *back-reflected* from the waveguide entrance.

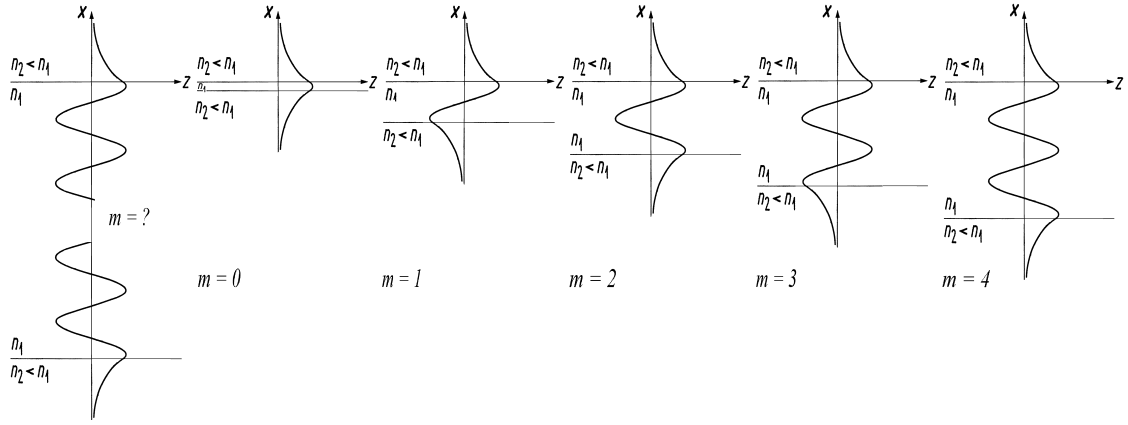


Fig. 2.8. Modes in a slab waveguide according to the results of Fig. 2.5(b). The upper boundary (UB) remains spatially fixed. The lower boundary (LB) is shifted to a distance $x_{\mathrm{UB}} - x_{\mathrm{LB}} = h$ such that the standing waves originating from the reflections at the UB and LB are in phase. From left to right: UB and LB at arbitrary distance, no phase match (m =?). LB at minimum distance (fundamental mode $m = 0$). LB moved to the next possible phase match positions ($m = 1, 2, 3, 4, \ldots$)

2.3.1 Eigenvalues in pictures

Figure 2.8 shows an *u*pper *b*oundary surface (UB) with the transverse field amplitude $\cos(k_{1x}x + \varphi_{E,H}/2)$ extracted from Fig. 2.5(b). Figure 2.8 is further supplemented with a mirror image version of Fig. 2.5(b) as a *l*ower *b*oundary surface (LB) together with its associated field dependence (leftmost picture, m =?). For the next graphs designated with $m = 0, 1, 2, 3, 4$, the UB remains fixed in position, while the LB is shifted to a distance $x_{\mathrm{UB}} - x_{\mathrm{LB}} = h$ such that the standing waves originating from the reflections at the UB and LB are in phase. If the LB is at its minimum distance, the field is called a fundamental mode $m = 0$. For the examples further to the right in Fig. 2.8, the LB is moved to the next possible phase match positions so that symmetric ($m = 0, 2, 4, \ldots$) and antisymmetric modes result (with a sign reversal of the field amplitude at the LB, $m = 1, 3, \ldots$). Obviously, the mode index m measures the number $m + 1$ of intensity maxima inside the core.

The modes constructed in Fig. 2.8 are called *guided modes* of the slab waveguide. They propagate without losses along the z-axis. Eventually, when the waveguide end is reached at $z = L \gg \lambda/n_1$, the trapped light rays with $\vartheta < \vartheta_T$ are radiated into the adjacent vacuum at a maximum angle γ_N, see Fig. 2.7. The angle γ_N defines the far-field light-shadow boundary of geometrical optics. Because γ_N may be measured outside the light waveguide, it is preferred to ϑ_T for characterizing the waveguiding properties. With Eq. (2.30), we have $n_1 \sin\vartheta = \sin\gamma$, and therefore

$$\sin\gamma_N = n_1 \sin\vartheta_T = \sqrt{n_1^2 - n_2^2}, \tag{2.36}$$

$$A_N = \sqrt{n_1^2 - n_2^2} = n_1\sqrt{2\Delta}\,, \quad \Delta = \frac{n_1^2 - n_2^2}{2n_1^2} \approx \{n_1 \approx n_2\} \approx \frac{n_1 - n_2}{n_1}$$

Two new symbols were introduced in Eq. (2.36), namely the numerical aperture A_N and the relative refractive index difference Δ. If Δ is large enough, the numerical aperture may become $A_N > 1$. For some rays then, we have total internal reflection at the waveguide end $z = L$, and $A_N = \sqrt{n_1^2 - n_2^2}$ cannot longer be associated with the maximum radiation angle $\sin\gamma_N \leq 1$.

2.3.2 Eigenvalue equation

For an analytical treatment, we use the results of Sect. 2.1.5 for $\vartheta < \vartheta_T$:

1. $\vec{\Psi}_s(\vec{r}) = \vec{\Psi}_s \exp(-\mathrm{j}\,\vec{k}_s \cdot \vec{r}\,)$ with $s = 1, 2$ in core and cladding for the electric or magnetic field vectors (components subscripted in the first place with the appropriate coordinates $q = x, y, z$, in the second place with s)

2. Polarization index $p = E, H$ marks TE- and TM-waves, respectively (E-polarized, H-wave and H-polarized, E-wave)

3. The propagation vectors of the superimposed plane waves are $\vec{k}_s = k_{sx}\vec{e}_x + k_{sz}\vec{e}_z$.

4. $\beta = k_{1z} = k_{2z}$ is the x-independent phase constant for the propagation of fields along the z-axis.

5. For the case of total internal reflection, the propagation vector components are real inside the core ($k_{1x} = \sqrt{k_1^2 - \beta^2}$, $\beta < k_1$), and imaginary in the cladding, i. e., outside the core ($k_{2x} = \pm \mathrm{j}\sqrt{\beta^2 - k_2^2}$, $\beta > k_2$).

6. n_e is an effective refractive index. A homogeneous plane wave in a homogeneous medium of refractive index n_e would propagate with the same phase velocity v as any guided wave with phase constant β.

In summary, we find:

$$\begin{aligned} &\beta = k_{1z} = k_1 \cos\vartheta = k_{2z} \quad (k_2 < \beta < k_1), \qquad n_e = \beta/k_0 \quad (n_2 < n_e < n_1), \\ &\qquad k_{1x} = k_1 \sin\vartheta = \sqrt{k_1^2 - \beta^2}, \qquad\qquad k_{2x} = \pm\mathrm{j}\sqrt{\beta^2 - k_2^2}. \end{aligned} \tag{2.37}$$

For simplifying the notation, we introduce the transverse core phase constant (core parameter "transversales Phasenmaß") u, the transverse cladding attenuation (cladding parameter "transversales Dämpfungsmaß") w, the normalized frequency (waveguide parameter) V, the relative refractive index difference Δ (see Eq. (2.36)), and the normalized propagation constants B, δ,

$$\begin{aligned} &u = k_{1x}\tfrac{h}{2} = \tfrac{h}{2}\sqrt{k_1^2 - \beta^2}, \qquad w = |k_{2x}|\tfrac{h}{2} = \tfrac{h}{2}\sqrt{\beta^2 - k_2^2}\,, \\ &V = \tfrac{h}{2}k_0 A_N = \sqrt{u^2 + w^2}, \qquad \Delta = \tfrac{n_1^2 - n_2^2}{2n_1^2} \approx \{n_1 \approx n_2\} \approx \tfrac{n_1 - n_2}{n_1}\,, \\ &B = \tfrac{\beta^2 - k_2^2}{k_1^2 - k_2^2} = \tfrac{w^2}{V^2} = 1 - \tfrac{u^2}{V^2} = 1 - \tfrac{\delta}{\Delta} \approx \{\Delta \ll 1\} \approx \tfrac{\beta - k_2}{k_1 - k_2}\,. \end{aligned} \tag{2.38}$$

As explained in Fig. 2.8, not all β are allowed for guided waves $k_2 < \beta < k_1$. For a transversely consistent field at the boundary planes $x = \pm h/2$, the standing waves

$$\cos(k_{1x}x) = \cos\left(\frac{2u}{h}\,x\right), \qquad \sin(k_{1x}x) = \sin\left(\frac{2u}{h}\,x\right)$$

originating from both interfaces have to match. Taking the phase shifts φ_p of Eq. (2.34) into account, the transverse phase difference after two total internal reflections must fulfill the resonance condition

$$-2k_{1x}h + 2\varphi_p = -2m\pi\,, \qquad -u + \varphi_p/2 = -m\pi/2\,. \tag{2.39}$$

With the explicit form of Eq. (2.34), this may be written as

$$\begin{aligned} \varphi_p &= 2\arctan\left(\sigma_p \frac{|k_{2x}^{(i)}|}{k_{1x}}\right), \\ u &= \frac{\varphi_p}{2} + m\frac{\pi}{2} = \arctan\left(\sigma_p \frac{w}{u}\right) + m\frac{\pi}{2}\,, \\ \tan\left(u - m\tfrac{\pi}{2}\right) &= \sigma_p \frac{w}{u}\,. \end{aligned} \tag{2.40}$$

For even and odd TE- and TM-waves, we find the result:

$$\begin{aligned} &\sigma_p w_{pm} = \sigma_p\sqrt{V^2 - u_{pm}^2} = \begin{cases} u_{pm}\tan u_{pm} & m = 0, 2, 4, \ldots \\ -u_{pm}\cot u_{pm} & m = 1, 3, 5, \ldots \end{cases} \\ &\qquad \text{TE-wave (H-wave):} \quad p = E, \qquad \sigma_E = 1 \\ &\qquad \text{TM-wave (E-wave):} \quad p = H, \qquad \sigma_H = n_1^2/n_2^2 \end{aligned} \tag{2.41}$$

The characteristic equation (2.41) (eigenvalue or dispersion equation, mathematical formulation of the graphical procedure in Fig. 2.8) defines allowed angles ϑ_m or so-called eigenvalues β_m. The waves $\vec{\Psi}_m(\vec{r}) = \vec{\Psi}_m(x)\exp(-\mathrm{j}\,\beta_m z)$ are eigenwaves or modes of the waveguide. Fields with other propagation constants cannot propagate, and are either evanescent along the z-direction, or radiate into the cladding.

In Fig. 2.9(a), the graphical representation of the functions to the left and to the right of the curly brace in Eq. (2.41) is depicted. For TE-modes, two circles with radii V, V' for two different operating points are sketched, while TM-modes are characterized by ellipses with a smaller half-axis V and a larger half-axis $\sigma_H V$. From the intersection points with the trigonometric functions $u\tan u$ and $-u\cot u$, the eigenvalues u_{pm} of Eq. (2.41) (or the propagation constants β_{pm}) for TE- and TM-waves are found, where $u_{Em} \le u_{Hm}$, $w_{Em} \ge w_{Hm}$, $\beta_{Em} \ge \beta_{Hm}$ is valid. This inequality can be also seen from Fig. 2.6: For the amplitude reflection coefficients the relation $r_E > r_H$ holds, therefore a TE-wave is always more confined to the core region than a TM-wave. As a consequence, the TE propagation constant must be closer to k_1 than in the TM case.

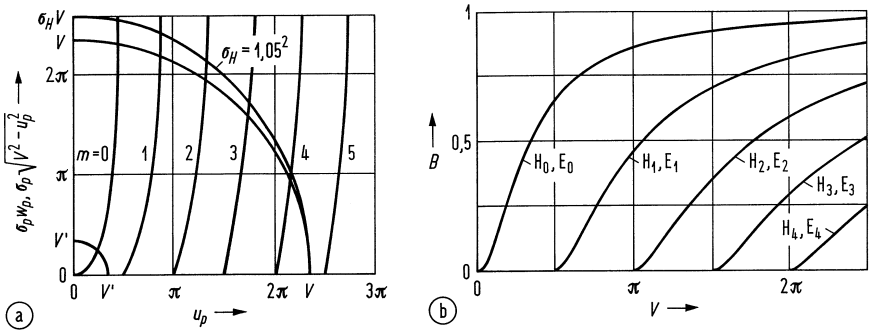


Fig. 2.9. Eigenvalues of the propagation constant β or the core parameter u. (a) Graphical solution of the eigenvalue equation for TE-modes (H-waves, $\sigma_E = 1$) and TM-modes (E-waves $\sigma_H = 1.05^2$), $V = 7.33$. $u_E \le u_H$, $\beta_E \ge \beta_H$, $B_E \ge B_H$ (b) Normalized propagation constant for weakly guiding LWG

The guided eigenmodes (or simply modes) are classified as TE_m- or H_m-modes (only H-component in z-direction) and as TM_m- or E_m-modes (only E-components in z-direction). The smaller V is, the less modes are guided. The fundamental modes H_0 (y-polarized) and E_0 (mainly x-polarized) can always propagate, no matter how low the frequency is or how small the waveguide height might be. Waveguides carrying only the fundamental mode are called singlemoded in contrast to multimode waveguides. The fundamental modes are essentially orthogonally polarized and have the largest possible β, B (the smallest possible u) as compared to modes with higher index m.

For weak guidance $\Delta \ll 1$, the polarization parameter is nearly the same for both polarizations, $\sigma_H \approx 1$. Therefore, the β-values for H- and E-waves approach each other asymptotically, and both modes propagate with virtually the same β. This is assumed for Fig. (2.9)b which shows the numerical solutions of Eq. (2.41) in form of a so-called dispersion diagram $B = B(V)$.

With the solution of the dispersion equation (2.41) we could immediately write down the field solutions. From Fig. 2.5(b) we know that the fields inside the core oscillate according to $\sin(2ux/h)$ and $\cos(2ux/h)$, while we have an exponential decay $\exp(\mp 2wx/h)$ in the cladding. However, in the following the wave equation (2.13) will be solved directly leading to the field solutions.

2.3.3 Vector solution

For the slab waveguide, the wave equation (2.12) may be simplified by assuming monochromatic waves (Eq. (2.13)) propagating as $\exp[\,\mathrm{j}\,(\omega t - \beta z)]$, and by observing $\partial/\partial y = 0$,

$$\nabla^2\Psi(t,\vec{r}) = \frac{n^2}{c^2}\frac{\partial^2}{\partial t^2}\Psi(t,\vec{r}) \Longrightarrow \left\{\begin{array}{l}\text{monochromatic waves}\\ \Psi(t,\vec{r}) = \Psi(\vec{r})\,\mathrm{e}^{\mathrm{j}(\omega t-\beta z)}\end{array}\right\} \tag{2.42}$$

$$\Big(\frac{\partial^2}{\partial x^2} - \beta^2\Big)\Psi(\vec{r}) = -n^2k_0^2\Psi(\vec{r}), \quad k_0 = \frac{\omega}{c}\,, \quad k_{1,2} = n_{1,2}k_0$$

For TE-waves, only the components E_y, H_x and H_z do not vanish, and for TM-waves it is H_y, E_x and E_z. Further, the refractive index n is constant inside the core ($n = n_1$) as well as in the cladding region ($n = n_2$). If we concentrate on TE-waves, we have to solve the simultaneous equations (subscripts $1, 2$; primes on function symbols mean derivatives with respect to the argument)

$$E_y''(x) \pm \Big(\frac{2v}{h}\Big)^2 E_y(x) = 0, \quad v = \left\{\begin{array}{ll} u & \text{core, upper sign,}\\ w & \text{cladding, lower sign.}\end{array}\right. \tag{2.43}$$

We look for solutions which oscillate inside the core, and which decay in the cladding exponentially which is needed for maintaining the field concentration, i. e., for the guiding of the wave. Such functions are well known,

$$E_y\big(|x| \le \tfrac{h}{2}\big) = E_{1y}\left\{\begin{array}{ll}\cos\big(\frac{2u}{h}x\big) & m \text{ even}\\ \text{or} & \\ \sin\big(\frac{2u}{h}x\big) & m \text{ odd}\end{array}\right. \tag{2.44}$$

$$E_y\big(|x| \ge \tfrac{h}{2}\big) = E_{2y}\left\{\begin{array}{ll}\exp\big(-\frac{2w}{h}|x|\big) & m \text{ even}\\ \text{or} & \\ \exp\big(-\frac{2w}{h}|x|\big)\,\mathrm{sgn}(x) & m \text{ odd}\end{array}\right. \tag{2.45}$$

The four unknown coefficients E_{1y}, E_{2y}, u, w have to be determined from the continuity conditions Eq. (2.29) for the transverse E- and H-fields E_y and H_z at the boundaries $x = \pm h/2$ observing that $|r_E| = 1$ and therefore $E_{3y} = E_{1y}$. For the tangential E-components of even modes we find the amplitude boundary condition at $x = +h/2$:

$$E_{1y}\cos\left(\frac{2u}{h}\frac{h}{2}\right) = E_{2y}\exp\left(-\frac{2w}{h}\frac{h}{2}\right) \quad\Longrightarrow\quad E_{2y} = E_{1y}\cos\left(\frac{2u}{h}\frac{h}{2}\right)\exp\left(\frac{2w}{h}\frac{h}{2}\right)$$

From Eq. (2.44) the field structure in the cladding results:

$$E_y\big(|x| \ge \tfrac{h}{2}\big) = E_{1y}\cos\left(\frac{2u}{h}\frac{h}{2}\right)\exp\left[-\frac{2w}{h}\left(|x| - \frac{h}{2}\right)\right] \tag{2.46}$$

From Maxwell's equations (2.1) we see that

$$\mathrm{curl}\,\vec{E} = -\mu_0\frac{\partial\vec{H}}{\partial t} \quad\Longrightarrow\quad H_z(x) = \mathrm{j}\,\frac{1}{\mu_0\omega}\,\frac{\partial E_y(x)}{\partial x}\,.$$

Applying again the continuity condition, this time for the transverse magnetic field of even modes at $x = +h/2$, results in:

$$H_z\left(x = \frac{h}{2}\right) = \underbrace{\mathrm{j}\,\frac{E_{1y}}{\mu_0\omega}\left(-\frac{2u}{h}\right)\sin\left(\frac{2u}{h}\frac{h}{2}\right)}_{\text{core Eq. (2.44)}} = \overbrace{\mathrm{j}\,\frac{E_{1y}}{\mu_0\omega}\left(-\frac{2w}{h}\right)\cos\left(\frac{2u}{h}\frac{h}{2}\right)}^{\text{cladding Eq. (2.46)}} \tag{2.47}$$

$$\Longrightarrow \quad w = u\tan u \quad \text{for } m \text{ even.}$$

Equation (2.47) corresponds to the first line of the eigenvalue equation (2.41). Performing similarly for odd TE-modes, and repeating the procedure for TM-waves, we would reproduce the dispersion equation (2.41) in full. With these results, the vector field solution for the slab waveguide problem is completed. A single wave function $\Psi_m(x) = E_{ym}(x), H_{xm}(x)$ (for TE-modes) and $\Psi_m(x) = E_{xm}(x), H_{ym}(x)$ (for TM-modes) represents the transverse components of $\vec{E}$, $\vec{H}$.

2.3.4 Scalar solution

There are classes of problems where the vector nature of fields is not relevant so that scalar optics may be employed:

1. For reflection and refraction at a plane boundary there is no difference between E- and H-polarization, if the relative refractive index difference is small, $\Delta \ll 1$, see text after Eq. (2.34) on Page 2.1.5.

2. Diffraction is independent of the polarization of the fields, if the diffracting objects are much larger than λ/n, and the fields in distances of the order λ/n away from the diffracting objects is not of interest.

3. The electromagnetic field which is associated with the passage of natural light through an optical instrument of moderate aperture and of conventional design is such that the intensity may be approximately represented in terms of a single complex scalar wave function.

Usually, the waveguides of interest are weakly guiding, so the z-components of the fields may be neglected, and the differences in TE- and TM-waves, i. e., the vector properties of the fields, disappear. This means that the relative n and $|\operatorname{grad} n|$ variations along a distance λ/n are small, $|\Delta n|_\lambda/n \ll 1$ and $|\Delta(\operatorname{grad} n)|_\lambda/|\operatorname{grad} n| \ll 1$. Terms with $\operatorname{grad} \ln n^2$ in Eq. (2.10) may then be neglected, and in Cartesian coordinates the differential equations are decoupled as if it was a homogeneous medium,

$$\nabla^2 \vec{E} - \frac{n^2(\vec{r})}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} = 0, \qquad \nabla^2 \vec{H} - \frac{n^2(\vec{r})}{c^2} \frac{\partial^2 \vec{H}}{\partial t^2} = 0. \tag{2.48}$$

For a (piecewise) homogeneous medium the wave equation may be solved, while for an inhomogeneous medium no general solution exists. As in Eq. (2.12), the scalar function $\Psi(t, \vec{r})$ represents the Cartesian field components, $\Psi(t, \vec{r}) = E_q(t, x, y, z), H_q(t, x, y, z)$, $q = x, y, z$, so $\Psi(\vec{r})$ is a solution of the scalar Helmholtz equation,

$$\begin{aligned} &\Psi(t, \vec{r}) = \Psi(\vec{r})\, \mathrm{e}^{\mathrm{j}\,\omega t}, && \mathcal{S}_{\mathcal{T}}(\vec{r}) = I(\vec{r}) = \tfrac{1}{2} n(\vec{r})\, |\Psi(\vec{r})|^2 , \\ &\left(\nabla^2 + k_0^2 n^2(\vec{r})\right) \Psi(\vec{r}) = 0 , && \tfrac{|\Delta n|_\lambda}{n}, \tfrac{|\Delta(\operatorname{grad} n)|_\lambda}{|\operatorname{grad} n|} \ll 1. \end{aligned} \tag{2.49}$$

The modulus $\mathcal{S}$ of the Poynting vector $\vec{\mathcal{S}} = \vec{E} \times \vec{H}$, averaged over an optical period (subscript $\mathcal{T}$), defines the intensity $I(\vec{r}) = \mathcal{S}_{\mathcal{T}}(\vec{r}) \sim |\Psi(\vec{r})|^2$. Each spectral component of the field is represented by a complex, intensity-normalized wave function $\Psi(t, \vec{r}) = \Psi(\vec{r}) \exp(\mathrm{j}\,\omega t)$ with complex amplitude $\Psi(\vec{r})$. Light interference effects may be described by a superposition of complex scalar fields.

In scalar optics , both TE- and TM-waves are *l*inearly *p*olarized all over the cross-section of the waveguide, and are therefore denominated as LP-modes. These approximate fields behave like *t*ransverse *e*lectro*m*agnetic TEM$_m$-waves. They may be described by a single wave function $\Psi_m(x) = E_{ym}(x)$, $H_{xm}(x)$, $E_{xm}(x)$, $H_{ym}(x)$ both for TE- and TM-modes. Naturally, the polar-

ization index p becomes meaningless now. In scalar approximation we write:

$$\Psi_m(x) = \Psi_{N\,m} \begin{cases} \left\{ \begin{matrix} \cos u_m & m \text{ even} \\ \sin u_m & m \text{ odd} \end{matrix} \right\} \mathrm{e}^{-\frac{2w_m}{h}(x-\frac{h}{2})} & \frac{h}{2} \le x < +\infty \\ \left\{ \begin{matrix} \cos(\frac{2u_m}{h}x) & m \text{ even} \\ \sin(\frac{2u_m}{h}x) & m \text{ odd} \end{matrix} \right\} & -\frac{h}{2} \le x \le \frac{h}{2} \\ \left\{ \begin{matrix} \cos u_m & m \text{ even} \\ -\sin u_m & m \text{ odd} \end{matrix} \right\} \mathrm{e}^{+\frac{2w_m}{h}(x+\frac{h}{2})} & -\infty < x \le -\frac{h}{2} \end{cases} \tag{2.50}$$

$$n_1 \int_{-\infty}^{+\infty} \Psi_m(x)\, \Psi^*_{m'}(x)\, \mathrm{d}x = \delta_{m\,m'}\,, \qquad \delta_{m\,m'} = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases} \tag{2.51}$$

The total cross-sectional power determines the normalizing modal constants $\Psi_{N\,m}$ of the orthogonality relation Eq. (2.51). The modal "intensity" (unit m^{-1} in this one-dimensional case) is $I_m(x) = \frac{1}{2} n_1 |\Psi_m(x)|^2$ ($n_1 \approx n_2$) according to equation (2.49) for scalar fields.

In Fig. 2.8 the lowest-order unnormalized ($\Psi_{N\,m} = 1$) wave functions of even and odd type are depicted according to Eq. (2.50).

2.3.5 Modal cutoff

Equation (2.37) gives the β-range of guided modes. The waveguiding limit is reached when the field extends into the whole cladding region $|x| < \infty$. Then, the cladding parameter is $w_p = 0$, $\beta = k_2$. For each mode m, we define a normalized cutoff frequency V_{mG} (cutoff wavelength λ_{mG}) for which $w_p = 0$. The total number of guided modes m with normalized cutoff frequencies $V_{mG} < V$ in two orthogonal polarizations is denoted by M_g. From Eq. (2.41) we find for $w_p = 0$ that $\tan u = 0$ and $\cot u = 0$ independently of the polarization p, i. e., $u_{mG} = V_{mG} = m\pi/2$ (see Fig. 2.9(a)). With Eq. (2.38) we have

$$V_{mG} = m\frac{\pi}{2}, \quad \lambda_{mG} = 2hA_N/m, \quad M_g = 2m(V_{mG}) = 2\,\frac{2}{\pi} V_{mG} \approx \Big\{ V \gg 1 \Big\} \approx \frac{4}{\pi} V. \tag{2.52}$$

The fundamental mode $m = 0$ has a cutoff frequency of zero (cutoff wavelength infinity). For unsymmetrical slab waveguides with different n for $x \le -h/2$ and $x \ge h/2$ the cutoff wavelength of the fundamental mode is finite. At the cutoff frequency $V = V_{mG} = u_{mG}$ we find with the help of Eqs. (2.52), (2.50) the relation $|\Psi_m(\pm h/2)/\Psi_{N\,m}| = 1$. For the limit $V, w_{pm} \to \infty$, Eqs. (2.41), (2.50) yield $\Psi_m(\pm h/2) = 0$.

In scalar approximation and for single-mode waveguides with $V \le V_{1G}$ the significant near-field extension x_M is at least the waveguide height h plus two times the penetration depth $(2w_m/h)^{-1}$ of the cladding field, i. e.,

$$u = V_{1G} = \frac{\pi}{2}\,, \quad w = \underbrace{\sqrt{V_{1G}^2 - u^2} = u \tan u}_{\substack{\text{solution:} \\ u \approx 0.934}} \quad \Longrightarrow \quad w \approx 1.26 \approx 1/0.8\,, \tag{2.53}$$

$$x_M \ge h(1 + 1/w) \approx 1.8\, h\,.$$

For $V < V_{1G}$, the cladding parameter w becomes smaller, i. e., the field penetrates more into the cladding, and x_M becomes larger than $1.8\, h$.

2.3.6 Group delay dispersion and transmission speed

We assume a weakly guiding slab so that the scalar approximation is valid. Following Eq. (2.21) ($t_g/L = \mathrm{d}k/\,\mathrm{d}\omega = n_g/c$) and Eq. (2.38) ($B \approx (\beta - k_2)/(k_1 - k_2)$, approximation for weak guidance),

the group delay time t_g of a mode after propagation through a waveguide length L is

$$\frac{t_g}{L} = \frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \frac{1}{c}\frac{\mathrm{d}\beta}{\mathrm{d}k_0} \approx \left\{ \begin{array}{c} \Delta \ll 1 \\ \frac{\mathrm{d}V}{\mathrm{d}k_0} \approx \frac{V}{k_0} \\ n_{1g} - n_{2g} \\ \approx n_1 - n_2 \end{array} \right\} \approx \underbrace{\frac{n_{2g}}{c}}_{\text{material}} + \underbrace{\frac{n_{1g}-n_{2g}}{c} \overbrace{\frac{\mathrm{d}(VB)}{\mathrm{d}V}}^{\substack{\text{group delay} \\ \text{factor}}}}_{\text{waveguide}} . \tag{2.54}$$

The first term describes the mode-independent group delay of a plane wave in a dispersive homogeneous material, the second term specifies the group delay because of mode-dependent waveguiding effects (the frequency dependence of the small difference $n_{1g} - n_{2g}$ is not consequential). The factor $\mathrm{d}(VB)/\,\mathrm{d}V$ is called group delay factor. At the cutoff frequency $V = V_{mG}$, the mode m fills the infinite cladding cross-section, therefore $t_g/L = n_{2g}/c$ and $\mathrm{d}(VB)/\,\mathrm{d}V = 0$. On the other hand, if the frequency becomes very high, $V \to \infty$, the field is totally concentrated to the core region, and $t_g/L = n_{1g}/c$, $\mathrm{d}(VB)/\,\mathrm{d}V = 1$.

Analogue to Eq. (2.27) we find the length-related group delay time difference $\Delta t_g/L$ of two signals λ and λ_1 propagating in the same mode m at optical carriers which differ in wavelength by $\Delta\lambda = \lambda - \lambda_1$, using Eq. (2.26), (2.54) and the approximations specified therein:

$$\begin{aligned} &\Delta t_g/L = \left[t_g(\lambda + \Delta\lambda, m) - t_g(\lambda, m)\right]/L = C\Delta\lambda = (M + W)\Delta\lambda\,, \\ &M \quad = M_s = \underbrace{\frac{1}{c}\frac{\mathrm{d}n_{sg}}{\mathrm{d}\lambda}}_{\substack{\text{material} \\ \text{dispersion}}} \quad (s = 1 \text{ or } 2), \qquad W = -\frac{n_{1g} - n_{2g}}{c\lambda}\, \underbrace{V\frac{\mathrm{d}^2(VB)}{\mathrm{d}V^2}}_{\substack{\text{dispersion} \\ \text{factor}}} \end{aligned} \tag{2.55}$$

Additionally, the first-order material dispersion coefficients in core (M_1) and cladding ($M_2 \approx M_1$) are assumed to be of similar value. The chromatic dispersion is expressed by the first-order coefficient C (unit ps / (km nm)) for a fixed reference wavelength λ_1. In analogy to the second-order material dispersion coefficient N in Eqs. (2.26), (2.27) we extend Eq. (2.55) by one more term and define a second-order dispersion coefficient D (unit ps / $\left(\mathrm{km\,nm^2}\right)$),

$$\begin{aligned} &\Delta t_g/L = C\,\Delta\lambda + D\,(\Delta\lambda)^2 + \ldots, \qquad C = M + W, \quad C = \frac{1}{L}\frac{\mathrm{d}\,t_{gm}}{\mathrm{d}\lambda}\,, \quad D = \frac{1}{L}\frac{1}{2}\frac{\mathrm{d}^2 t_{gm}}{\mathrm{d}\lambda^2}\,, \\ &C(\lambda_C) = 0\,, \qquad \lambda_C \text{ is the zero first-order dispersion wavelength for mode } m. \end{aligned} \tag{2.56a}$$

When for a certain reference wavelength $\lambda_1 = \lambda_C$ the first-order chromatic dispersion C becomes zero, the total dispersion is determined by the second-order dispersion coefficient D.

With a slight re-ordering of Eq. (2.56a) we can define a wavelength-dependent chromatic dispersion factor $C_\lambda(\lambda) = C + D\,\Delta\lambda$ which is approximated by a straight line near the reference wavelength λ_1,

$$\Delta t_g/L = C_\lambda(\lambda)\,\Delta\lambda = (C + D\,\Delta\lambda)\;\Delta\lambda = C\,\Delta\lambda + D\,(\Delta\lambda)^2\,, \quad D = \frac{\mathrm{d}C_\lambda(\lambda)}{\mathrm{d}\lambda}\,. \tag{2.56b}$$

Comparing D in Eq. (2.56a), (2.56b) could lead to confusion because of the factor 1/2. In Eq. (2.56a), the dispersion *coefficients* $C \equiv C(\lambda_1)$, $D \equiv D(\lambda_1)$ are constants of the Taylor expansion for the function $t_{gm}(\lambda)$ at a certain reference wavelength $\lambda = \lambda_1$. Therefore, $\mathrm{d}C(\lambda_1)/\,\mathrm{d}\lambda = 0$ holds by definition, and $D(\lambda_1) \neq \mathrm{d}C(\lambda_1)/\,\mathrm{d}\lambda$. On the other hand, the dispersion *function* $C_\lambda(\lambda)$ may be linearly expanded, and its so-called dispersion slope $D = \mathrm{d}C_\lambda(\lambda)/\,\mathrm{d}\lambda|_{\lambda_1}$ is then well defined.

Basically, three types of dispersion in light waveguides have to be discriminated, see Fig. 1.4 and 1.5 on Pages 3 and 4:

Intramodal dispersion Group delay dispersion (or *g*roup *v*elocity *d*ispersion, GVD) for a fixed mode m is described by Eq. (2.56a). Parameters are the chromatic dispersion coefficient $C = M + W$, the material dispersion coefficient M, the waveguide dispersion coefficient W, the dispersion factor $V\,\mathrm{d}^2(VB)/\,\mathrm{d}V^2$, and the dispersion slope D.

Intermodal dispersion (or simply modal dispersion) From Fig. 2.9(b) we see that for a fixed V the slopes of $B(V) \mathrel{\widehat{=}} \beta(\omega)$ which correspond to the group delay vary between different modes. This is to be expected when inspecting the congruences (zigzag rays) in Fig. 2.7: Different modes correspond to differently inclined rays. Between planes $z = 0$ and $z = L$, the geometrical *and* the optical ray path lengths differ. A short impulse at $z = 0$ generates a series of separated impulses at $z = L$ corresponding in number to the excited modes. There is a spread of the group delays *between different* modes m and $m + \Delta m$,

$$\Delta t_g / L = \left[t_g(\lambda, m + \Delta m) - t_g(\lambda, m) \right] / L = G \Delta m \,. \tag{2.57}$$

The quantity G is called the modal dispersion coefficient. For weak guidance $\Delta \ll 1$, the maximum spread of allowed angles for guided modes is diminished according to Eq. (2.36), and so is the modal dispersion.

Polarization mode dispersion (PMD) As discussed on Page 23, orthogonally polarized modes have slightly different propagation constants and hence differ also in group delay, see Fig. 2.9(a).

The effect of dispersion is to increase the width of an input impulse by Δt_g at the waveguide output, Fig. 1.4(a) on Page 3. If $C = 50\,\mathrm{ns}\,/\,(\mathrm{km\ nm})$ is specified, a 150 ns wide input impulse emitted by a source with a spectral width of $\Delta\lambda = 1\,\mathrm{nm}$ would be broadened to a width of 200 ns after a hypothetical propagation length of $L = 1\,\mathrm{km}$.

PMD becomes important for integrated-optics waveguides with large Δ, or for optical fibres where short light impulses travel over long distances.

Transmission speed Group delay spreading becomes a serious restriction on transmission speed, because neighbouring impulses that overlap interfere with each other (Fig. 1.5 on Page 4), making it impossible to decide about a logical 1 or a logical 0. Consequently, impulses have to be separated by more than Δt_g. One rough guideline for estimating the maximum bit rate is that the interval between impulses should be four times the dispersion. The achievable bit rate-length product (clock frequency f_t) is

$$f_t L = \frac{1}{4 \Delta t_g / L} \,. \tag{2.58}$$

So far we have considered the effects of dispersion on (hypothetical) impulses of zero width. As is well known from the Fourier transform relation Eq. (2.2) on Page 7, a rectangular time function $\Psi(t)$ which exists only in an interval $-\Delta t_H/2 \le t < \Delta t_H/2$ where it is Ψ_0 has a Fourier spectrum

$$\breve{\Psi}(f) = \int_{-\Delta t_H/2}^{+\Delta t_H/2} \Psi_0 \, \mathrm{e}^{-\mathrm{j}\, 2\pi f t} \, \mathrm{d}t = \Psi_0 \Delta t_H \, \frac{\sin\left(\pi f \Delta t_H\right)}{\pi f \Delta t_H} \,. \tag{2.59}$$

The spectral half-maximum half-width $f_H = \Delta f_H/2$ is defined by $\breve{\Psi}(f_H)/\breve{\Psi}(0) = 1/2$. Because $(\sin x)/x = 1/2$ for $x \approx 0.60 \times \pi \approx \pi/2$ holds, this leads to $\pi \left(\Delta f_H/2\right) \Delta t_H \approx \pi/2$ resulting in

$$\Delta t_H \, \Delta f_H \approx 1 \,. \tag{2.60}$$

The product of temporal width and spectral width is of the order of 1. This is a general property of Fourier-transform related functions.

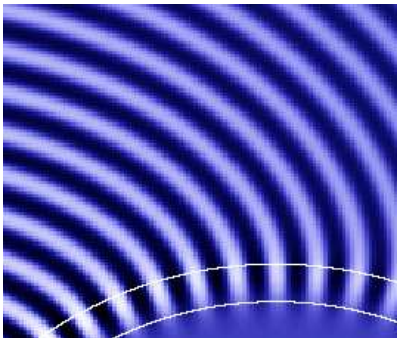
For a zero-width δ-function $\Psi(t) = \Psi_0 \, \delta(t)$ the spectrum $\breve{\Psi}(f) = \Psi_0$ would be infinitely broad and constant, so this is not a physical situation. In real systems, the initial impulse is usually Gaussian-shaped and starts with a half-maximum temporal full-width of $\Delta t_{H\,\mathrm{in}}$ (half-maximum spectral half-width $f_{H\,\mathrm{in}}$). The waveguide transfer function in the spectral domain may be approximated with a Gaussian low-pass having a half-maximum spectral half-width of $f_H \approx 0.441/\Delta t_H$

and a corresponding impulse response with a temporal half-maximum full-width of $\Delta t_H \approx \Delta t_g$. In this case the impulse's temporal half-maximum full-width $\Delta t_{H\,\mathrm{out}}$ at the fibre output and its half-maximum spectral half-width $f_{H\,\mathrm{out}}$ can be shown to be

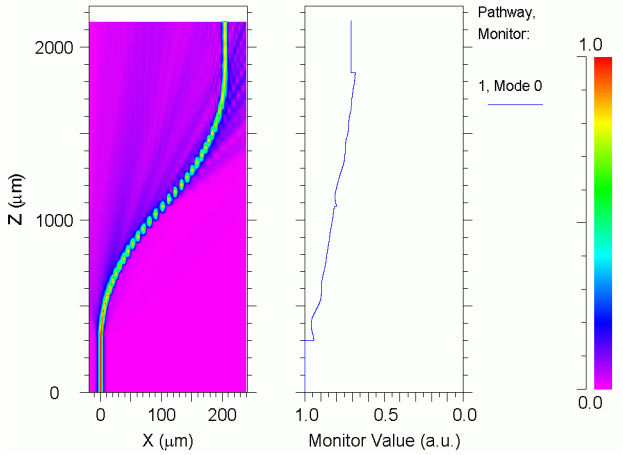
$$(\Delta t_{H\,\mathrm{out}})^2 = (\Delta t_H)^2 + (\Delta t_{H\,\mathrm{in}})^2, \quad f_{H\,\mathrm{out}}^{-2} = f_H^{-2} + f_{H\,\mathrm{in}}^{-2}, \quad f_H \approx 0.441/\Delta t_H\,. \tag{2.61}$$

2.3.7 Bend

For lossless dielectrics, the mode in a straight slab waveguide does neither change its shape nor does it loose energy for any arbitrary propagation length. The surfaces of constant phase are planes oriented perpendicular to the waveguide axis z. This may be seen from the structure of the solution Eq. (2.42) on Page 24: The transverse field components $\Psi(\vec{r})$ have the same constant phase in planes $z = \mathrm{const}$, Eqs. (2.44), (2.45). The situation is different as soon as the waveguide is bent. As to be seen[20] in Fig. 2.10(a), the plane wavefronts of the mode have to tilt according to the bend, and an observation point on the wavefront far away from the bend circle centre moves much faster than a point positioned at the same wavefront but nearer to the bend circle centre. When moving the far point more and more outwards, the velocity of the phase front at this point eventually exceeds the velocity of light in the cladding medium, the wavefronts start lagging, and power spills out. The mode is called "leaky"and becomes lossy.



(a) Phase fronts for a bent waveguide



(b) Electric field magnitude and waveguide power for an S-bend

Fig. 2.10. Slab waveguide bends in scalar approximation. (a) Phase surfaces for a circular waveguide bend[20] (b) Waveguide height $h = 5\,\mu$m, initial and ending straight section lengths $300\,\mu$m, bend radii 3 mm, angle of oblique section $15\,^\circ$, refractive index difference $n_1 - n_2 = 0.007$, cladding refractive index $n_2 = 1.5$, vacuum wavelength $\lambda = 1.55\,\mu$m. Excitation of fundamental guided mode $m = 0$ (Fig. 2.8) in straight waveguide at $z = 0$. (left) Contour plot of electric field magnitude with colour scale at rightmost side (right) Total power in waveguide cross-section along the waveguide axis (1, Mode 0, blue)

Looking at the S-bend displayed in Fig. 2.10(b)(left) we see that the mode is well confined for the first $300\,\mu$m. Then it hits the right turn, and power starts to leak out to the left. After the inclination point, when the left turn starts, power is spilling out to the right. The power leakage is well to be seen in the power plot Fig. 2.10(b)(right). When the wave enters the final straight waveguide section, the power remains again constant, but has only about 70 % of the input power level.

[20] Hammer, M.; Hiremath, K.; Stoffer, R.: Analytical approaches to the description of optical microresonator devices. International School of Quantum Electronics "Microresonators as Building Blocks for VLSI Photonics", Erice, Sicily, Italy, October 18–25, 2003. Slide 21

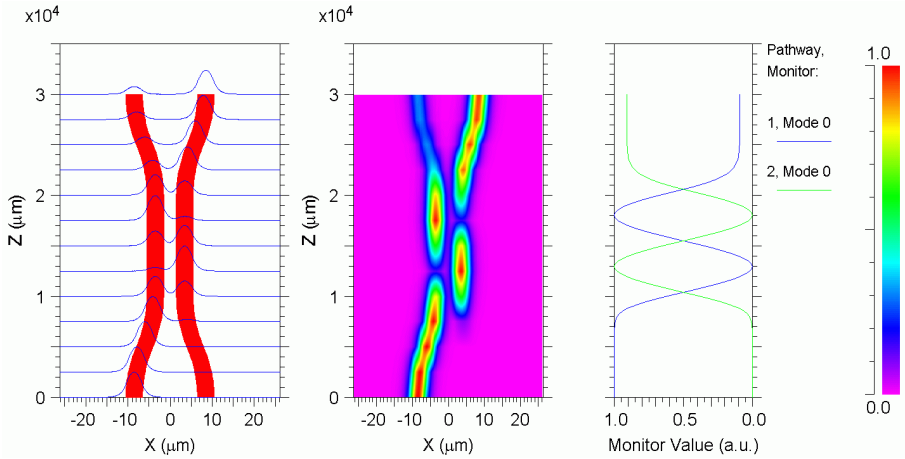


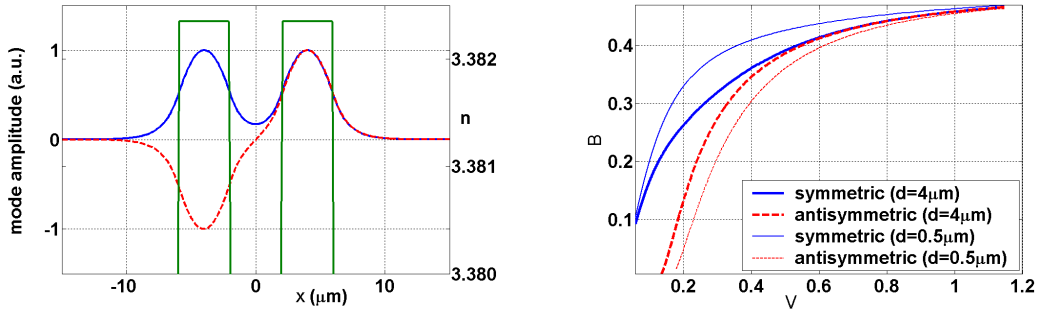
Fig. 2.11. Slab waveguide directional coupler in scalar approximation. Waveguide height $h = 4\,\mu$m, separation of inner core-cladding boundaries $3\,\mu$m, refractive index difference $n_1 - n_2 = 0.0067$, cladding refractive index $n_2 = 3.378$, vacuum wavelength $\lambda = 1.3\,\mu$m. Excitation of fundamental guided mode $m = 0$ (Fig. 2.8) in straight waveguide at lower left coupler input. Mind the vast scale differences for the x- and the z-directions; the curved section is angled only at about 6 °. (left) Slice plot of electric field magnitude (middle) Contour plot of electric field magnitude with colour scale at rightmost side (right) Total power in waveguide cross-sections along the axes of left-hand side waveguide (1, Mode 0, blue) and of right-hand side waveguide (2, Mode 0, green)

2.3.8 Directional coupler

Optical waveguides which run parallel to each other interact with one another by way of their evanescent fields. This interaction becomes significant when the evanescent fields from both waveguides overlap sufficiently. In Fig. 2.11(left) such a coupling structure with two slabs of equal heights $h = 4\,\mu$m is displayed. At $z = 0$ the slabs are far apart, but with increasing z they come nearer to each other such that the inner core-cladding boundaries are separated by $3\,\mu$m. The curvature radius is so large compared to the S-bend Fig. 2.10 that no power leakage can be seen. The straight waveguide at the lower left corner is excited by its fundamental mode $m = 0$. The result is a power transfer from the field in the left waveguide to the overlapping mode in the right slab. However, as soon as the power is completely transferred to the right waveguide, the system starts coupling back to the left waveguide, so we have a periodic energy exchange between both coupled waveguides. When the coupling zone ends and the waveguides start separating again, the actual power distribution gives the coupling ratio of this directional coupler. No significant part of the wave is seen to be propagating into the backward direction, and consequentially no power is leaving the lower-right waveguide end.

The oscillation of the power between both waveguides is clearly to be seen in the slice plot Fig. 2.11(left) and in the contour plot Fig. 2.11(middle). The power in both waveguides as traced in Fig. 2.11(right) oscillates according to spatial sinusoidals, which are offset by half their spatial period Λ. This behaviour may be mathematically described by coupled-mode theory,[21,22,23,24] but we follow here a more intuitive approach. Figure 2.12(a) shows the transverse refractive index profile of an infinitely long double-slab waveguide which is homogeneous along the propagation direction z. The eigenfields of this super-structure are called "supermodes". The fundamental symmetric and the next higher-order antisymmetric supermode are depicted in Fig. 2.12(a), the associated dispersion diagrams corresponding to Eq. (2.38) and Fig. 2.9 on Page 23 are to be seen in Fig. 2.12(b). Obviously, the (normalized) propagation constants B_s, β_s and B_a, β_a for the $s$ymmetric and $a$ntisymmetric modes differ for any fixed normalized frequency V. Assume that symmetric and

[21] Unger, H.-G.: Planar optical waveguides and fibres. Oxford: Clarendon Press 1977. Sect. 3.6 Page 252 ff.
[22] Huang Hung-chia: Coupled mode theory. Utrecht: VNU Science Press 1984
[23] Liu, M. M.-K.: Principles and applications of optical communications. Chicago: McGraw-Hill 1996. Sect. 11.2.3 and App. 14-A on Pages 540 ff. and 742 ff.
[24] Huang Hung-chia: Microwave approach to highly irregular fiber optics. New York: John Wiley & Sons 1998



(a) Electric field magnitudes and refractive index

(b) Dispersion diagrams

Fig. 2.12. Directional coupler with two infinitely long parallel slabs. Waveguide height $h = 4\,\mu$m, separation of inner core-cladding boundaries $d = 4\,\mu$m, refractive index difference $n_1 - n_2 = 0.0067$, cladding refractive index $n_2 = 3.378$, vacuum wavelength $\lambda = 1.3\,\mu$m. Fundamental symmetric (blue, ——) and next higher-order antisymmetric supermode (red, – – –) (a) Electric field magnitudes, and refractive index profile $n(x)$ (rectangular shapes, green, right-hand side axis) (b) Dispersion diagrams for wider ($d = 4\,\mu$m, thick lines) and narrower separated waveguides ($d = 0.5\,\mu$m, thin lines). Normalized frequency $V = \frac{h}{2} k_0 (n_1^2 - n_2^2)^{1/2}$

antisymmetric mode are excited with the transverse electric field magnitudes $E_s(x)$ and $E_a(x)$. The superposition of both fields is then

$$\begin{aligned} E_-(x,z) &= E_s(x)\,\mathrm{e}^{-\mathrm{j}\,\beta_s z} - E_a(x)\,\mathrm{e}^{-\mathrm{j}\,\beta_a z}, \\ E_+(x,z) &= E_s(x)\,\mathrm{e}^{-\mathrm{j}\,\beta_s z} + E_a(x)\,\mathrm{e}^{-\mathrm{j}\,\beta_a z}. \end{aligned} \tag{2.62}$$

Because the mode profiles of $E_s(x)$ and $E_a(x)$ are very similar in the regions of large fields strengths, the superposition $E_-(x,z)$ concentrates virtually all power in the left slab, while the superposition $E_+(x,z)$ has all the power in the right slab. Both supermodes propagate with different phase velocities $v_s = \omega/\beta_s$ and $v_a = \omega/\beta_a$, so the transverse field shapes $E_\mp(x,z)$ change along the propagation direction z. If the phase difference acquired is 2π, the constellation at $z = 0$ repeats. This happens after the so-called beat length Λ,

$$|\beta_s - \beta_a|\,\Lambda = 2\pi\,, \qquad \Lambda = \frac{2\pi}{|\beta_s - \beta_a|}\,. \tag{2.63}$$

However, at $z = \Lambda/2$ the phase shift is π, and the superposition $E_-(x,z)$ has changed to $E_+(x,z)$, and vice versa. At this point virtually all power has been transferred from one waveguide to the other, compare Fig. 2.11(right). The longer the beat length Λ is, i. e., the longer the interaction between both waveguides needs to be, the weaker the coupling has become. This can be understood from Fig. 2.12(b)(thin lines). If the waveguide separation is reduced from $h = 4\,\mu$m to $h = 0.5\,\mu$m, the propagation constants β_s and β_a differ more than before, and therefore the beat length Λ is reduced. However, for a tight coupling the field shapes $E_s(x)$ and $E_a(x)$ become more different even in the regions of large field strengths, and so the superpositions Eq. (2.62) cannot establish a true zero-power condition in one of the waveguides. If only the two lowest supermodes of Fig. 2.12(a) are guided, the power incident on the lower-left slab will be partially transferred to modes radiated into the cladding, but partially the power will be also reflected. These reflected waves excite backward running supermodes, and consequently some small amount of power will be seen to leave the lower-right slab.

The ratio of power leaving the coupler at the "wrong" lower-right port, and of power entering at the lower-left port is dubbed directivity. For a practical device, the directivity should be as small as possible. Typically, it is in the order of $-30 \ldots -20\,$dB.

2.3.9 Y-branch

For quite a few applications a full directional coupler is neither needed nor wanted. When a source should excite two waveguides with equal powers, then a Y-shaped branch as in Fig. 2.13 suffices. If

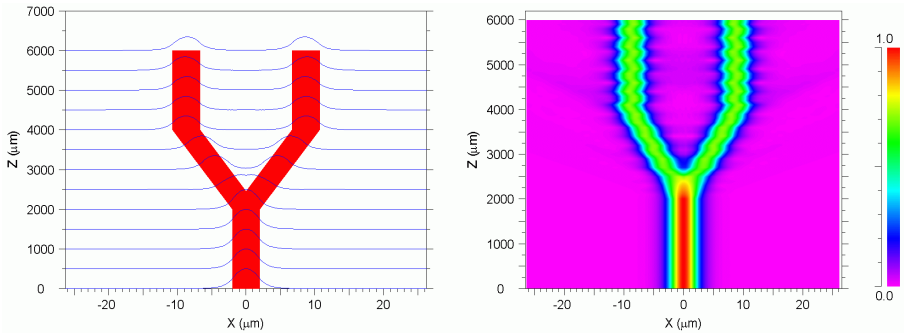


Fig. 2.13. Slab waveguide Y-branch in scalar approximation. Waveguide height $h = 4\,\mu$m, refractive index difference $n_1 - n_2 = 0.0067$, cladding refractive index $n_2 = 3.378$, vacuum wavelength $\lambda = 1.3\,\mu$m. Excitation of fundamental guided mode $m = 0$ (Fig. 2.8) in vertical "stem" of the Y-branch. Mind the vast scale differences for the x- and the z-directions; the branch section is angled at only 0.25. (left) Slice plot of electric field magnitude (right) Contour plot of electric field magnitude with colour scale at rightmost side

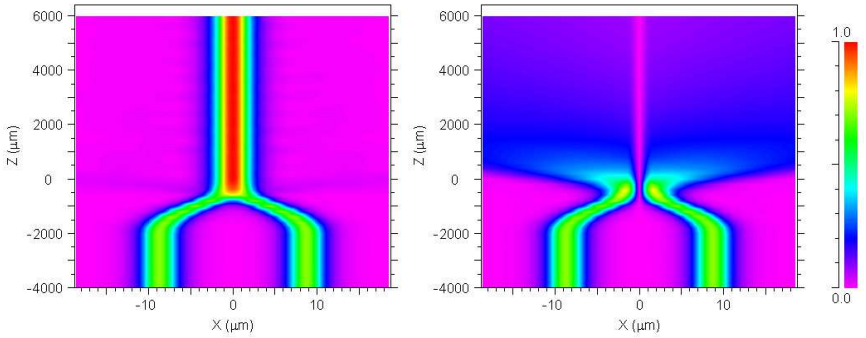


Fig. 2.14. Slab waveguide Y-branch in scalar approximation. Waveguide height $h = 3\,\mu$m, refractive index difference $n_1 - n_2 = 0.0067$, cladding refractive index $n_2 = 3.378$, vacuum wavelength $\lambda = 1.3\,\mu$m. Mind the vast scale differences for the x- and the z-directions; the branch section is angled at only 0.25. Contour plot of electric field magnitude with colour scale at rightmost side. (left) Symmetric-mode excitation (right) Antisymmetric-mode excitation

the axes of the waveguides in the branch section deviate only slightly from the feeding-waveguide axis, not much power will be radiated into the cladding, and the input power is equally split between the two output arms.

The reverse operation of a Y-branch functioning as a power combiner is not so straightforward, because its operation depends strongly on the excitation. Figure 2.14(left) displays the effect when the parallel-waveguide section is excited by its *symmetric supermode* as to be seen in Fig. 2.12(a). When the waveguides come nearer to each other gradually ("adiabatically"), the supermode adapts to the new structure without significant radiation into the cladding. At the Y-junction the two fields of both waveguides merge, and because they look alike and are in phase, the symmetric fundamental guided mode $m = 0$ (Fig. 2.8 on Page 21) of the straight slab is excited. Practically all power from the two Y-branches is combined into one output waveguide.

When the *antisymmetric supermode* is excited as in Fig. 2.14(right), the merger of the fields at the Y-junction looks like an antisymmetric mode $m = 1$ (Fig. 2.8), and tries to excite the corresponding mode in the output waveguide. However, the branch is operated such that the first antisymmetric mode $m = 1$ of the straight waveguide is in its cut-off state, and therefore next to all power is radiated into the cladding. A small part is also reflected into the backward direction, as it is the case whenever a waveguide inhomogeneity is involved.

2.4 Strip waveguide

Finite-sized integrated optics waveguides have to trap the light in two spatial directions, vertically as in a slab waveguide, and laterally, thereby forming a strip waveguide. In Fig. 2.15(a) the cross-section of a symmetric strip waveguide is depicted. The core region has a refractive index n_1, height h, and lateral width b. The surrounding medium with refractive index $n_2 < n_1$ "buries" the strip.

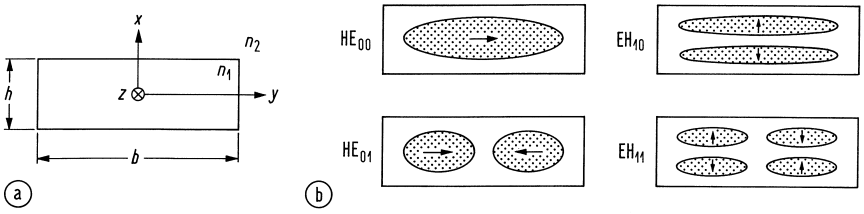
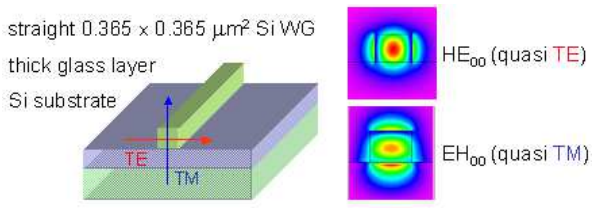


Fig. 2.15. Cross-section of a buried strip waveguide (a) geometric arrangement (b) transverse electric field and near-field intensity (dotted areas correspond to high intensity)

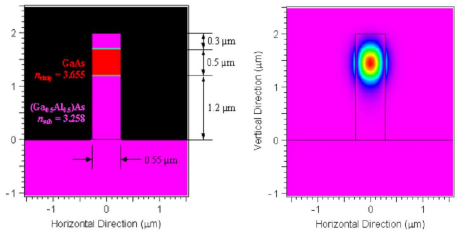
Guided modes exist, if the field is totally reflected at lateral boundaries $y = \pm b/2$, too, so the field oscillates spatially also in y-direction. This condition may be fulfilled if, e. g., two E_m-modes of a slab waveguide propagate with small angles $\pm\vartheta_S$ with respect to the z-axis, measured in the yz-plane, see Fig. 2.7). A *lateral* phase condition equivalent to the vertical phase condition Eq. (2.39) holds. A dispersion equation for u_S and w_S (subscript S like strip) similar to Eq. (2.41) may be calculated.

Consequently, a second, lateral mode index $n = 0, 1, 2, \ldots$ counts the $n + 1$ intensity maxima inside the strip along the y-direction. The slight tilt of the E_m-waves generates a H-component in z-direction in addition to the already existing E-component. Such modes with longitudinal *E- and H*-components are called *hybrid modes* or EH_{mn}-modes. The same ideas lead to the derivation of hybrid HE_{mn}-modes from H_m-waves of the slab waveguide.

Schematic examples for transverse electric field structures and near-field intensities are shown in Fig. 2.15(b). The fields are built by four congruences of equal intensity and appropriate polarization. Each pair of congruences consists of two orthogonally polarized congruences. The lateral fields $\Psi_n(y)$ have a structure as in Eq. (2.50) with the replacements $m \to n$, $x \to y$, $h \to b$, $u \to u_S$, $w \to w_S$. The vertical field $\Psi_m(x)$ of the strip is given by Eq. (2.50) with the substitutions $u \to u_S$

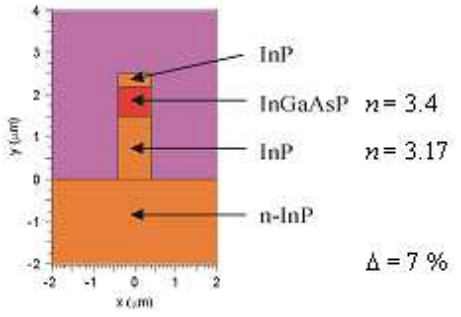


(a) SOI strip waveguide with fundamental mode field magnitudes

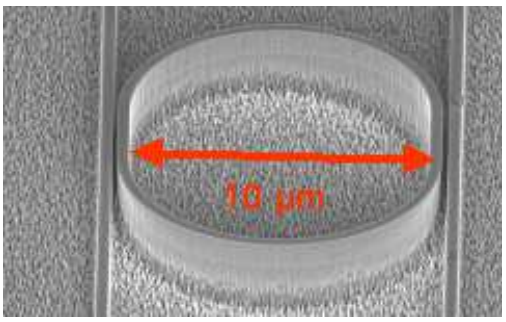


(b) GaAs/GaAlAs pedestal waveguide. Left: index profile. Right: quasi-TE field magnitude

Fig. 2.16. Unsymmetric strip waveguides as used in integrated optics. Insets: colour coded electric field magnitudes (red: maximum; violet: zero) of fundamental transverse electric mode (quasi-TE, HE_{00}-mode, electric field dominantly in substrate plane) and of fundamental transverse magnetic mode (quasi-TM, EH_{00}-mode, electric field dominantly normal to substrate plane) (a) Silicon strip waveguide ($n = 3.5$) with glass-layer substrate ($n = 1.44$) deposited on Silicon wafer (SOI, Silicon-on-insulator) (b) GaAs pedestal waveguide ($n_{\mathrm{GaAs}} = 3.655$) in $(\mathrm{Ga}_{0.5}\mathrm{Al}_{0.5})\mathrm{As}$ ridge ($n_{(\mathrm{Ga}_{0.5}\mathrm{Al}_{0.5})\mathrm{As}} = 3.258$)



(a) InGaAsP/InP pedestal waveguide, refractive index profile



(b) InGaAsP/InP pedestal waveguide, race-track ring filter

Fig. 2.17. InGaAsP/InP pedestal waveguide structure (a) Refractive index profile (b) Scanning electron microscope (SEM) picture of race-track ring resonator with two straight bus waveguides. Waveguide widths 400 nm, coupling gaps 100 nm (left) and 200 nm (right). At ring resonances, all power is transfered from one bus to the other. Signals at other frequencies pass unaffected

and $w \to w_S$. For weak guidance, the near field and its orthogonality relation are

$$\begin{aligned} &\Psi_{mn}(x,y) = \Psi_{N\,nm}\Psi_m(x)\,\Psi_n(y), \\ &n_1 \int_{-\infty}^{+\infty} \Psi_{mn}(x,y)\,\Psi^*_{m'n'}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \delta_{m\,m'}\,\delta_{n\,n'}\,. \end{aligned} \tag{2.64}$$

In integrated optics, mostly unsymmetrical strip waveguides are used, Fig. 2.16. Here, a low-index substrate like glass with $n = 1.44$ carries a high-index strip (e. g., Si with $n = 3.5$), which is covered either with air or with a low-index polymer for protection. For materials where such a large refractive index difference between strip ad substrate cannot be achieved, the waveguiding region must be elevated above the substrate plane. This is called a pedestal waveguide. Lateral index guiding is again controlled by an semiconductor-air boundary. For these high-index contrast waveguides, scalar optics is insufficient, and fully numerical vectorial optics must be employed.

Figure 2.16(a) shows a strip waveguide in Silicon-on-insulator (SOI) technology, and a GaAs/GaAlAs pedestal waveguide, Fig. 2.16(b). Insets in Fig. 2.16(a) and at the right hand side of Fig. 2.16(b) depict the colour coded electric field magnitudes of the fundamental transverse electric mode (quasi-TE, HE_{00}-mode, electric field dominantly in substrate plane) and of the fundamental transverse magnetic mode (quasi-TM, EH_{00}-mode, electric field mostly normal to substrate plane).

In Fig. 2.17 a practical example of an InGaAsP/InP pedestal waveguide structure is shown. The scanning electron microscope (SEM) picture Fig. 2.17(b) of a race-track ring resonator with two straight bus waveguides reveals the necessary precision. Waveguide widths of 400 nm, coupling gaps of 100 nm (left) and 200 nm (right) are typical. The straight section of the race-track has a length of 3 μm. The bus waveguides have measured attenuations of 4 dB / mm (TM) and 7 dB / mm (TE) which is tolerable for the short propagation distances involved. The differences in attenuation for TE and TM fields stem from the lithographical roughness of the vertical sidewalls. This root-mean-square (RMS) roughness was estimated to be $\sigma = (7 \pm 3)$ nm and had a correlation length of $L_c = (54 \pm 5)$ nm. For frequencies where the ring is in resonance, all power is transferred from one bus to the other. Signals at other frequencies pass unaffected.

2.5 Fibre waveguides

For transmission lengths in the 100 . . . 1 000 km range, low production cost, low attenuation (0.2 dB / km), and low group delay dispersion are important, so for this purpose quartz glass fibres are the only choice. For 100 m sized local areas very inexpensive *p*lastic *o*ptical *f*ibres (POF) can be also used.

Fibre waveguides (in short: fibres) consist of an isotropic, cylindersymmetric core with an diameter of $2a = 10 \ldots 100\,\mu$m, and an isotropic, cylindersymmetric cladding with lower refractive index n_2 and diameter $2b = 125\,\mu$m. For usual operating conditions, the fields are well confined to the core and do not reach far into the cladding. Therefore, the cladding may be safely assumed to be infinitely extended. Core and cladding region are described in cylindrical coordinates r, φ, z. The refractive index profile is assumed to be independent of φ, z. The cladding region is matched to the core index, i. e., there is no refractive index step at the core-cladding interface,

$$n^2(r) = \begin{cases} n_1^2\,[1 - 2\Delta g(r/a)]\,, & 0 \le r < a, \\ n_1^2\,[1 - 2\Delta] = n_2^2, & a \le r < \infty, \end{cases} \qquad g(r/a) = \begin{cases} 0, & r = 0, \\ 1, & r \ge a. \end{cases} \tag{2.65}$$

The refractive index on the fibre axis is denoted as n_1. The relative refractive index difference Δ was already defined in Eq. (2.38). For the profile function $g(r/a)$ the class of power low profiles is important,

$$g(r/a) = (r/a)^q, \qquad 0 \le q < \infty. \tag{2.66}$$

For $q \to \infty$ a step-index profile results ($n = n_1$ for $0 \le r < a$, $n = n_2$ for $r \ge a$). The choice of $q = 2$ defines a cladded parabolic profile. Fibres with non-constant core refractive index are called graded-index fibres. The profile exponent q is usually chosen such that intermodal dispersion becomes minimum.

In Sect. 2.3, Eq. (2.54), the group delay was discussed for the example of a slab waveguide. Figure 2.7 may be also understood as the cross-section of a fibre with core radius $a = h/2$ and constant core refractive index n_1. Modes correspond to rays which are differently inclined with respect to the fibre axis. The modal dispersion can be reduced, if the core refractive index changes

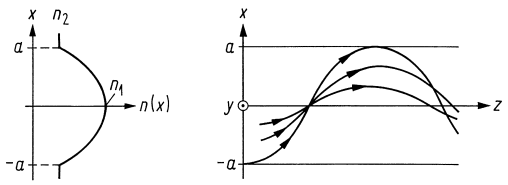


Fig. 2.18. Graded-index waveguide

gradually from the on-axis value n_1 to its cladding value n_2. In this case, the rays with larger geometrical path length propagate longer in regions with low refractive index so that the optical path lengths are equalized. Figure 2.18 shows the index profile $n^2(r)$ and the ray paths of such a graded-index waveguide.

2.5.1 Modal fields in a fibre

The vector wave equations (2.10) for the step-index fibre profile can be solved analytically by tabulated functions. For the cladded parabolic-index profile closed-form solutions are available only for the φ-independent TE-modes. The calculation for the step-index profile follows the procedure for the slab waveguide outlined in Sect. 2.3.3 on Page 24. However, the details of the calculations are rather involved. Therefore, we present only some graphical results, Fig. 2.19.

In analogy to the modes of a strip waveguide the fibre modes are in general hybrid modes and have to be classified by two subscripts (mode numbers), namely $\nu' = 0, 1, 2, \ldots$ for the azimuthal ($\cos \nu'\varphi$)- and ($\sin \nu'\varphi$)-dependencies of the vector field components, where $2\nu'$ counts the number of field extrema for $0 \le \varphi < 2\pi$, and $\mu = 1, 2, 3, \ldots$ counting the radial field extrema. For each pair of mode numbers (ν', μ) there are two field distributions denoted by $\mathrm{HE}_{\nu'\mu}$ and $\mathrm{EH}_{\nu'\mu}$, each with a $\cos \nu'\varphi$- or $\sin \nu'\varphi$-dependence. Consequently, four modes are associated with each pair of mode numbers having $\nu' \neq 0$, and two modes each for rotational symmetrical fields $\nu' = 0$. In contrast

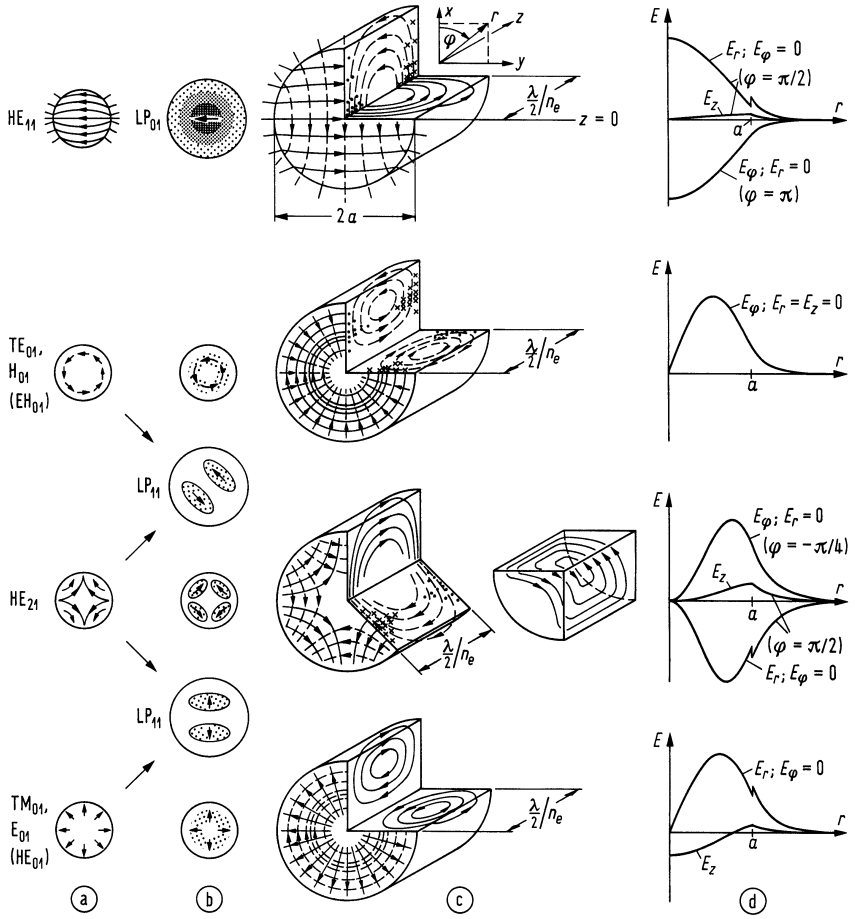


Fig. 2.19. Fields in a step-index fibre. Electric field strength (——); magnetic field strength (– – –); intensity (shaded). (a) electric field of vector modes and (b) electric field and intensity of vector and LP–modes, viewed in $-z$–direction (c) graphs of electric and magnetic fields, $B = 0.9$ (d) radial dependency of electric field strength, $V = 4.75$, $\Delta = 4\,\%$; E_r, E_φ (depicted in plane $z = 0$) are cophasal to each other and in quadrature to E_z (depicted in plane $z = \lambda/(4n_e)$)

to the field properties of a strip waveguide, a HE-mode (EH-mode) is not necessarily nearly a H-mode (E-mode). Only waves with $\nu' = 1$ have an on-axis intensity maximum, all other modes show an on-axis zero field.

The fundamental mode is denoted as HE_{11}-mode, Fig. 2.19. It has a zero cutoff frequency and propagates with identical propagation constants in two different shapes, one of which is rotated around the z-axis by 90 °. The mode looks similar to the transverse electric field of the fundamental H_{11}-mode of a circular waveguide with metallic boundaries. Unfortunately, this choice has some drawbacks. The $HE_{0\mu}$-modes for $\nu' = 0$ are $E_{0\mu}$-modes (i. e., not $H_{0\mu}$-modes!), and $EH_{0\mu}$-modes are $H_{0\mu}$-modes, see Fig. 2.19(a). All other modes are hybrid modes indeed.

For hybrid modes, the transverse components of $\vec{E}$ in planes $z = \text{const}$ are not perpendicular to the transverse components of $\vec{H}$ in the same planes. This can be easily seen by projecting the transverse field components $E_{x'}, H_{y'}$ in a local Cartesian coordinate system x', y', z' with varying angles between the z'-axis and the z-axis. If the z'-axis and the z-direction are not parallel, the fields $E_{x'}, H_{y'}$ develop in general both a longitudinal or z-component (become hybrid in the coordinate system without primes), and the angle between the projections of $E_{x'}, H_{y'}$ unto planes $z = \text{const}$ is no longer 90 °. The components E_r, E_φ are orthogonal and in phase, so the transverse

components are linearly polarized at each point of the cross-section area. However, the direction of polarization changes from point to point, see Fig. 2.19(c). The longitudinal component E_z is shifted by 90 ° with respect to the transverse components, Fig. 2.19(c). Because the normal component of the dielectric displacement $\vec{D}$ (the radial component in this case) must be continuous at the core-cladding interface $r = a$, the radial component of E_r is discontinuous, Eq. (2.8). Therefore we see kinks of the electric field lines at $r = a$ (Fig. 2.19(a),(c)), and discontinuities of E_r as in Fig. 2.19(d). For high index contrast strip waveguides, this behaviour is very pronounced, Fig. 2.16.

For weak guidance ($\Delta \ll 1$) the locally plane waves associated with the ray congruences propagate nearly in the direction of the fibre axis, therefore the longitudinal field components may be neglected so that the assumptions of scalar optics apply. Only $\mathrm{HE}_{1\mu}$-modes have an intensity maximum at the fibre axis. The propagation constants β of the mode pairs ($\mathrm{HE}_{2\mu}$, $\mathrm{H}_{0\mu}$), ($\mathrm{HE}_{2\mu}$, $\mathrm{E}_{0\mu}$) and ($\mathrm{HE}_{\nu'+2,\mu}$, $\mathrm{EH}_{\nu'\mu}$) with $\nu' \geq 1$ become asymptotically identical for $\Delta \to 0$ and fixed ν', μ, compare Page 23.

Modes with identical β superimpose to a resulting mode which does not change during propagation. Linear combinations of such modes with very similar β form approximative TEM-modes of the weakly guiding fibre where the fields are *l*inearly *p*olarized over the whole cross-section. Therefore, one single scalar quantity $\Psi_{\nu\mu}(r, \varphi)$ characterizes so-called $\mathrm{LP}_{\nu\mu}$-modes, which are eigenmodes of a weakly guiding fibre.

For the mode number sets $\nu \neq 0$, μ there are four $\mathrm{LP}_{\nu\mu}$-modes, namely fields with azimuthal dependencies $\cos\nu\varphi$, $\sin\nu\varphi$ (rotated against each other by 90 °) in two orthogonal linear polarizations each. The $\mathrm{LP}_{0\mu}$-modes correspond directly to the vector fields $\mathrm{HE}_{1\mu}$, and exist also in two polarizations each. The superposition of $\mathrm{LP}_{\nu\mu}$-modes by vectors modes is shown in Fig. 2.19(a),(b) with two examples.

For actual weakly guiding fibres the propagation constants of $\mathrm{HE}_{\nu+1,\mu}$-waves and $\mathrm{H}_{0\mu}$-, $\mathrm{E}_{0\mu}$- or $\mathrm{EH}_{\nu-1,\mu}$-waves differ slightly by an amount of, say, $\Delta\beta$. Therefore a "modal beat" is to be observed (see discussion near Eq. (2.63) on Page 31), i. e., the LP-modes rotate around the z-axis while propagating. After the beat length Λ, the azimuthal position of the initial modal field reproduces,

$$\Lambda = \frac{2\pi}{\Delta\beta} = \frac{\lambda}{n_e}\frac{\beta}{\Delta\beta}. \tag{2.67}$$

In the vicinity of the cutoff frequency of the $\mathrm{LP}_{\nu 1}$-modes of a step-index profile fibre, the smallest beat length amounts to $\Lambda \approx 3.4\,\lambda/(n_1\Delta^2)$. For $\lambda = 1.3\,\mu$m, $n_1 = 1.45$, $\Delta = 1\,\%$ this results in $\Lambda \approx 30$ mm. If $V \gg V_{\nu\mu G}$ the difference $\Delta\beta$ decreases and Λ increases; for a slab waveguide, this may be seen by inspection of Fig. 2.9(a): Intersections of $u \tan u$ for $m = 0$ with curves $V = \mathrm{const}$ are asymptotically found at the same abscissa values $u_{E,H} = \pi/2$.

2.5.2 Weakly guiding fibre in scalar approximation

For weak guidance $\Delta \ll 1$, only $\mathrm{LP}_{\nu\mu}$-modes (and not the complete vector solutions) are of interest. Therefore we do not solve Eq. (2.48), but rather the scalar Helmholtz equation (2.49) for a weakly inhomogeneous medium. As discussed in Sect. 2.5.1, the $\mathrm{LP}_{\nu\mu}$-modes may be characterized by only *one* transverse field component, e. g., by the component $\Psi_{\nu\mu}(\vec{r})$ in the direction of the linearly polarized electric field vector $\vec{E}$. In cylindrical coordinates r, φ, z, we make the separation ansatz

$$\begin{aligned}
&\Psi_{\nu\mu}(r,\varphi,z) = \Psi_{\nu\mu}(r,\varphi)\,\mathrm{e}^{-\mathrm{j}\,\beta_{\nu\mu}z}, \quad \nu,\nu' = 0,1,2,\dots\,, \quad \mu,\mu' = 1,2,3,\dots\,,\\
&\Psi_{\nu\mu}(r,\varphi) = \Psi_{\mu}^{(\nu)}(r)\Psi^{(\nu)}(\varphi), \qquad \Psi^{(\nu)}(\varphi) = \tfrac{1}{\sqrt{1+\delta_{\nu 0}}}\left\{\begin{matrix}\cos\nu\varphi\\ \text{or}\\ \sin\nu\varphi\end{matrix}\right\},\\
&n_1\int_{\varphi=0}^{2\pi}\int_{r=0}^{\infty}\Psi_{\nu\mu}(r,\varphi)\,\Psi_{\nu'\mu'}(r,\varphi)\,r\,\mathrm{d}r\,\mathrm{d}\varphi = \delta_{\nu\nu'}\delta_{\mu\mu'}\,.
\end{aligned} \tag{2.68}$$

$I_{\nu\mu}(r,\varphi) = \frac{1}{2}n_1|\Psi_{\nu\mu}(r,\varphi)|^2$ represents the normalized field intensity ($n_1 \approx n_2$). An orthogonality relation normalizes the total cross-sectional power to 1/2. Subscripts and superscripts ν, μ are

omitted if there is no danger of misunderstanding. The unit vectors in radial, azimuthal and axial direction are $\vec{e}_r$, $\vec{e}_\varphi$ and $\vec{e}_z$. The propagation vector of a locally plane wave (Sect. 2.2) in a weakly inhomogeneous medium is $\vec{k} = k_r\vec{e}_r + k_\varphi\vec{e}_\varphi + k_z\vec{e}_z$, $k_z = \beta$, $|\vec{k}| = k = k_0 n(r)$. The refractive index profile is rotationally symmetric, Eq. (2.65). For the wave function $\Psi_\mu^{(\nu)}(r)$ (having plane wave fronts because of Sect. 2.2, and therefore being real) we find from the scalar Helmholtz equation (2.49) together with Eq. (2.68) (primes on function symbols denote derivatives with respect to the argument)

$$\begin{aligned} \Psi''(r) + \frac{1}{r}\Psi'(r) + k_r^2(r)\,\Psi(r) = 0, \qquad & \Psi(r) \equiv \Psi_\mu^{(\nu)}(r), \\ k_r^2(r) = k_0^2 n^2(r) - k_\varphi^2 - \beta^2, \qquad & |k_\varphi| = \nu/r. \end{aligned} \tag{2.69}$$

The relation for $|k_\varphi|$ represents the azimuthal consistency or resonance condition for Ψ (well known from the discussion of the slab waveguide Eq. (2.39)): When moving around the fibre axis in a distance r, the accumulated phase shift must be an integer multiple of 2π, if the modal field is to remain stationary (not changing during propagation), and we find (Eq. (2.69))

$$-2\pi r|k_\varphi| = -2\nu\pi\,. \tag{2.70}$$

In analogy to Eq. (2.38) we define the core parameter u, the cladding parameter w (also expressed by the cladding penetration depth r_w), the normalized frequency V, the relative refractive index difference Δ, the factor κ (an abbreviation for a combination of modified Bessel functions K_ν of order ν) and the quantities B, δ (normalized propagation constants),

$$\begin{aligned} u &= a\sqrt{k_1^2 - \beta^2}, & w &= a\sqrt{\beta^2 - k_2^2} = 2a/r_w, \\ V &= ak_0 A_N = \sqrt{u^2 + w^2}, & A_N &= \sqrt{n_1^2 - n_2^2} = n_1\sqrt{2\Delta}, \\ \kappa(w) &= \mathrm{K}_\nu^2(w)/[\mathrm{K}_{\nu-1}(w)\ \mathrm{K}_{\nu+1}(w)], & \beta^2 &= k_1^2(1-2\delta), \\ B &= \frac{\beta^2 - k_2^2}{k_1^2 - k_2^2} = \frac{w^2}{V^2} = 1 - \frac{u^2}{V^2} = 1 - \frac{\delta}{\Delta} \approx \{\Delta \ll 1\} \approx \frac{\beta - k_2}{k_1 - k_2}. \end{aligned} \tag{2.71}$$

2.5.3 Step-index fibre

Equation (2.69) may solved for fibres with a step-index profile, $q \to \infty$ in Eqs. (2.65), (2.66). The light-shadow boundary is located at the core-cladding interface $r = a$. In analogy to Eq. (2.43) we find from Eq. (2.69) inside the core and the cladding two Bessel differential equations, from which a complicated-looking dispersion equation analogously to Eq. (2.41) may be derived. The resulting dispersion diagram is shown in Fig. 2.20. The cutoff frequencies $V_{\nu\mu G}$ of the $\mathrm{LP}_{0\mu}$- and $\mathrm{LP}_{2,\mu-1}$-modes for $\mu \geq 2$ are identical, see Fig. 2.20(a). At $V = V_{\nu\mu G}$ the field inside the core varies similar to a sine function.

At very high frequencies $k_0 \to \infty$ there is no evanescent field, and the mode is completely confined to the core. The propagation constant is then $B = 1$ or $\beta = n_1 k_0$ with a cladding parameter $w = V \to \infty$ approaching infinity. There are μ radial intensity maxima (compare to the slab waveguide Eq. (2.50). The radial extension of the fundamental fibre mode is at least the $1.6 \times a$. This may be compared to the equivalent relation $x_M \geq 1.8\,h$ (Eq. (2.53)) for the slab waveguide, where $x_M \mathrel{\widehat{=}} 2r_M$ and height $h \mathrel{\widehat{=}} 2a$.

For large normalized frequencies $V \gg 1$, the number M_g of guided modes may be estimated. Each pair of modal numbers (ν, μ) stands for 4 $\mathrm{LP}_{\nu\mu}$-modes or 4 vector modes (see Sect. 2.5.1). The number of guided modes is (without derivation, compare to Eq. (2.52) for the slab waveguide)

$$M_g \approx 4\left(\frac{1}{\pi}V - \frac{1}{4}\right)^2 \approx \frac{4}{\pi^2}V^2 \approx \frac{V^2}{2}, \qquad V \gg 1\,. \tag{2.72}$$

In the relation of the length-normalized group delay time t_g/L (Eq. (2.54), (2.71)) we can compute the group delay factor $\mathrm{d}(VB)/\,\mathrm{d}V$. For the group delay dispersion $\Delta t_g/L$ (Eq. (2.55),

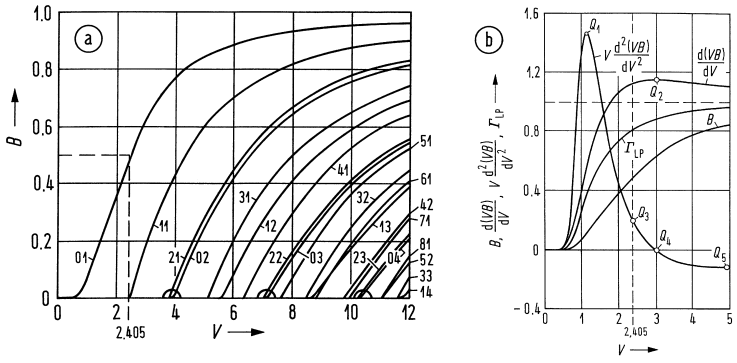


Fig. 2.20. Propagation of $\mathrm{LP}_{\nu\mu}$–modes in a step-index fibre. (a) Normalized propagation constant B; cutoff frequencies $V_{0\mu G} = V_{2,\mu-1,G}$ for $\mu \geq 2$ marked by semicircles, $V_{11G} = 2.405$, $V_{21G} = V_{01G} = 3.832$ (b) fundamental mode mode LP_{01}: normalized propagation constant B, group delay factor $\mathrm{d}(VB)/\mathrm{d}V$, dispersion factor $V\,\mathrm{d}^2(VB)/\mathrm{d}V^2$, and field confinement factor Γ_{LP}. $Q_1 = (1.15, 1.46)$, $Q_2 = (3, 1.14)$, $Q_3 = (2.405, 0.2)$, $Q_4 = (3.04, 0)$, $Q_5 = (4.95, -0.113)$

(2.71)) the dispersion factor $V\,\mathrm{d}^2(VB)/\,\mathrm{d}V^2$ may be derived. In Fig. 2.20(b) these quantities are displayed for the fundamental LP_{01} mode as a function of V. The field distribution for a fixed (ν, μ) may be integrated separately in the core and cladding cross-section to determine the relative modal powers in the respective cross-section areas. The total cross-sectional power is denoted P, the portions transported in the core and cladding region are P_1 and P_2, respectively. A so-called field confinement factor Γ_{LP} gives a measure of the field concentration,

$$\Gamma_{\mathrm{LP}} = \frac{P_1}{P} = 1 - \frac{P_2}{P}\,. \tag{2.73}$$

The ratio Γ_{LP} for the fundamental LP_{01} mode is also displayed in Fig. 2.20(b).

For $V \gg 1$ a step-index fibre guides a large number of modes, typically for $V = 75 \ldots 250$ a number of $M_g = 10^3 \ldots 10^5$ according to Eq. (2.72). From Eq. (2.54) we see that t_g is a monotonically increasing function of the group delay factor. The largest group delay difference of a strongly multimoded step-index fibre is approximately

$$\frac{\Delta t_{g\,\max}}{L} = \frac{n_{1g} - n_{2g}}{c}\left(1 - \frac{2}{V}\right) \approx \frac{n_1 - n_2}{c}\left(1 - \frac{2}{V}\right). \tag{2.74}$$

When $V \to \infty$, then $\Delta t_{g\,\max}$ corresponds to the group delay difference of homogeneous plane waves propagating in a medium n_1 or n_2, respectively, therefore $\Delta t_{g\,\max}/L \sim \Delta$ holds. The quantity $(n_1 - n_2)/c$ is the length-related group delay difference of a ray parallel to the fibre axis, and a ray which intersects the axis at an angle $\vartheta = \vartheta_T$ being defined as the critical angle of total internal reflection for a plane wave incident at a plane interface as in Fig. 2.5, $\cos\vartheta_T = n_2/n_1$.

For a typical step-index multimode fibre with $n_1 = 1.5$ and $\Delta = 1\,\%$ the relative group delay spread amounts to $\Delta t_{g\,\max}/L \approx n_1\Delta/c = 50\,\mathrm{ns}\,/\,\mathrm{km}$.

Conventional single-mode fibre (CSF)

Conventional single-mode fibres (CSF, also called *n*on*d*ispersion-*s*hifted *f*ibres NDSF) are characterized by a large core diameter $2a \approx 9\,\mu\mathrm{m}$ and a low relative refractive index difference in the order of $\Delta = 0.1 \ldots 0.3\,\%$. As an example,[25] we consider a step-index fibre with $n_1 = 1.450840$, $n_2 = 1.446918$, $a = 4.1\,\mu\mathrm{m}$ and $\Delta = 0.27\,\%$. The cutoff wavelength of this fibre is $\lambda_{11G} = 1.142\,\mu\mathrm{m}$, the first-order zero dispersion wavelength λ_C is given by $C(\lambda_C = 1.325\,\mu\mathrm{m}) = 0$. The dispersion

[25] Ghatak, A.; Thyagarajan, K.: Introduction to fiber optics. Cambridge: University Press 1998. Table 10.2 Page 189

$\lambda/\,\mu$m	1.1	1.3	1.56
V	2.497	2.113	1.761
$V\,\mathrm{d}^2(VB)/\,\mathrm{d}V^2$	0.150	0.370	0.710
$M\,/\,\frac{\mathrm{ps}}{\mathrm{km\,nm}}$	−23	+1.6	+22
$W\,/\,\frac{\mathrm{ps}}{\mathrm{km\,nm}}$	−1.8	−3.7	−5.9
$C\,/\,\frac{\mathrm{ps}}{\mathrm{km\,nm}}$	−25	−2.1	+16
$D(\lambda_C = 1.325\,\mu\mathrm{m}) = 0.083\,\mathrm{ps}\,/\,(\mathrm{km\,nm^2})$			

Table 2.1. Dispersion characteristics[25] of a step-index CSF with $n_1 = 1.450840$, $n_2 = 1.446918$, $a = 4.1\,\mu$m and $\Delta = 0.27\,\%$. Wavelength of zero chromatic dispersion $C(\lambda_C) = 0$ is $\lambda_C = 1.325\,\mu$m, cutoff at $\lambda_{11G} = 1.142\,\mu$m

characteristics are listed in Table 2.1 and displayed in Fig. 2.21(a). Such fibres, e. g., SMF-28™, are referred to as G.652 type by the International Telecommunication Union (ITU).

As discussed in Sect. 2.3.6, the effect of chromatic dispersion is to increase the width of an input impulse at the waveguide output, Fig. 1.5(a) on Page 4. If $C = 16\,\mathrm{ps}\,/\,(\mathrm{km\ nm})$ is specified, a 17 ps wide input impulse emitted by a source with a spectral width of $\Delta\lambda = 0.49\,\mathrm{nm}$ (corresponding to $\Delta f = (17\,\mathrm{ps})^{-1} = 60\,\mathrm{GHz}$ at a wavelength of $1.55\,\mu$m) would be broadened to a width of roughly (Eq. (2.61) on Page 29) $\sqrt{17^2 + (16 \times 0.49 \times 5)^2}\,\mathrm{ps} = \sqrt{17^2 + 39^2}\,\mathrm{ps} = 43\,\mathrm{ps}$ after a propagation length of only $L = 5\,\mathrm{km}$. This means that a strong intersymbol interference would be observed for a signal with a bit rate of 40 Gbit/s (period $T_t = 25\,\mathrm{ps}$), Fig. 1.5(b) on Page 4.

Dispersion shifted fibre (DSF)

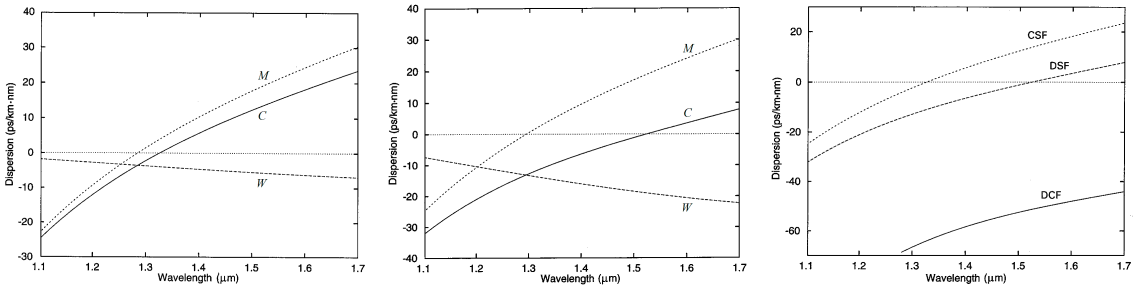
At a wavelength larger than the zero λ_0 of the first-order material dispersion $M(\lambda_0) = 0$, we see from Fig. 2.4 that $M > 0$. Because of $n_{1g} - n_{2g} > 0$ and Fig. 2.20(b), the first-order waveguide dispersion in the whole single-mode range $V < V_{11G} = j_{0,1} = 2.405$ has the opposite sign as M, $W = -(n_{1g} - n_{2g})/(c\lambda) \cdot V\,\mathrm{d}^2(VB)/\,\mathrm{d}V^2 < 0$ (Eq. (2.55)). Choosing $W(\lambda_C)$ suitably reduces therefore the first-order chromatic dispersion to zero, $C(\lambda_C) = 0$ for a certain wavelength $\lambda_C > \lambda_0$. Preferably, the minimum attenuation wavelength $\lambda_C = \lambda_\alpha = 1.55\,\mu$m is chosen (see Fig. 2.3 on Page 15). Such fibres are denoted as *d*ispersion *s*hifted *f*ibres (DSF). With a constant $V = ak_1\sqrt{2\Delta}$, the waveguide dispersion $|W|$ may be enlarged by diminishing a and increasing Δ simultaneously. Table 2.2 lists and Fig. 2.21(b) displays the dispersion properties of a DSF with $n_1 = 1.457893$, $n_2 = 1.446918$, $a = 2.3\,\mu$m and $\Delta = 0.75\,\%$. The cutoff wavelength of this fibre is $\lambda_{11G} = 1.073\,\mu$m, the first-order zero dispersion wavelength λ_C is given by $C(\lambda_C = 1.523\,\mu\mathrm{m}) = 0$.

$\lambda/\,\mu$m	1.1	1.3	1.56
V	2.346	1.985	1.654
$V\,\mathrm{d}^2(VB)/\,\mathrm{d}V^2$	0.223	0.476	0.845
$M\,/\,\frac{\mathrm{ps}}{\mathrm{km\,nm}}$	−25	+0.6	+22
$W\,/\,\frac{\mathrm{ps}}{\mathrm{km\,nm}}$	−7.3	−13.2	−20
$C\,/\,\frac{\mathrm{ps}}{\mathrm{km\,nm}}$	−32	−12.6	+2
$D(\lambda_C = 1.523\,\mu\mathrm{m}) = 0.048\,\mathrm{ps}\,/\,(\mathrm{km\,nm^2})$			

Table 2.2. Dispersion characteristics[26] of a step-index DSF with $n_1 = 1.457893$, $n_2 = 1.446918$, $a = 2.3\,\mu$m and $\Delta = 0.75\,\%$. Wavelength of zero chromatic dispersion $C(\lambda_C) = 0$ is $\lambda_C = 1.523\,\mu$m, cutoff at $\lambda_{11G} = 1.073\,\mu$m

The small core size reduces the significant modal extension $r_M \approx 1.8 \times a$ (Page 38 and Eq. (2.53)) so that coupling to fibres becomes more difficult. Further, the increased doping for

[26] See Table 10.3 Page 191 in reference Footnote 25 on Page 39



(a) CSF: $\Delta = 0.27\,\%$, $a = 4.1\,\mu$m, $\lambda_C = 1.325\,\mu$m, $\lambda_{11G} = 1.142\,\mu$m (Table 2.1)

(b) DSF: $\Delta = 0.75\,\%$, $a = 2.3\,\mu$m, $\lambda_C = 1.523\,\mu$m, $\lambda_{11G} = 1.073\,\mu$m (Table 2.2)

(c) DCF: $\Delta = 2\,\%$, $a = 1.5\,\mu$m, $\lambda_{11G} = 1.158\,\mu$m (Table 2.3). Comparison of CSF, DCF and DCF

Fig. 2.21. Dispersion characteristics of basic step-index fibre types with cladding index $n_2 = 1.446918$ as described in Tables 2.1–2.3. (a), (b) Upper curves: Material dispersion M. Lower curves: Waveguide dispersion W. Middle curves: Total chromatic dispersion C (all in units of ps /(km nm)), see Eq. (2.55) on Page 27. (c) Comparison of CSF (upper curve), DSF (middle curve) and DCF (lower curve) [adapted from Figs. 10.2–10.4 in reference Footnote 25 on Page 39]

a larger Δ causes more attenuation, Sect. 2.1.4. Usually, these fibres have therefore a triangular profile for reducing the necessary Δ and to achieve a lower attenuation. They are classified as G.653 fibres by the ITU.

Dispersion-compensating fibre (DCF)

In many countries, tens of millions of kilometers of CSF already exist in the underground ducts operating at a zero-dispersion ($M(\lambda_0) = 0$, $C(\lambda_C) = 0$) wavelength $\lambda_0 \approx \lambda_C \approx 1.3\,\mu$m. One could increase the transmission capacity by operating these fibres at the minimum attenuation wavelength $\lambda_\alpha = 1.55\,\mu$m (see Fig. 2.3), and using *w*avelength *d*ivision *m*ultiplexing (WDM, employing multiple optical wavelengths) together with optical amplifiers.

$\lambda/\,\mu$m	1.1	1.3	1.56
V	2.531	2.141	1.784
$C\,/\,\frac{\text{ps}}{\text{km nm}}$	−97	−67	−50

Table 2.3. Dispersion characteristics[27] of a step-index DCF with $n_1 = 1.476754$, $n_2 = 1.446918$, $a = 1.5\,\mu$m and $\Delta = 2\,\%$, cutoff at $\lambda_{11G} = 1.158\,\mu$m

But, then there will be a significant residual (positive) dispersion, see Table 2.1. On the other hand, replacing these fibres by DSF would involve huge costs. As such, in recent years, there has been considerable work in upgrading the installed $1.3\,\mu$m-optimized optical fibre links for operation at $1.55\,\mu$m. This is achieved by developing fibres with very large negative first-order dispersion coefficients C, a few hundred meters to a kilometer, which can be used to compensate for dispersion over tens of kilometers of the fibre in the link. These types of fibres are known as *d*ispersion-*c*ompensating *f*ibres (DCF).[28]

They are realized by making the waveguide dispersion still stronger than for DSF, so that the total dispersion becomes negative at $\lambda_\alpha = 1.55\,\mu$m. Table 2.3 and Fig. 2.21(c) review the dispersion properties of a DCF in comparison to CSF and DSF. The fibre data are $n_1 = 1.476754$, $n_2 = 1.446918$, $a = 1.5\,\mu$m, $\Delta = 2\,\%$, and a cutoff wavelength of $\lambda_{11G} = 1.158\,\mu$m.

DCF are designed with a large Δ for a dispersion as high as $C = -300\,\text{ps}\,/\,(\text{km nm})$, therefore attenuation becomes significant as discussed above. Hence, for DCF a measure of the dispersion

[27] See Fig. 10.4 Page 191 in reference Footnote 25 on Page 39

[28] See Sect. 15.3 Page 325 in reference Footnote 25 on Page 39

compensation efficiency is given by the *f*igure *o*f *m*erit (FOM, unit ps / (dB nm)), which is defined as the ratio of the dispersion coefficient $|C|$ (unit ps / (km nm)) and the total length-related loss a/L_{DCF} (unit dB / km, Eq. (2.23)) of the DCF with length L_{DCF},

$$\mathrm{FOM} = \frac{|C|}{a/L_{\mathrm{DCF}}}\,, \qquad P(L_{\mathrm{DCF}}) = P_0\,\mathrm{e}^{-\alpha L_{\mathrm{DCF}}}, \quad \frac{a}{L_{\mathrm{DCF}}} = \frac{10}{L_{\mathrm{DCF}}}\,\mathrm{lg}\,\frac{P_0}{P(L_{\mathrm{DCF}})} = 4.34\,\alpha\,. \quad (2.75)$$

For a dispersion of $C = -300\,\mathrm{ps}\,/\,(\mathrm{km\,nm})$ the FOM is typically $\mathrm{FOM} = 400\,\mathrm{ps}\,/\,(\mathrm{dB\,nm})$.

Dispersion flattened fibre (DFF)

More complicated profile functions (e. g., a triple-clad step-index fibre having n_2, n_3, n_4 as depicted in Fig. 2.22(b)) allow more zeros of C. The design goal of a optimum flat $C(\lambda)$ *and* low doping

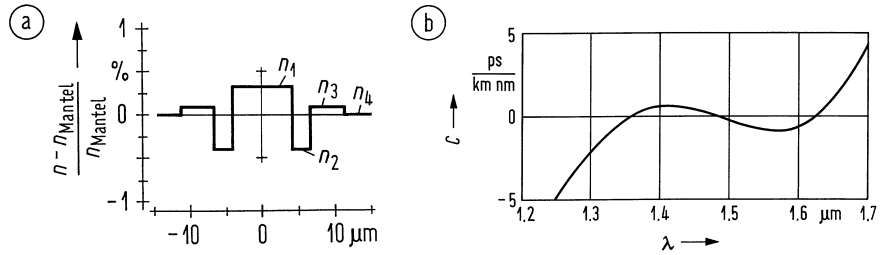


Fig. 2.22. Dispersion characteristics of a triple-clad step-index DFF. (a) Refractive index profile (b) Chromatic dispersion. Quartz glass core with F-doped claddings (computed) $a_1 = 3.8\,\mu$m, $a_2 = 7\,\mu$m, $a_3 = 12.95\,\mu$m. Refractive indices at $\lambda = 1.064\,\mu$m are $n_1 = 1.45$, $n_2 = 1.4383$, $n_3 = 1.4471$, $n_4 = 1.4442$. Three zeros $C(\lambda_{C\,1,2,3}) = 0$ at $\lambda_{C\,1} = 1.36\,\mu$m, $\lambda_{C\,2} = 1.48\,\mu$m, $\lambda_{C\,3} = 1.625\,\mu$m. (Mantel = cladding)

with $\Delta < 1\,\%$ for low attenuation (see Sect. 2.1.4) can be simultaneously achieved. The dispersion factor my be chosen for three zeros $C(\lambda_C) = 0$ in the wavelength range $1.3\,\mu\mathrm{m} \le \lambda \le 1.65\,\mu$m, Fig. 2.22(b). For fabricated fibres a value of $|C| \le 2\,\mathrm{ps}\,/\,(\mathrm{km\,nm})$ is feasible. A completely flat chromatic dispersion $C(\lambda) = 0$ in a certain wavelength range may be achieved in principle. However, the required refractive index profiles have so many accurately specified steps that its production becomes unmanageable. Fibres of the type Fig. 2.22 are called *d*ispersion *f*lattened *f*ibres (DFF) or *d*ispersion *f*lattened *s*ingle*m*ode fibres (DFSM).

Singlemode fibre data

As has been shown previously, the rate of dispersion change is expressed as dispersion slope, Eq. (2.56b) on Page 27. For high speed multi-channel DWDM networks, a lower dispersion slope enables more uniform and optimum performance across the entire wavelength band. Another advantage to a low dispersion slope is around the nonlinear phenomenon of four wave mixing (FWM). Very low dispersion, which can happen at the lower end of the C band in nonzero-dispersion fibre (NZDF) with high dispersion slope, can result in FWM and degradation of multi-channel DWDM system performance. With a low dispersion slope, a fibre allows its minimum dispersion to be increased in this region to better suppress FWM, while keeping the fibre's maximum dispersion small enough for signals to travel over long distances with minimum need for costly dispersion compensation, see Table 2.4.

Sophisticated multilayer core designs as the one in Fig. 2.22 can reduce the dispersion slope to $0.045\,\mathrm{ps}\,/(\mathrm{km\ nm^2})$ (Lucent TRUEWAVE_1550, $A_{\mathrm{eff}} = 59\,\mu\mathrm{m}^2$, mode diameter $8.6\,\mu$m), but there are trade-offs. An important one is that reduced dispersion-slope designs tend to have smaller mode-field diameters than conventional ones with a larger dispersion slope of $0.086\,\mathrm{ps}\,/(\mathrm{km\ nm^2})$ (Corning SMF28_1550, $A_{\mathrm{eff}} = 85\,\mu\mathrm{m}^2$, mode diameter $10.4\,\mu$m). For smaller A_{eff}, more optical power is concentrated in a smaller fibre cross-section thereby raising the power density and increasing the strength of nonlinear effects, which can cause crosstalk between WDM channels.

Fiber Type	Attenuation	Chromatic Dispersion		MFD	Polarization
	α_0 [dB/km]	D [ps/nm/km]	D' slope [ps/nm^2/km]	A_{eff} [µm^2]	PMD [ps/√km]
PirelliWIDELIGHT_1550	0.24	-6.85	0.157	51	≤ 0.1
PirelliWIDELIGHT_1625	0.25	-0.1	0.107	51	≤ 0.1
PirelliFREELIGHT_1550	0.23	4.3	0.114	72	≤ 0.1
PirelliFREELIGHT_1625	0.25	11.2	0.11	72	≤ 0.1
PirelliDEEPLIGHT_1550	0.23@1560nm	-2.2 @1560nm	0.12	70	≤ 0.1
CorningSMF28_1310	0.34	0 @1313nm	0.086	66.5	≤ 0.1
CorningSMF28_1550	0.19	16	0.086	85	≤ 0.1
CorningSMF28e_1310	0.34	0 @1313nm	0.086	66.5	≤ 0.1
CorningSMF28e_1550	0.19	16		85	≤ 0.1
CorningLEAF	0.2 @1550nm	4 @1550nm	0.1 @1550nm	72	≤ 0.1
CorningLEAF_submarine	0.2 @1550nm	0 @1580nm	0.11 @1580nm	71	≤ 0.1
FurukawaSMF332_1310	0.32	0	0.092	68	≤ 0.5
FurukawaSMF332_1550	0.18	18	0.092	86.5	≤ 0.5
AlcatelSMF_1310	0.3	0	0.086	63.6	≤ 0.1
AlcatelSMF_1550	0.2	16		81.67	≤ 0.1
AlcatelTERALIGHT_1550	0.205	8	0.058	65	≤ 0.1
AlcatelTERALIGHT_1620	0.22	10.9 @1600nm	0.058	65	≤ 0.1
LucentTRUEWAVE_1600	0.2	4.5	0.045	55	≤ 0.1
LucentTRUEWAVE_1550	0.2	7	0.045	59	≤ 0.1
LucentALLWAVE_1310	0.3	0 @1312nm	0.088	66	≤ 0.1
LucentALLWAVE_1550	0.2	0 @1312nm	0.088	80	≤ 0.1
SumitomoZ_1550	0.17	18.5	0.056	80	
SumitomoZPLUSa_1550	0.168	20.5	0.059	110	

Table 2.4. Singlemode fibre data from various commercial suppliers. D and D' correspond to the respective quantities C and D in Eq. (2.56b) on Page 27. The mode-field diameter (MFD) is expressed by the effective mode field area A_{eff}. Group delay differences between orthogonally polarized fundamental fibre modes are specified by the PMD parameter (*p*olarization-*m*ode *d*ispersion, see also Page 28). [Table after OptSim 4.5 Models Reference Volume I: Sample Mode (2005). RSoft Design Group, Inc., Physical Layer Division, Ossining, NY 10562. `http://www.rsoftdesign.com`]

*P*olarization *m*ode *d*ispersion (PMD) becomes important for integrated-optics waveguides with large Δ, or for optical fibres where short light impulses (e. g., $\Delta t_H = 17\,\text{ps}$ for a bit rate of 40 Gbit/s, bit period $T_t = 1/40\,\text{GHz} = 25\,\text{ps}$) travel over long distances. Because of fibre imperfections which cause group delay differences and induce couplings between the two orthogonally polarized fundamental modes, there exists a PMD group delay spread, which increases less than linearly with transmission length L,

$$\Delta t_{gp} = D_p\sqrt{L}\,. \tag{2.76}$$

For a typical PMD of $D_p = 0.1\,\text{ps}/\sqrt{\text{km}}$ and a propagation length of $L = 625\,\text{km}$, impulses in orthogonal polarizations would be displaced in time by 2.5 ps, i. e., by a tenths of a bit period T_t. This uncertainty could prove prohibitive for a real high-capacity system.

2.5.4 Parabolic-index fibre

An analytical solution of Eq. (2.69) exists also for an (unphysically) *infinitely extended* parabolic refractive index profile Eq. (2.65) with $g(r/a) = (r/a)^2$ in $0 \le r < \infty$. The results are discrete eigenvalues β, δ for the Gauss-Laguerre modes. A LP_{01} mode has the transverse dependency of a

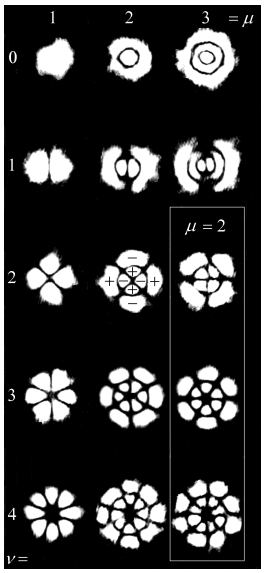


Fig. 2.23. Gauss-Laguerre modes of a parabolic-index fibre with parameters $a = 23\,\mu\text{m}$, $A_N = 0.2$, $V = 46$ at $\lambda = 0.6328\,\mu\text{m}$. The beam radius of the fundamental LP_{01}-mode is $w_0 = 4.8\,\mu\text{m}$.

rotationally symmetric Gaussian near-field with field radius w_0. The solutions describe approximately the modes of a physical parabolic fibre *with an actual cladding* (Eqs. (2.65), (2.66)), as long as the solution fields do not reach significantly into the cladding beyond the physical core-cladding interface at $r = a$. This is true for most of the modes if $V \gg V_{\nu\mu G}$. The cutoff frequency $V_{\nu\mu G}$ may be estimated observing that $\beta = k_2$ at cutoff. Proceeding as in Eq. (2.72) the number M_g of guided modes is found to be

$$V_{\nu\mu G} = 2m = 2(\nu + 2\mu - 1), \qquad M_g \approx m^2_{\max 2} = \frac{V^2}{4}, \qquad V \gg 1\ . \tag{2.77}$$

For very short wavelengths $k_0 \to \infty$ the fields are totally concentrated to the fibre axis. In contrast, the modes of a step-profile fibre (or a slab waveguide) always fill the total cross-section, Fig. 2.19. Figure 2.23 displays the intensity of the 12 lowest-order Gauss-Laguerre modes in a parabolic-index fibre. The fields have either a $(\sin \nu\varphi)$- or a $(\cos \nu\varphi)$-dependency on the azimuthal angle φ. Identical intensity distributions stem from two orthogonal polarizations. Regions adjacent to each other in the radial or azimuthal direction have alternating signs of the associated wave function $\Psi_{\nu\mu}(r, \varphi)$ of Eqs. (2.68), (2.69), see Fig. 2.23 for $(\nu = 2, \mu = 2)$. The field radii increase with increasing mode order.

The maximum group delay difference of guided modes with $0 < \delta < \Delta$ ($n_1 \approx n_{1g}$, $\mathrm{d}\delta/\,\mathrm{d}\omega \approx -n_{1g}\delta/k_1$) becomes

$$\frac{t_g}{L} = \frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \frac{n_{1g}}{c}\left(1 + \frac{\delta^2}{2}\right), \qquad \frac{\Delta t_{g\,\max}}{L} = \frac{n_{1g}}{c}\frac{\Delta^2}{2}. \tag{2.78}$$

Compared to $\Delta t_{g\,\max}/L \sim \Delta$ for a step-index profile (Eq. (2.74)), the group delay dispersion for a parabolic-index fibre is reduced to $\Delta t_{g\,\max}/L \sim \Delta^2/2$.

For a typical graded-index fibre with $n_1 = 1.5$ and $\Delta = 1\,\%$ the relative group delay spread amounts to $\Delta t_{g\,\max}/L \approx n_1\Delta^2/(2c) = 0.2\,\text{ns}\,/\,\text{km}$, much smaller than the spread of $\Delta t_{g\,\max}/L \approx n_1\Delta/c = 50\,\text{ns}\,/\,\text{km}$ for an equivalent step-index fibre, Eq. (2.74) on Page 39.

2.5.5 Orthogonality and coupling efficiency

A weakly guiding, lossless fibre be excited in plane $z = z'$ by a scalar light field $\Phi(r, \varphi, z)$ having a total power of P_Φ. The source field $\Phi(r, \varphi, z \geq z')$ may be expanded into a set of orthonormal

guided modes $\Psi_{\nu\mu}(r,\varphi,z)$ ($n_1 \approx n_2 \approx n(r)$) *and* non-guided modes (symbolically NG),

$$\Phi(r,\varphi,z) = \sum_{\nu,\mu} c_{\nu\mu}(z')\,\Psi_{\nu\mu}(r,\varphi,z) \;+\; \mathrm{NG}, \quad P_\Phi = \frac{n_1}{2}\int_0^{2\pi}\int_0^{\infty} |\Phi(r,\varphi,z)|^2\, r\,\mathrm{d}r\,\mathrm{d}\varphi\,. \tag{2.79}$$

The coefficients $c_{\nu\mu}$ are called excitation coefficients, coupling coefficients, or Fourier coefficients. By multiplying Eq. (2.79) at $z = z'$ with $\Psi^*_{\nu'\mu'}(r,\varphi,z')$ and integrating over the cross-sectional area, we write (disregarding for the moment the non-guided modes NG):

$$\int_0^{2\pi}\int_0^{\infty} \Phi(r,\varphi,z')\,\Psi^*_{\nu'\mu'}(r,\varphi,z')\,r\,\mathrm{d}r\,\mathrm{d}\varphi = \sum_{\nu,\mu} c_{\nu\mu}(z') \underbrace{\int_0^{2\pi}\int_0^{\infty} \Psi_{\nu\mu}(r,\varphi,z')\,\Psi^*_{\nu'\mu'}(r,\varphi,z')\,r\,\mathrm{d}r\,\mathrm{d}\varphi}_{\delta_{\nu\nu'}\delta_{\mu\mu'}/n_1}$$

Because of the orthogonality of guided modes ($\Psi_{\nu\mu}(r,\varphi,z)$, Eq. (2.68)) and the non-guided modes (NG, not proven here), the sum reduces to one element, $\sum_{\nu,\mu} c_{\nu\mu}(z')\,\delta_{\nu\nu'}\delta_{\mu\mu'}/n_1 = c_{\nu'\mu'}(z')/n_1$, so that

$$c_{\nu\mu}(z') = n_1 \int_0^{2\pi}\int_0^{\infty} \Phi(r,\varphi,z')\,\Psi^*_{\nu\mu}(r,\varphi,z')\,r\,\mathrm{d}r\,\mathrm{d}\varphi\,, \quad P = \frac{1}{2}\sum_{\nu,\mu} |c_{\nu\mu}(z')|^2. \tag{2.80}$$

The sum of the modal powers $|c_{\nu\mu}|^2/2$ equals the total guided power P because of the mode orthogonality. If a (generally complex) field $\Phi(r,\varphi,z=0) \equiv \Phi(r,\varphi)$ is launched at the fibre entrance plane $z' = 0$, the fraction $\eta_{\nu\mu}$ of the total light source power P_Φ which becomes coupled to guided modes $\Psi_{\nu\mu}(r,\varphi,z=0) \equiv \Psi_{\nu\mu}(r,\varphi)$ is of considerable interest. For this power coupling coefficient we find with the help of Eqs. (2.79), (2.80), (2.68) and $c_{\nu\mu} \equiv c_{\nu\mu}(z'=0)$

$$\eta_{\nu\mu} = \frac{|c_{\nu\mu}|^2}{2P_\Phi} = \frac{\left|\int_0^{2\pi}\int_0^{\infty} \Phi(r,\varphi)\,\Psi_{\nu\mu}(r,\varphi)\,r\,\mathrm{d}r\,\mathrm{d}\varphi\right|^2}{\int_0^{2\pi}\int_0^{\infty} |\Phi(r,\varphi)|^2\,r\,\mathrm{d}r\,\mathrm{d}\varphi \;\int_0^{2\pi}\int_0^{\infty} \Psi^2_{\nu\mu}(r,\varphi)\,r\,\mathrm{d}r\,\mathrm{d}\varphi}\,. \tag{2.81}$$

The reciprocal quantity $\eta^{-1}_{\nu\mu}$ is denoted as power coupling attenuation or insertion loss, and is measured in dB,

$$a_{\eta_{\nu\mu}} = 10\lg \eta^{-1}_{\nu\mu}\,.$$

Only for $\Phi(r,\varphi) = \Psi_{\nu\mu}(r,\varphi)$ the power coupling coefficient $\eta_{\nu\mu}$ reaches the maximum value one. For two orthogonal fields $\Phi(r,\varphi) = \Psi_{\nu'\mu'}(r,\varphi)$, $\nu' \neq \nu$ or $\mu' \neq \mu$ the power coupling is zero, $\eta_{\nu\mu} = 0$. If a Gaussian near-field $\Phi(r,\varphi) = \Phi_0 \exp(-r^2/w_0^2)$ excites a singlemode fibre $\Psi_{01}(r,\varphi) = \Psi_1^{(0)}(r)$, the power coupling η_{01} becomes maximum for a certain beam radius $w_0 = r_G$. The Gaussian *m*ode *f*ield *r*adius (MFR) r_G minimizes the mean square deviation between the Gaussian and the fundamental mode Ψ_{01} so that

$$\int_0^{\infty} \left[\Phi_0 \exp(-r^2/r_G^2) - \Psi_1^{(0)}(r)\right]^2 r\,\mathrm{d}r = \min\,.$$

Chapter 3

Light sources

Among the variety of optical sources, optical fiber communication systems almost always use semiconductor-based light sources such as light-emitting diodes (LED) and laser diodes (LD) because of the several advantages such sources have over the others. These advantages include compact size, high efficiency, required wavelength of emission, and the possibility of direct modulation at high speeds. Besides the references given in the preface, several books covering the topic can be recommended,[1,2,3,4] sorted according to complexity.

Laser is an acronym for *l*ight *a*mplification by *s*timulated *e*mission of *r*adiation. Therefore, our first task is to understand what is meant by stimulated (synonym: induced) emission and under what conditions one can achieve amplification of light by stimulated emission. Laser — the device — may be defined as a highly monochromatic, coherent source of optical radiation. In this sense it is analogous to an electronic oscillator, which is a source of electromagnetic waves in the lower frequency range of the electromagnetic spectrum. The acronym "laser" contains the word "amplification", and obviously the optical amplifier and the laser are as closely related as the "transistor amplifier" and the "transistor oscillator". Historically, the advent of lasers preceded that of optical amplifiers, so the chapter on lasers is placed ahead that of optical amplifiers.

A laser consists of an active medium that is capable of providing optical amplification. This medium may be a collection of microsystems like atoms, molecules, or ions in the solid, liquid, or gaseous form. Placed around the amplifying medium there is an optical resonator that provides the necessary optical feedback, Fig. 3.1. For an optical amplifier, this feedback is sufficiently suppressed in a certain range of the gain. The optical resonator in its simplest form consists of two plane

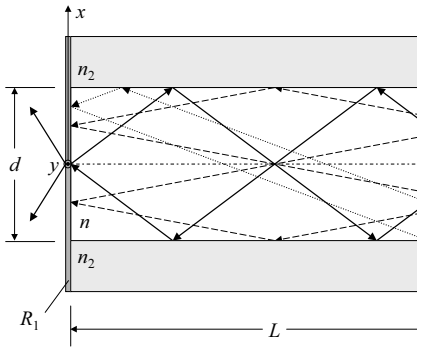


Fig. 3.1. Laser resonator modes. Resonator length L, strip waveguide height d (corresponds to h in Fig. 2.7), strip waveguide height b along y-axis, active volume $V = Lbd$, mirrors with power reflection factors $R_{1,2}$

[1] See Chapter 11 in reference Footnote 25 on Page 39

[2] Hecht, J.: Understanding fiber optics, 4. Ed. Upper Saddle River: Prentice Hall 2002. Chapter 9

[3] Agrawal, G. P.: Lightwave technology. Vol. 1: Components and devices. Hoboken: John Wiley & Sons 2004. Chapter 5

[4] Iizuka, K.: Elements of photonics, Vol. I and II. New York: John Wiley & Sons 2002. Vol. II Chapter 13 and 14

mirrors aligned suitably to confine the optical energy as light propagates back and forth between the mirrors. Such a structure is called a Fabry-Perot[5] resonator[6,7,8] It might consist of a transverse strip waveguide with height d and width b as in Fig. 2.15(a) on Page 33 (formerly, the waveguide height d was denoted as h), and two plane mirrors with power reflection coefficients $R_{1,2}$ at $z = 0, L$. The waveguide has a refractive index n and is surrounded by a cladding with index n_2. With semiconductor lasers it is common to cleave the crystal at $z = 0, L$ perpendicular to the z-axis. For a refractive index $n \approx 3.6$ (GaAs), the power reflection coefficient for perpendicular incidence at such a cleaved plane surface is $R = 32\,\%$ according to Eq. (2.35). Figure 3.1 shows a few closed ray paths for visualizing possible *modes* in a laser cavity. The active volume amounts to $V = dbL$.

Having established the number and density of modes of the electromagnetic field in a finite volume, we first review the basic emission and absorption processes in a microsystem (an atom, a molecule, or an electron in the conduction band of a semiconductor), and then discuss the conditions for light amplification and laser oscillation in a semiconductor.

Number of states Inside the homogeneous resonator medium with refractive index n the wave equation (2.12) is solved by monochromatic homogeneous plane waves with complex amplitudes $\widetilde{\Psi}(k_x, k_y, k_z)$, real angular frequency ω and real propagation vector $\vec{k}$, Eq. (2.19), (2.14). If the components of $\vec{k}$ are fixed, the separation condition $|\vec{k}| = n\omega/c$ determines the frequency, Eq. (2.17). A superposition of such waves defines all possible standing or propagating fields.

The propagation constants into the directions of the coordinates $q = x, y, z$ of Fig. 3.1 are denoted as k_q with $k^2 = \sum_q k_q^2 = (n\omega/c)^2$ from Eq. (2.17). Further, the lengths $L_x = d$, $L_y = b$, $L_z = L$ and the integers $m_{x,y,z} = 0, 1, 2, \ldots$ and $l_q = 0, \pm 1, \pm 2, \ldots$ are introduced for convenience. In addition to the two *transverse* field consistency conditions $2k_{x,y}L_{x,y} = m_{x,y} \times 2\pi$ discussed in Sects. 2.3, 2.4, a third *longitudinal* resonance condition $2k_zL_z = m_z \cdot 2\pi$ fits the modal phase along the z-axis. Obviously, the possible values of $0 \le k_q \le k$ are discrete and describe standing waves or modes,

$$k_q = m_q \delta k_q\,, \quad \delta k_q = \pi/L_q\,, \qquad q = x, y, z\,, \quad m_q = 0, 1, 2, \ldots \tag{3.1}$$

Figure 3.26 on Page 93 displays a typical spectrum of resonator lines. As an example for computing the number m_q of possible laser resonator modes assume the following: A box-shaped active volume with lengths L_q, an amplification half-power bandwidth Δf_H; a modal frequency spacing δf_q, and no dispersion of the refractive index n. From Eq. (3.1) we find the estimate for the mode numbers (see also Eq. (3.68) on Page 79)

$$\delta k_q = 2\pi \frac{n}{c}\,\delta f_q = \frac{\pi}{L_q}\,, \quad \delta f_q = \frac{c}{2nL_q}\,, \quad m_{q\,\max} = \max\left(1, \frac{\Delta f_H}{\delta f_q}\right), \quad M_{\rm tot} = 2\prod_{q=x,y,z} m_{q\,\max}\,. \tag{3.2}$$

The max-function guarantees a modal count of at least 1, $m_q \ge 1$. The total number of modes $M_{\rm tot}$ results from multiplying the maximum mode numbers $m_{q\,\max}$ for all three coordinate directions in 2 polarizations: Each, e. g., longitudinal mode m_z can appear in a number of $m_{x\max} \times m_{y\max}$ varieties of transverse modes, and in 2 orthogonal polarizations. Usually, semiconductor lasers are transversely singlemoded and oscillate in one polarization only, so that $M_{\rm tot} = m_{z\,\max}$.

Because standing waves are made up of *two* counterpropagating fields, we can express the standing wave condition Eq. (3.1) by allowing also negative k_q-values, i. e., by doubling both the

[5] Charles Fabry, French physicist, ⋆ 1867, † 1945. — Alfred Pérot, French physicist, ⋆ 1863, † 1925

[6] Pérot, A. and Fabry, C.: On the application of interference phenomena to the solution of various problems of spectroscopy and metrology. Astrophys. J. 9 (1899), 87–115. http://dx.doi.org/10.1086/140557
Pérot, A. and Fabry, C.: Théorie et applications d'une novelle méthode de spectroscopie interférentielle. Ann. Chim. Phys. 16 (1899) 115–44.

[7] Born, M.; Wolf, E.: Principles of optics, 6. Ed. Oxford: Pergamon Press 1980

[8] See Sect. 9.6.1 Page 368 in reference Footnote 1 on Page 7

k_q-interval to $-k \le k_q \le k$ and the k_q-increments to $\Delta k_q = 2\delta k_q$. For a resonator we have to impose periodic boundary conditions[9] $\exp[-\mathrm{j}\, k_q q] = \exp[-\mathrm{j}\, k_q(q + L_q)]$ for the phase factors of the transverse and longitudinal fields enforcing $k_q L_q = l_q \cdot 2\pi$,

$$k_q = l_q \Delta k_q\,, \quad \Delta k_q = 2\pi/L_q\,, \qquad q = x, y, z\,, \quad l_q = 0, \pm 1, \pm 2, \ldots \tag{3.3}$$

With the separation condition $k^2 = \sum_q k_q^2 = 4\pi^2 \sum_q l_q^2/L_q^2 = (2\pi f n/c)^2$, the points (l_x, l_y, l_z) in l-space lie on a three-axial ellipsoid $\sum_q l_q^2/(L_q f n/c)^2 = 1$. Because both signs are allowed for l_q, the total volume $V_f = \frac{4\pi}{3} \prod_q L_q f n/c$ divided by the elementary volume $(l_x l_y l_z)_{\mathrm{elem}} = 1$ specifies the number of possible l_q-values for running waves with frequencies $0 \ldots f$ and therefore the number of states with a fixed polarization. The total number of states in two orthogonal polarizations each is the integer part of $M_{\mathrm{tot}} = 2V_f$ (correction formulae[10] apply for small state numbers, which is especially true for transversely singlemoded semiconductor lasers),

$$M_{\mathrm{tot}} = 2\,\frac{4\pi}{3} \prod_{q=x,y,z} \frac{L_q f n}{c} = 2\,V_c\,\frac{4\pi}{3}\left(\frac{f n}{c}\right)^3 = 2\,V_c\,\frac{4\pi}{3}\left(\frac{k}{2\pi}\right)^3, \quad V_c = L_x L_y L_z\,. \tag{3.4}$$

Recalling the de Broglie[11] relation between the wavelength $\lambda = 2\pi/k$ of an electromagnetic field and its associated mechanical photon momentum p,

$$p = \hbar k\,, \qquad \hbar = h/(2\pi), \tag{3.5}$$

we see from Eq. (3.4) that $(4\pi/3)[hk/(2\pi)]^3 = (4\pi/3)\,(\hbar k)^3 = (4\pi/3)p^3 = V_p$ represents a volume in momentum space for a maximum momentum $p = \hbar k$ (maximum photon energy hf), while V_c is the volume in configuration space. In a space with g generalized coordinates there are $2g$ degrees of freedom, and the so-called phase space volume is $V_\phi = V_c V_p$. The total number of states M_{tot} with a (polarization) degeneracy of 2 in a $2g$-dimensional phase space volume V_ϕ for photon energies up to hf may therefore be written as

$$\begin{aligned} &M_{\mathrm{tot}} = 2\,V_\phi/h^g, \quad V_\phi = V_c V_p\,; \\ &\quad g = 3\,, \quad V_c = L_x L_y L_z\,, \quad V_p = \frac{4\pi}{3}\left(\frac{\hbar\omega}{c/n}\right)^3 = \frac{4\pi}{3}\,(\hbar k)^3 = \frac{4\pi}{3}\,p^3. \end{aligned} \tag{3.6}$$

The first line of Eq. (3.6) is valid in general,[12] while the second line reflects the special conditions of the problem under investigation.[13]

The density of states $\varrho_{\mathrm{tot}}(f)$ is the state number $\mathrm{d}M_{\mathrm{tot}}$ per frequency interval $\mathrm{d}f$ and configuration volume V_c. The density increases in proportion to f^2 (i. e., to the squared photon energy $(hf)^2$; n_g is the group refractive index, see Eq. (2.21)),

$$\varrho_{\mathrm{tot}}(f) = \frac{1}{V_c}\frac{\mathrm{d}M_{\mathrm{tot}}}{\mathrm{d}f} = \frac{8\pi}{c^3}(f n)^2 n_g\,, \quad n_g = n + f\,\frac{\mathrm{d}n}{\mathrm{d}f} = n - \lambda\,\frac{\mathrm{d}n}{\mathrm{d}\lambda}\,. \tag{3.7}$$

Again, this formula must not be applied uncritically in the case of transversally singlemoded semiconductor lasers, see the remarks before Eq. (3.4).

---

[9] Grau, G.; Freude, W.: Optische Nachrichtentechnik, 3. Ed. Berlin: Springer Verlag 1991. Since 1997 out of print. Corrected reprint from University Karlsruhe 2005, available in electronic form via W. F. (`W.Freude@kit.ed`). Sect. 2.1.7 Eqs. (2.40), (2.41)

[10] Weyl, H.: Gesammelte Abhandlungen (1911–1915), Band I, Abhdlg. Nr. 13, 16–19, 22. Ed.: K. Chandraskharan. Berlin: Springer Verlag 1986. Cited after: Baltes, H. P.; Hilf, E. R.: Spectra of finite systems. Mannheim: B.I.-Wissenschaftsverlag 1976

[11] See Footnote 3 on Page 1

[12] See Sect. 2.1.7 Eqs. (2.47), (2.48) and Sect. 3.2.1 Eq. (3.10) in reference Footnote 9 on Page 49

[13] See reference Footnote 10 on Page 49

3.1 Luminescence and laser radiation

The excitation energy W_2 of any microsystem may be released by a transition to a state of lower energy W_1. This transition can be radiative by emission of a photon with energy $hf = W_2 - W_1$ (Planck's constant[14] h, frequency f), or nonradiative. The transition probability depends on the quantum mechanical properties of the microsystem, and on the interaction with, e. g., an electromagnetic field. The microsystem may also gain energy making an upwards (radiative)

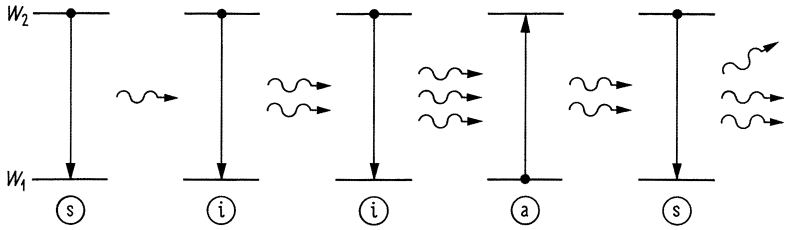


Fig. 3.2. Interaction of a two-level microsystem with electromagnetic radiation, photon energy $hf = W_2 - W_1$. (a) absorption, (s) spontaneous emission, and (i) induced (= stimulated) emission of photons

transition by absorbing an amount of electromagnetic energy $hf = W_2 - W_1$. The released photon energy is emitted into any mode of the electromagnetic field. For our present purpose the term "mode" may be associated with the resonating modes in the active volume V_R. Nonradiative transitions transfer the same amount of energy to or from phonons (thermal vibrations of the crystal lattice) or other degrees of freedom of the interacting substances. Referring to Fig. 3.2, there are three types of interactions:

Absorption A microsystem in its ground state W_1 can absorb radiation at a frequency $f = (W_2 - W_1)/h$ and make an upward transition to the higher energy level W_2. This absorption process is therefore *induced* or *stimulated* by an existing electromagnetic field. The absorption rate depends on the electromagnetic energy density, and on the number of microsystems in the ground state, Fig. 3.2(a).

Spontaneous emission An excited microsystem in level W_2 can make a downward transition to the ground state W_1 "spontaneously" (apparently without any interaction) by emitting a photon with energy $hf = W_2 - W_1$, Fig. 3.2(s). The spontaneous emission rate depends only on the number of excited microsystems.

The term "spontaneous" needs more explanation. In a *semiclassical theory*, the microsystem is treated as a quantum mechanical system, while the field description is classical. As a result, an excited microsystem would stay in its excited state W_2 for an infinite time period.

Experimentally it is observed that microsystems release their energy at random after a certain average spontaneous lifetime τ_{sp}. In *quantum electrodynamics* the particle (photon) nature of radiation is formulated by a quantization of the electromagnetic field. The outcome is that the electric and the magnetic fields $\vec{E}$ and $\vec{H}$ are connected by an uncertainty relation. Therefore, the simultaneous states $\vec{E}(t,\vec{r}) = 0$ and $\vec{H}(t,\vec{r}) = 0$ for all $t, \vec{r}$ are impossible. However, the quantum electrodynamical vacuum, as defined by the expected values $\overline{\vec{E}(t,\vec{r})} = 0$, $\overline{\vec{H}(t,\vec{r})} = 0$ for all $t, \vec{r}$, *is* allowed. For this case, the average energy density $\overline{\epsilon_0 \vec{E}^2/2 + \mu_0 \vec{H}^2/2}$ is finite such that the total mean energy in each state of the electromagnetic field with a certain polarization amounts to $hf/2$ (zero point energy).

This energy cannot be extracted from the system (it cannot be used to fry eggs), but the fields fluctuating around the expectation zero represent a perturbation for an excited microsystem, and may therefore *induce random transitions* to the ground state. These "spontaneously"

[14] See Footnote 2 on Page 1

emitted photons will be found with equal probability in any possible mode of the electromagnetic field, because all modes possess the same zero point energy $hf/2$. A spontaneously emitted photon modifies an already existing field in a mode by superimposing an additional field with a random phase thus establishing a *noise* signal. Incoherent radiation of this type is called luminescence.

Spontaneous absorption (absorption induced by the zero point field fluctuation) is not an allowed process, because the zero point energy cannot be extracted from the electromagnetic vacuum and therefore not be transferred to a microsystem.

Induced emission A microsystem in an excited level W_2 can also make a downward transition to the ground state W_1 in the presence and induced (synonym: stimulated) by an external radiation of frequency $f = (W_2 - W_1)/h$. As in the case of (induced) absorption, the emission rate depends on the electromagnetic energy density, and on the number of microsystems in the excited state, Fig. 3.2(i). In contrast to spontaneous emission processes (= transitions induced by zero point fluctuations) the emitted radiation is phase coherent with the stimulating radiation. Therefore, the induced radiation adds with the same polarization and phase to the stimulating field and becomes amplified much like by an electronic amplifier.

3.1.1 Lifetime and linewidth

As explained above, spontaneous absorption is impossible. Further, a microsystem in its ground state does not possess the energy to radiate a photon. Therefore, the lifetime of the ground state is infinite. From quantum theoretical considerations the system energy may be determined inside the observation time τ_2 with an uncertainty ΔW_2 ($\hbar = h/(2\pi)$),

$$\Delta W_2 \, \tau_2 \geq \hbar/2 \,, \quad \Delta f_2 \geq 1/(4\pi\tau_2) \,. \tag{3.8}$$

If the excited state energy is not exactly known because of the finite lifetime $\tau_2 \leq \tau_{\rm sp}$ (by induced emission the excited state lifetime may become smaller than $\tau_{\rm sp}$), the spontaneously emitted photon energy hf with expectation $\overline{hf_0} = \overline{W_2 - W_1}$ is uncertain by $\Delta W_2 \geq \frac{1}{2}\hbar/\tau_{\rm sp}$. Therefore, the probability density of the spontaneous emission (luminescence spectrum) has a lineshape $\rho(f)$ with a maximum at $f = f_0$ and a linewidth of $\Delta f \sim 1/\tau_{\rm sp}$.

3.1.2 Laser action

In the presence of external radiation, the stimulated absorption probability per microsystem is the same as the stimulated emission probability. A net emission is possible if the number of excited state microsystems exceeds the number of ground state systems. This does not correspond to the normal population distribution in thermal equilibrium, where the number of ground state microsystems is larger than the number of excited systems, and is therefore called *population inversion*. Such an inversion has to be provided by some kind of a "pump" mechanism, and is a prerequisite for light amplification. For concentrating next to all of the emitted photons in a narrowband spectral range smaller than $\Delta f_2 = 1/(4\pi\tau_{\rm sp})$, a resonance structure as in Fig. 3.1 provides the necessary means.

Amplification and oscillation The modes of an optical resonator having a lossless transverse guiding structure and partially transmitting mirrors at $z = 0, L$ are characterized at a certain light resonance frequency f_S by a quality factor $Q = f_S/\Delta f_S$, which defines a resonator bandwidth Δf_S and a photon lifetime $\tau_P \sim 1/\Delta f_S$, which is caused mainly by transmission losses of the mirrors. If the resonator volume V_R contains excited microsystems, the spontaneously emitted photons are collected by all resonator modes specified by k_q from Eq. (3.3). If the maximum of the spontaneous emission line $\rho(f_0)$ is centred at a resonator mode $f_S = f_0$, this special mode collects a larger number N_P of photons than other modes. Because the induced emission is in proportion to N_P, the emission probability increases with N_P. For a population inversion condition the induced

absorptions are less than the stimulated emissions, and a net stimulated emission rate results causing a coherent amplification of the light in mode f_S.

With increasing pump rate the gain becomes higher. When the resonator losses are just compensated, the so-called threshold of oscillation is reached. With increasing pump rate, the photon number increases at first exponentially, and so does the probability $1/\tau_2$ of stimulated emissions per second. Each additionally excited microsystem releases its energy practically immediately after an effective lifetime $1/\tau_{2\,\mathrm{eff}} = 1/\tau_2 + 1/\tau_{\mathrm{sp}}$, and the probability $1/\tau_{\mathrm{sp}}$ of spontaneous emissions per second into this mode becomes less and less important, $1/\tau_{2\,\mathrm{eff}} \approx 1/\tau_2$.

Because spontaneous emission is reduced compared to induced emission, the field becomes more coherent. The number of photons N_P in the dominant resonator mode stabilizes at such a high stationary level N_{P0} that practically all microsystems, which are excited additionally by the pump, release their energy immediately by induced (coherent) emission, and the gain gets clamped at the threshold level. The stimulated-emission photons compensate the total resonator losses. A light field develops having a near-sinusoidal time dependence $E(t) = A(t)\cos[\omega_L t + \varphi(t)]$, where amplitude $A(t)$ and phase $\varphi(t)$ vary very slowly on the scale of an optical period $1/f_S$ (e. g., for a semiconductor laser $f_S = 193\,\mathrm{THz}$ at $\lambda = 1.55\,\mu\mathrm{m}$, $1/f_S = 5.2\,\mathrm{fs}$). This leads to a very narrow spectral line width Δf_S. For a so-called *d*istributed *f*eed-*b*ack (DFB) semiconductor laser a typical value is $\Delta f_S = 40\,\mathrm{MHz}$.

If there was a momentary increase in photon number $N_P > N_{P0}$, the increased stimulated emission would deplete the population inversion, and the gain would decrease followed by a reduction of the photon number to $N_P < N_{P0}$. This being the case, stimulated emission is below its stationary value, the pump rebuilds the inversion, and a so-called relaxation oscillation of photon number and inversion (gain) is to be expected. This corresponds to an energy exchange between two energy reservoirs, like with inductor and capacitor in a resonant circuit. For semiconductor lasers, relaxation oscillations of the optical intensity occur at microwave frequencies in the order of $f_r = 1 \ldots 30\,\mathrm{GHz}$.

If there were no longitudinal resonator ($R_{1,2} = 0$ in Fig. 3.1), a wave travelling through the active medium would be amplified. Residual mirror reflectivities $R_1, R_2 \neq 0$ could lead to a regenerative oscillation and have to be avoided in this case.

Modulation By changing the pump rate, the population inversion may be modified. If the light source has no optical resonator and therefore emits only spontaneous radiation, the number of excited microsystems cannot decrease faster than their lifetime τ_{sp}, and the maximum light intensity modulation frequency is in the order of $1/\tau_{\mathrm{sp}}$ if nonradiative transitions are excluded. When the lifetime τ_{sp} is reduced by additional nonradiative processes, $1/\tau_{2\,\mathrm{eff}} = 1/\tau_{\mathrm{nr}} + 1/\tau_{\mathrm{sp}}$, the modulation corner frequency increases to the order of $1/\tau_{2\,\mathrm{eff}}$, but the radiation efficiency with respect to the pump rate decreases.

In a laser above threshold, the effective lifetime $1/\tau_{2\,\mathrm{eff}} = 1/\tau_2 + 1/\tau_{\mathrm{sp}} \gg 1/\tau_{\mathrm{sp}}$ of the excited microsystems can be made very much smaller than the spontaneous lifetime τ_{sp} by the mechanism of stimulated emission, without paying the prize of a reduced radiation efficiency.

Coupling efficiency Laser light may me concentrated in only *one* transverse resonator mode $\Phi(r,\varphi)$. Using a lossless optical transformation $\Phi(r,\varphi) \to \Phi'(r,\varphi) = \Psi_{01}(r,\varphi)$ (e. g., with a lens system) the laser power can be completely transferred to a singlemode fibre mode $\Psi_{01}(r,\varphi)$ so that the power coupling efficiency as given by Eq. (2.81) on Page 45 becomes unity,

$$\eta_{01} = \frac{\left|\int_0^{2\pi}\int_0^{\infty} \Phi'(r,\varphi)\,\Psi_{01}(r,\varphi)\,r\,\mathrm{d}r\,\mathrm{d}\varphi\right|^2}{\int_0^{2\pi}\int_0^{\infty} |\Phi'(r,\varphi)|^2\,r\,\mathrm{d}r\,\mathrm{d}\varphi \;\int_0^{2\pi}\int_0^{\infty} \Psi_{01}^2(r,\varphi)\,r\,\mathrm{d}r\,\mathrm{d}\varphi} = \{\Phi'(r,\varphi) = \Psi_{01}(r,\varphi)\} = 1\,. \quad (3.9)$$

Luminescent light in a bandwidth Δf is emitted in *all* $M_{\mathrm{tot}} = V\int_{-\Delta f/2}^{+\Delta f/2} \varrho_{\mathrm{tot}}(f)\,\mathrm{d}f$ transverse and longitudinal resonator modes (Eq. (3.7)). However, it is impossible by principle[15] to transform the

[15] See Sect. 2.1.7 on Page 32 and Eq. (2.43) in reference Footnote 9 on Page 49

field $\Phi(r, \varphi)$ resulting from the superposition of independent transverse resonator modes into a single transverse fibre mode.

Noise The noise properties of a luminescent device and a laser are completely different. Spontaneous emission represents a noise signal of the electromagnetic field with an expectation of zero and with a non-zero intensity, very much like thermal noise from a resistor (e. g., an incandescent lamp). On the other hand, a laser signal resembles a sinusoidal field perturbed by the noise of spontaneous emissions. The amplitude fluctuations are relatively small because of the nonlinear amplitude control described in Sect. 3.1.2 on Page 52 (gain clamping). The magnitude of the phase (or frequency) fluctuation depends mainly on the resonator bandwidth.

3.1.3 Laser active materials

For a microsystem in thermal equilibrium the occupation probabilities $w(W_i)$ of the various energy levels W_i at any absolute temperature T are given by the Maxwell-Boltzmann statistics (ground state W_1, degeneracy g_i of level W_i, Boltzmann[16] constant $k = 1.380\,658 \times 10^{-23}\,\mathrm{Ws\,/\,K}$)

$$w(W_i) = g_i\,\mathrm{e}^{-W_i/(kT)} \Big/ \sum\nolimits_i g_i\,\mathrm{e}^{-W_i/(kT)}\ . \tag{3.10}$$

Two-level systems

For nondegenerate two-level microsystems as in Fig. 3.2, the population numbers $N_{1,2}$ of microsystem with energy states $W_{1,2}$ in thermal equilibrium are related as

$$\frac{N_2}{N_1} = \mathrm{e}^{-(W_2-W_1)/(kT)}, \qquad N = N_1 + N_2\,. \tag{3.11}$$

N is the total number of microsystems. In thermal equilibrium the excited state is less densely populated than the ground state by an exponential factor depending on the difference energy $hf = W_2 - W_1$ with respect to the thermal energy kT. With induced absorption as described in Fig. 3.2(a) the population number N_2 may be increased in proportion to the photon number N_P which is available in a resonator mode of frequency f, and in proportion to the time t. On the other hand, spontaneous emission reduces N_2 in proportion to t, and stimulated emission diminishes N_2 in proportion to N_P and t. Therefore, in the presence of an electromagnetic field of photon energy hf a dynamic equilibrium will be reached for which the number of spontaneous and induced emissions equals the number of stimulated absorptions. If N_P is so large that spontaneous emission may be neglected, an dynamic equilibrium state $N_2 = N_1$ (with spontaneous emission: $N_2 \leq N_1$) may be reached so that the number of stimulated emissions equals the number of stimulated absorptions. The medium is called transparent in this case. However, with a strict two-level system it is impossible to achieve a gain by population inversion.

Three-level systems

The situation is improved with a three-level system, Fig. 3.3(a). If we pump the system at an absorption frequency $f^{(\mathrm{a})} = (W_3 - W_1)/h$, microsystems can get excited from level W_1 to level W_3, where in the nondegenerate case the occupation numbers are related by $N_3 \leq N_1$. The excited microsystems make a downward transition (releasing their energy radiatively or noradiatively) also to level W_2. If for a given pumping rate the transition rate $1/\tau_{32}$ for $W_3 \to W_2$ exceeds that of the transition rates $1/\tau_{31}$ for $W_3 \to W_1$ and $1/\tau_{21}$ for $W_2 \to W_1$, the microsystems in energy level W_3 deplete (thereby reducing the occupation number of the ground state $N_1 \approx N_3$), and accumulate in energy level W_2, so that a population inversion with $N_2 > N_1$ and an associated net gain for the signal emission frequency $f^{(\mathrm{e})} = (W_2 - W_1)/h$ becomes possible. For reaching the gain threshold a

[16] Ludwig Boltzmann, Austrian physicist, ⋆ Wien 20.2.1844, † Duino (Duino-Aurisina, near Trieste) 5.9.1906 (suicide). Professor in Graz, Wien, München, Leipzig

very high pump rate is necessary because the occupation number of the ground state is very high, Eq. (3.11). The pump efficiency cannot be larger than $hf^{(e)}/(hf^{(a)})$,

$$\eta_p \leq \frac{W_2 - W_1}{W_3 - W_1} = \frac{hf^{(e)}}{hf^{(a)}} \,. \tag{3.12}$$

Practically, η_p is much smaller, because not all pump photons excite microsystems with an energy W_3, and not all excited systems end up in level W_2.

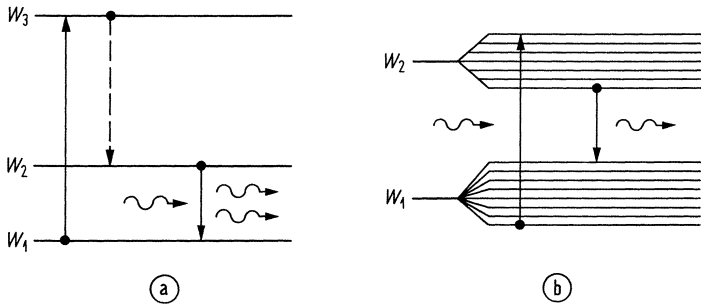


Fig. 3.3. Pump mechanism using energy levels (a) outside (three-level laser system) or (b) inside the energy level group of the laser transition (pseudo-four-level laser system)

Four-level systems and semiconductors

A more efficient pumping scheme can be realized by a four-level system with pump levels $W_{0,3}$ and laser levels $W_{2,1}$ because the final state W_1 for the lasing transition is different from the densely populated ground state W_0. A pseudo-four-level scheme is depicted in Fig. 3.3(b). The lasing levels $W_{2,1}$ split up into closely neighbouring sublevels. According to the equilibrium distribution Eq. (3.11), the occupation probability of the lowest energy level ($\widehat{=} W_0$) is highest, and of the highest energy levels ($\widehat{=} W_3$) lowest, so that absorption from the lowest energy states to the highest ones is a most probable process. On the other hand, emission from a strongly populated level ($\widehat{=} W_2$) to a sparsely populated level ($\widehat{=} W_1$) is very probable. Therefore, the maximum for absorption is found at higher frequencies (shorter wavelengths) than the maximum for luminescence, and effective pumping may be achieved at a slightly shorter wavelength then the lasing emission wavelength. This mechanism can be used in Erbium-doped fibre amplifiers (EDFA), where an absorbed pump power at $\lambda^{(a)} = 1.48\,\mu$m produces an optical gain at the emission signal wavelength $\lambda^{(e)} = 1.53\,\mu$m. The maximum pump efficiency is $\eta_p = 1.48\,\mu\text{m}\,/1.53\,\mu\text{m} = 97\,\%$.

The scheme of Fig. 3.3(b) may be also applied to the case of a semiconductor device. Levels W_2 and W_1 are to be associated with conduction and valence band states, respectively. Pump light with a photon energy $hf^{(a)}$ ($\widehat{=} W_3 - W_0$) can be absorbed for producing electron-hole pairs in the appropriate energy levels. This could be also achieved with a forward biased semiconductor pn-diode by injecting electrons and holes into the conduction and valence band, respectively, to achieve a population inversion. For a temperature $T = 0$ the "pump energy" $eU = hf^{(a)}$ given by the forward voltage U (elementary charge $e = 1.602 \times 10^{-19}\,$C), would define the energetic difference at which electrons and holes could be injected, i. e., the difference of the quasi Fermi[17] levels $W_{Fn} - W_{Fp} = eU = hf^{(a)}$ for electrons in the conduction band (W_{Fn}) and for holes in the valence band (W_{Fp}), respectively. The energy $hf^{(e)}$ of the emitted photons is therefore smaller than $W_{Fn} - W_{Fp}$, but necessarily larger than the bandgap W_G,

$$W_G < hf^{(e)}, \qquad hf^{(e)} \leq hf^{(a)}, \qquad hf^{(a)} = W_{Fn} - W_{Fp} = eU \,. \tag{3.13}$$

[17] Enrico Fermi, Italian physicist, ⋆ Rome 29.9.1901, † Chicago (Illinois) 28.11.1954. Professor in Rome and later in USA. Nobel prize 1938

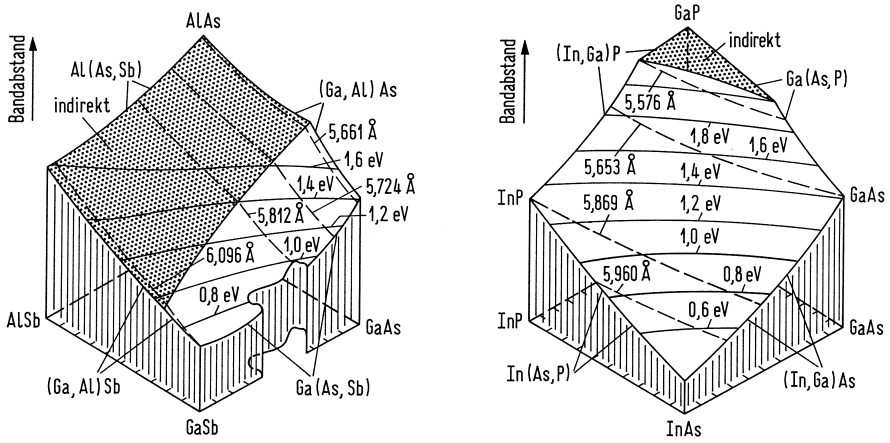


Fig. 3.4. Material systems $(Ga_{1-x}Al_x)(As_ySb_{1-y})$ and $(In_{1-x}Ga_x)(As_yP_{1-y})$, bandgaps and lattice constants. Dotted region: indirect semiconductor (Bandabstand $\widehat{=}$ bandgap)

Semiconductor	$W_G/\,\mathrm{eV}$ $(\lambda_G/\,\mu\mathrm{m})$	n at λ_G	$a/\mathrm{Å}$
GaSb, direct	0.726 (1.708)	3.82	6.096
GaAs, direct	1.424 (0.871)	3.655	5.653
AlSb, indirect	1.58 (0.785)	3.4	6.135
AlAs, indirect	2.163 (0.573)	3.178	5.660
$(Ga_{1-x}Al_x)As$ direct: $x \leq 0.3$	$1.424 + 1.247\,x$ $1.424 \ldots 1.798$ $(0.871 \ldots 0.69)$	$3.59 - 0.71\,x +$ $+0.091\,x^2$ (at $\lambda = 0.9\,\mu\mathrm{m}$)	$5.653 + 0.027\,x$
$(Ga_{1-x}Al_x)(As_ySb_{1-y})$ lattice-matched to GaSb direct: $x \leq 0.24$ $y = x/1.11$	$0.726 + 0.834\,x +$ $+1.134\,x^2$ $0.726 \ldots 0.991$ $(1.708 \ldots 1.25)$	?	6.096

Table 3.1. Material system $(Ga_{1-x}Al_x)(As_ySb_{1-y})$. W_G bandgap, $\lambda_G = hc/W_G$ bandgap wavelength, n refractive index, a lattice constant

Such a laser diode could be also used as a photodetector for photon energies $hf^{(\mathrm{a})} > W_G$. Each photo-generated electron-hole pair is separated by the field inside the pn-junction and induces an induction current in the external circuit with a time integral e.

3.1.4 Compound semiconductors

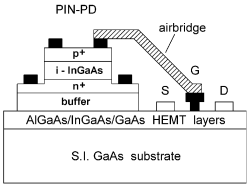
In Sect. 3.1.3 we had discussed which properties a material should have for the generation and amplification of light. Here, we specify important properties of the III-V compound semiconductors (Ga,Al)(As,Sb) and (In,Ga)(As,P). The bandgap W_G (and hence the bandgap wavelength λ_G and the refractive index n) depend on the composition. The lattice constant may be chosen to match the lattice constant of a binary substrate semiconductor. Lattice matching is very important for several reasons:

- A close lattice match is necessary in order to grow high-quality crystal layers.
- Excess lattice mismatch between the heterostructure layers results in crystalline imperfections which lead to nonradiative recombination and thus prevent lasing.
- Lattice mismatch causes degradation in devices during operation.

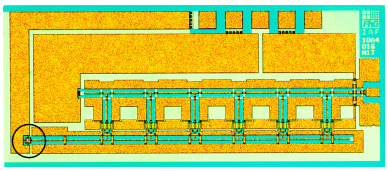
Elemental semiconductors as Si and Ge have a diamond structure, while compound semiconductors as GaAs or InP have a zinc-blende structure. Having no inversion centre, the crystals of the zinc-blende type show a linear electro-optic effect, so these substances my be also used to construct modulators and switches.

Semiconductor	W_G/ eV (λ_G/ μm)	n at λ_G	a/Å
InAs, direct	0.36 (3.444)	3.52	6.058
InP, direct	1.35 (0.918)	3.45	5.869
GaAs, direct	1.424 (0.871)	3.655	5.653
GaP, indirect	2.261 (0.548)	3.452	5.451
$(In_{0.49}Ga_{0.51})P$, direct lattice-matched to GaAs	1.833 (0.676)	3.451 ?	5.653
$(In_{0.53}Ga_{0.47})As$, direct lattice-matched to InP	0.75 (1.653)	3.61	5.869
$(In_{1-x}Ga_x)(As_yP_{1-y})$ lattice-matched to InP direct: $y \leq 1$ $x = y/(2.2091 - 0.06864\,y)$	$1.35 - 0.72\,y + 0.12\,y^2$ $1.35 \ldots 0.75$ $(0.918 \ldots 1.653)$	$3.45 + 0.256\,y - 0.095\,y^2$ $3.45 \ldots 3.61$	5.869

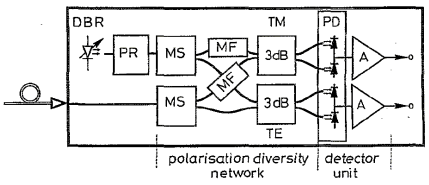
Table 3.2. Material system $(In_{1-x}Ga_x)(As_yP_{1-y})$. W_G bandgap, $\lambda_G = hc/W_G$ bandgap wavelength, n refractive index, a lattice constant



(a) InGaAs PIN photodiode on GaAs. Air bridge to HEMT amplifier[18]



(b) Photodiode (◯) with 30 Ω coplanar transmission lines and 6 interconnecting cascoded HEMT[18]



(c) Polarization diversity heterodyne receiver integrated on InP substrate with local laser, couplers, detectors, and electronics[19]

Fig. 3.5. Advanced OEIC designs. (a) Top-illuminated $(In_{0.53}Ga_{0.47})As$ pin photodiode on a semi-insulating (S.I.) GaAs substrate. Diameter of light-sensitve area 10 μm, sensitivity $S = 0.4\,A/W$, dark current 9 nA at -2 V bias (b) 40 Gbit/s, 6-stage travelling wave amplifier, HEMT gate length 0.15 μm. Processing time 3 months, chip size 2.5 mm ×1 mm on a 3 in wafer (c) Optical polarization diversity heterodyne receiver with tunable local distributed Bragg reflector laser (DBR), polarization rotator (PR), mode splitters (MS), mode filters (MF), 3 dB couplers, photodiodes (PD), and electronic amplifiers (A) consisting of JFET and load resistors. Processing time 5 months, 25 mask steps, 7 epitaxy steps, 170 main processing steps, 2 in wafer, 100 receiver chips 9 mm ×0.6 mm, carrying 17 sub-components, yield > 50 %

Figure 3.4 shows the bandgap W_G and the lattice constant a for two compound material systems. Tables 3.1 and 3.2 summarize the numerical values. With ternary compound crystals (see Table 3.1, Fig. 3.4), active $(Ga_{1-x}Al_x)As$ layers slightly mismatched to a GaAs substrate may be grown for laser diode emission wavelengths $\lambda = 0.69 \ldots 0.87\,\mu$m. Using quaternary compound crystals $(Ga_{1-x}Al_x)(As_ySb_{1-y})$ lattice-matched to a GaSb-substrate, laser diodes can befabricated emitting at $\lambda = 1.25 \ldots 1.71\,\mu$m; this material is also well suited for long-wavelength detectors. For photodetectors, indirect semiconductors are applicable, and lattice-matched compound crystals on GaSb may be grown, leading to an absorption energy $W_G = 0.726 \ldots 1.6$ eV, $\lambda_G = 1.71 \ldots 0.78\,\mu$m. There is an miscibility gap of unknown extent for compound crystals with similar concentrations of As and Sb.

With the material system $(In_{1-x}Ga_x)(As_yP_{1-y})$ laser diodes and photodiodes are grown on lattice-matched InP substrates ($\lambda = 0.92 \ldots 1.65\,\mu$m). With GaAs substrates, emission wavelength in the region $\lambda = 0.87\,\mu$m (GaAs) down to $\lambda = 0.68\,\mu$m $(In_{0.49}Ga_{0.51}P)$ become possible.

High-quality GaAs substrates are available with relatively large wafer diameters of 3 in (8 cm) and 4 in (10 cm). GaAs-based integrated circuits are a standard technique, and so the construction of optoelectronic integrated circuits with (In,Ga)(As,P), lattice-mismatched to a GaAs substrate, $\lambda = 1.3\,\mu$m $\widehat{=}$ 0.95 eV could mature to become a cost-saving alternative. Much more expensive is the processing of (In,Ga)(As,P) on typically 2 in (5 cm) InP substrates, $\lambda = 1.55\,\mu$m $\widehat{=}$ 0.8 eV. Figure 3.5 shows two examples of *o*pto-*e*lectronic *i*ntegrated *c*ircuits (OEIC):

GaAs Fig. 3.5(a),(b). Hybridly integrated 40 Gbit / s pin HEMT receiver with $(In_{0.53}Ga_{0.47})As$ pin diode, diameter 10 μm, sensitivity $S = 0.4\,A/W$, dark current 9 nA at -2 V bias, 4- to 6-stage travelling wave amplifier, HEMT gate length 0.15 μm. Processing time 3 months,

chip size 2.5 mm ×1 mm on a 3-in wafer.[18]

InP Fig. 3.5(c). Monolithically integrated optical heterodyne receiver with tunable local distributed feedback (DFB) laser, polarization diversity reception and detector unit. Processing time 5 months, 25 mask steps, 7 epitaxy steps, 170 main processing steps, 100 receiver chips 0.6 mm ×9 mm with 17 components each per 2-in wafer, yield > 50 %.[19,20]

---

[18]Hurm, V; Benz, W.; Bronner, W.; Hülsmann, A.; Jakobus, T.; Köhler, K.; Leven, A.; Ludwig, M.; Raynor, B.; Rosenzweig, J.; Schlechtweg, M.; Thiede, A.: 40 Gbit/s 1.55 μm monolithic integrated GaAs-based PIN-HEMT photoreceiver. Proc. 24th Europ. Conf. Opt. Commun. Madrid (ECOC'98), Vol. 3, 121–123, 1998

[19]Kaiser, R.; Trommer, D.; Fidorra, F.; Heidrich, H.; Malchow, S.; Franke, D.; Passenberg, W.; Rehbein, W.; Schroeter-Janssen, H.; Stenzel, R.; Unterbörsch, G.: Monolithically integrated polarisation diversity heterodyne receivers on GaInAsP/InP. Electron. Lett. 30 (1994) 1446–1447

[20]Final Report on the Joint Research Program „Optical Signal Processing" funded by the Ministry of Research and Technology (BMFT) in the framework of the focus program PHOTONIK (4/90–3/94). July (1994) D.4-1–D.4-7

3.2 Semiconductor physics

The simplest laser diode structure is a pn-homojunction biased with a forward current I, Fig. 3.6. Spontaneously emitted light leaves the active layer in all possible directions. The field is guided in the active region strip waveguide and reflected from the cleaved end facets at $z = 0, L$, which form a Fabry-Perot laser. The other surfaces at $x = \pm d/2$ and $y = \pm b/2$ are saw-cut to prevent parasitic resonances. To understand the device properties, we have to recall some semiconductor basics in the following.

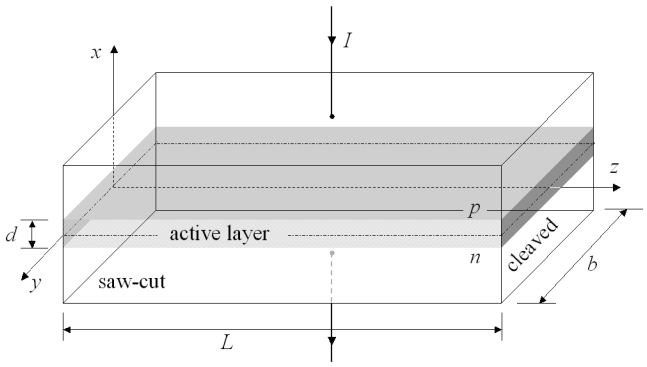


Fig. 3.6. Forward biased semiconductor pn-homojunction acting as a laser diode. Side-walls are saw-cut, the end facets are cleaved. Typical dimensions: $d = 0.1 \ldots 0.2\,\mu$m (active layer), $b = 3 \ldots 6\,\mu$m, $L = 200 \ldots 600\,\mu$m

3.2.1 Energy bands and density of states

Let us consider a linear crystal consisting of a chain of N equal atoms spaced a lattice constant a apart with a length $L = Na$. Each isolated atom contributes a single bound electron with energy levels W_i'. If the N atoms interact, the N electrons do not remain bound to a fixed atom, but belong collectively to the crystal and assume N energy states $W_{i\mu}$ near W_i'. This splitting of energies is in analogy to the coupling of identical resonance circuits to form a bandpass filter. The probability $w_{i\mu}(\vec{r})$ to find an electron of energy $W_{i\mu}$ at a position $\vec{r}$ is given by the modulus squared of the quantum mechanical wave function $\Psi_{i\mu}(\vec{r})$ (probability density amplitude, Schrödinger function),

$$\Psi_{i\mu}(x) = \frac{1}{\sqrt{L}}\, u_i(k_\mu, x)\, \mathrm{e}^{\,\mathrm{j}\,k_\mu x}, \qquad u_i(k_\mu, x) = u_i(k_\mu, x + a), \qquad k_\mu = \mu\,\frac{2\pi}{Na}\,, \tag{3.14}$$

$$\int_0^L |\Psi_{i\mu}(x)|^2\,\mathrm{d}x = 1, \qquad W_{i\mu} = W_i(k_\mu), \qquad N \text{ values for } \mu = 0, \pm1, \pm2, \ldots, (N-1)/2\,.$$

The functions differ from each other by a parameter $k_\mu = \mu \cdot 2\pi/L$ having N discrete values. This is analogous to the number of longitudinal modes in a resonator, see Eq. (3.3). The higher the electron energy becomes, the less it is influenced by the periodic atomic potentials, and the lattice-periodic function $u_i(k_\mu, x)$ approaches asymptotically one. A free electron of mass m moving in a constant potential W_0 has an energy $W_0 + p_\mu^2/2m$ given by its mechanical momentum p_μ. It may be described by a probability density wave with a de Broglie wavelength $\lambda_\mu = h/p_\mu$, so that

$$k_\mu = \frac{2\pi}{\lambda_\mu} = \frac{2\pi}{h}\,p_\mu = \frac{p_\mu}{\hbar}, \quad \hbar k_\mu = p_\mu, \quad W = W_0 + \frac{p_\mu^2}{2m} = W_0 + \frac{\hbar^2 k_\mu^2}{2m}\,. \tag{3.15}$$

For a free electron, the product $\hbar k_\mu$ denotes the mechanical momentum p_μ, which justifies the plane-wave ansatz Eq. (3.14). For crystal electrons the quantity $\hbar k_\mu$ cannot be interpreted as an electron momentum, but in interactions with photons it represents an invariant together with the photon momentum $p = h/\lambda = \hbar k$ for a field of wavelength λ and propagation constant k. Therefore

$\hbar k_\mu$ is called the crystal or pseudo momentum. The number of possible mutual exclusive electron energy states is $2N$ regarding the spin degeneracy. The N energy eigenvalues $W_i(k_\mu)$ represent the bandstructure of the band i, which resulted from the level W_i' of the isolated atom. The function $W_i(k_\mu)$ is periodic in k_μ with a period $2\pi/a$ and may be therefore restricted to a region $-\pi/a < k_\mu \leq \pi/a$ (first Brillouin zone). The bandstructure Fig. 3.7 has the symmetry property $W_i(k_\mu) = W_i(-k_\mu)$ and possesses extrema at the borders of the Brillouin zone. The topmost band

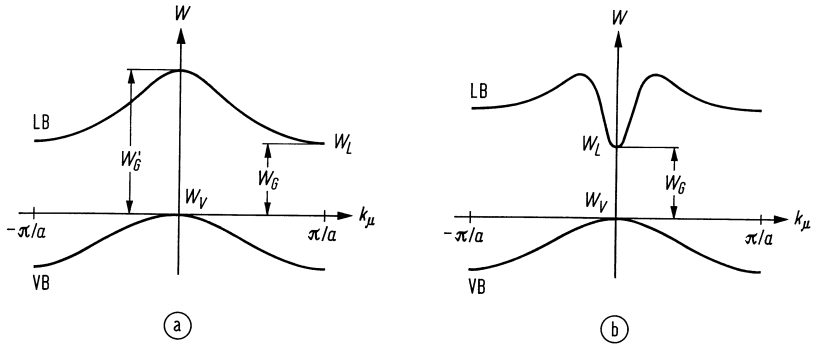


Fig. 3.7. Bandstructures of conduction band (CB, "Leitungsband" LB) and valence band (VB) of an (a) indirect semiconductor and of a (b) direct semiconductor. Minimum bandgap energy W_G, bandgap energy W_G' for an indirect semiconductor at $k_\mu = 0$

which is fully occupied at $T = 0$ is called valence band (VB), the lowest band which is empty for $T = 0$ is called conduction band (CB, "Leitungsband" LB). The difference between the lowest energy level of the CB and the highest energy state of the VB is denoted as bandgap energy $W_G = W_C - W_V$. Because of thermal excitation at $T > 0$ the lowest CB states are occupied while the highest VB states are empty. For a transition $W_C \to W_V$ in an indirect semiconductor the crystal momentum changes by an amount of $\hbar k_\mu = \hbar\pi/a$, which cannot be transferred to the emitted photon with momentum $\hbar k = \hbar \cdot 2\pi/\lambda$ because the lattice constant is much smaller than the wavelength, $a \ll \lambda$. So a phonon as a third interaction particle with sufficient momentum is necessary, but this tree-particle scattering is less probable than a two-particle interaction. Therefore the emission (and the absorption) of photons at energies slightly larger than $hf = W_G$ is a very unlikely process.

The elemental semiconductors Ge and Si are indirect semiconductors and therefore unsuitable for efficient light sources, Fig. 3.7(a). However, they may be successfully used for photodetectors, because the low indirect-semiconductor absorption probability may be effectively increased by an extended interaction length (i. e., by a longer absorption region). If the photon energy becomes larger than W_G', the absorption becomes very likely, because electron transitions with constant momentum near $k_\mu = 0$ are possible. For the bandgap energies the following values hold:

$$W_G = \begin{cases} 0.67\,\text{eV} \mathrel{\widehat{=}} 1.85\,\mu\text{m (Ge)} \\ 1.13\,\text{eV} \mathrel{\widehat{=}} 1.10\,\mu\text{m (Si)} \end{cases} \quad W_G' = \begin{cases} 0.8\,\text{eV} \mathrel{\widehat{=}} 1.55\,\mu\text{m (Ge)} \\ 3.4\,\text{eV} \mathrel{\widehat{=}} 0.36\,\mu\text{m (Si)} \end{cases} \tag{3.16}$$

In direct semiconductors, Fig. 3.7(b), the transitions from the lower CB edge to the upper VB edge and vice versa are possible for a constant crystal momentum $\hbar k_\mu = 0$. These processes are therefore very likely, so that direct semiconductors may be used both for light sources and for detectors.

Consider a direct semiconductor Fig. 3.7(b). In the vicinity of $k_\mu = 0$ the actual bandstructure may be approximated by a parabola,

$$W = W_i(k_\mu) = W_i(0) + \frac{1}{2}\frac{\mathrm{d}^2 W_i}{\mathrm{d}k_\mu^2} k_\mu^2 = W_0 + \frac{p_\mu^2}{2m_{\text{eff}}} = W_0 + \frac{\hbar^2 k_\mu^2}{2m_{\text{eff}}}\,, \tag{3.17}$$

defining an effective mass m_{eff} which should be attributed to a free electron at the same energy. The effective mass is negative for crystal electrons at the upper VB edge. An electron with charge

$(-e)$ and effective electron mass $m_n = m_{\text{eff}} < 0$ can be equivalently replaced by a so-called hole (a missing state *not* occupied by an electron) with positive charge $(+e)$ and effective hole mass $m_p = |m_{\text{eff}}| > 0$.

For laser action the population of the CB and VB edges of direct semiconductors near $k_\mu = 0$ is important. If Z stands for the number of electron states with 2 spin directions and a modulus of the crystal momentum up to $p_\mu = \hbar k_\mu$, we have in analogy to Eq. (3.6)

$$Z = \frac{2V_\phi}{h^3}, \qquad V_\phi = VV_p, \qquad V_p = \frac{4\pi}{3}(\hbar k_\mu)^3, \qquad \hbar k_\mu = p_\mu = \sqrt{p_x^2 + p_y^2 + p_z^2}\,. \tag{3.18}$$

In k_μ-space the differential volume in a k_μ-radius interval $k_\mu \ldots k_\mu + \mathrm{d}k_\mu$ is $\mathrm{d}^3\,k_\mu = 4\pi k_\mu^2\,\mathrm{d}\,k_\mu$, and the differential number of states $\mathrm{d}Z$ which corresponds to an energy interval $W \ldots W + \mathrm{d}\,W$ amounts to

$$\mathrm{d}Z = 2V\,\frac{1}{(2\pi\hbar)^3}\,4\pi(\hbar k_\mu)^2\,\mathrm{d}(\hbar k_\mu) = 2V\,\frac{1}{(2\pi)^3}\,\mathrm{d}^3\,k_\mu = V\rho\,(W)\,\mathrm{d}W\,. \tag{3.19}$$

Using Eq. (3.17) we find for the so-called density of states (DOS) $\rho\,(W)$, see Eq. (3.7),

$$\rho(W) = \frac{1}{V}\frac{\mathrm{d}Z}{\mathrm{d}W} = \frac{1}{V}\frac{\mathrm{d}Z}{\mathrm{d}p_\mu}\frac{\mathrm{d}p_\mu}{\mathrm{d}W} = \frac{1}{2\pi^2}\left(\frac{2|m_{\text{eff}}|}{\hbar^2}\right)^{3/2}\sqrt{\pm(W - W_0)}\,. \tag{3.20}$$

For the CB we take the positive sign $\sqrt{+(W - W_0)}$ and $W_0 = W_C$, $|m_{\text{eff}}| = m_n$, and for the VB the negative sign $\sqrt{-(W - W_0)}$, $W_0 = W_V$, $|m_{\text{eff}}| = m_p$. The effective density of states N_B is defined as

$$\frac{\mathrm{d}Z}{\mathrm{d}W}\,kT = V\frac{2}{\sqrt{\pi}}N_B\sqrt{\pm\frac{W - W_0}{kT}}, \qquad N_B = 2\left(\frac{2\pi|m_{\text{eff}}|kT}{h^2}\right)^{3/2}. \tag{3.21}$$

Because $\int_{W_0}^{W_0+kT}\sqrt{\frac{W-W_0}{kT}}\,\frac{\mathrm{d}W}{kT} = \int_0^1\sqrt{W'}\,\mathrm{d}\,W' = \frac{2}{3}$ holds, we find $\frac{1}{V}\int_{W_0}^{W_0+kT}\mathrm{d}Z = \frac{2}{\sqrt{\pi}}\frac{2}{3}N_B \approx 0.752 \times N_B$. Therefore the effective density of states N_B specifies approximately the density of states inside an energy interval $W_0 \ldots W_0 \pm kT$ measured from the bandedge energy W_0. The effective DOS near the conduction and valence band edges are N_C and N_V, respectively. With the free electron rest mass m_0 and the effective carrier masses for GaAs and InP at $T = 293\,\mathrm{K}$ we

	m_n/m_0	m_p/m_0	$N_C/\,\mathrm{cm}^{-3}$	$N_V/\,\mathrm{cm}^{-3}$
vacuum	1	–	2.42×10^{19}	–
GaAs	0.067	0.48	4.20×10^{17}	8.05×10^{18}
InP	0.077	0.64	5.17×10^{17}	1.24×10^{19}

Table 3.3. Examples for effective masses and effective DOS ($T = 293$ K)

find the values specified in Table 3.3. A doped semiconductor[21] in the saturation range[22] is called degenerately doped, if the dopant concentration is larger than the effective DOS N_B; in this case the Fermi level moves *into* the band.

[21] Singh, J.: Physics of semiconductors and their heterostructures. New York: McGraw-Hill 1993

[22] See Sect. 8.4.2 Page 270 in Reference 21 on Page 60: „In general, there are three regions of interest for doped (extrinsic) semiconductors. At very low temperatures, the electrons (holes) are trapped at the donor (acceptor) levels and the free carrier density goes to zero. This region is called the *freeze-out* range. At higher temperatures, the shallow levels are ionized and there is little change in free carrier density. This region is called the *saturation* range. Finally, at very high temperatures, the intrinsic carrier density exceeds the doping levels and the carrier density ($n \approx p$) increases exponentially as for an *intrinsic* material. The higher the bandgap, the higher the temperature where this regime takes over. However, electronic devices cannot operate in this regime."

3.2.2 Filling of electronic states

In a quantum mechanical treatment,[23] particles fall in two categories: fermions and bosons. Particles like photons and phonons are bosons having integer spins $0, \hbar, 2\hbar, \ldots$ Particles such as electrons are fermions with spins $\hbar/2, 3\hbar/2, 5\hbar/2 \ldots$ This subtle difference forces a very important distinction on the occupation statistics. Only one fermion can occupy a quantum state, while any number of bosons can be placed in a particular state. This is the reason why electromagnetic fields can be amplified.

At the absolute temperature $T = 0$ the electrons fill the lowest energy states. At $T > 0$ the distribution which minimizes the free energy of the system is, for fermions, the Fermi-Dirac[24] distribution (Fermi[25] function for short),

$$f(W) = \frac{1}{1 + g \exp\left(\frac{W - W_F}{kT}\right)}, \qquad g = \begin{cases} 1 & \text{band states} \\ 1/2 & \text{donor states} \\ 2 & \text{acceptor states} \end{cases} . \tag{3.22}$$

For impurities a degeneracy factor[26] g has to be taken into account. Figure 3.8 displays the Fermi function for band states. Here W_F is called the chemical potential or Fermi energy, and it represents the energy where the occupation probability $f(W)$ becomes $1/2$ at all temperatures. The transition from a large to a low occupation probability ($0.88 \geq f(W) \geq 0.12$) takes place in a region $4kT$ centred at the Fermi energy W_F (at $T = 293\,\mathrm{K}$ we have $kT = 25\,\mathrm{meV}$ or $\Delta f = 2kT/h = 12.1\,\mathrm{THz}$). In addition to the quantum statistics Eq. (3.22), we also have the classical statistics of Boltzmann[27] which can be derived from Eq. (3.22),

$$\begin{aligned} f(W) &\approx g \exp\left(-\frac{W - W_F}{kT}\right) \qquad \text{for } W - W_F > 3kT\,, \\ f(W) &\approx 1 - g \exp\left(\frac{W - W_F}{kT}\right) \qquad \text{for } W - W_F < -3kT\,. \end{aligned} \tag{3.23}$$

The residual error is smaller than 5 % ($e^3 \approx 20$), if the Fermi level has an energetic distance from the bandedges W_C, W_V of at least three times the thermal energy kT. Especially for undoped semiconductors (see Sect. 3.2.3 and Eq. (3.26)) the approximation is very good.

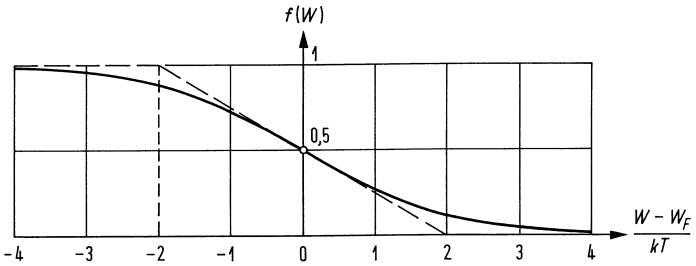


Fig. 3.8. Fermi function for band energy states ($g = 1$)

[23] See Sect. 1.5 in reference Footnote 21 on Page 60

[24] See Footnote 13 on Page 12

[25] See Footnote 17 on Page 54

[26] A *donor* electron has one of two possible spin orientations. For the occupation of the state with an electron there are two favourite cases (spin $+\hbar/2$, spin $-\hbar/2$) out of three possibilities (electron at donor, i. e., spin $+\hbar/2$, spin $-\hbar/2$; no electron at donor). Therefore we have $f(W_F) = 2/3$ for a donor at $W = W_D = W_F$, Eq. (3.22). *Acceptors* bind an electron with a well-defined spin to fill up the outmost shell. For the transition of an electron to an acceptor atom there is one favourite case (fitting spin) out of three possible cases (electron transition to acceptor with fitting spin, no VB electron available with fitting spin, no electron transition to acceptor at all). As a consequence, $f(W_F) = 1/3$ for an acceptor at $W = W_A = W_F$, Eq. (3.22).

[27] See Footnote 16 on Page 53

3.2.3 Impurities and doping

The density n_T of CB electrons and the density p of VB holes may be calculated with the help of the DOS Eq. (3.20), the effective DOS Eq. (3.21) and the Fermi distribution Eq. (3.22),

$$n_T = \int_{W_C}^{\infty} \rho_C(W)\, f(W)\, \mathrm{d}W \qquad p = \int_{-\infty}^{W_V} \rho_V(W)\big[1 - f(W)\big]\, \mathrm{d}W\,. \tag{3.24}$$

With the Boltzmann approximations (valid for $n_T \ll N_C$, $p \ll N_V$ only, i. e., for non-degenerate doping) the integrals Eq. (3.24) can be solved,

$$\left.\begin{aligned} n_T &= N_C \exp\left(-\frac{W_C - W_F}{kT}\right) \\ p &= N_V \exp\left(-\frac{W_F - W_V}{kT}\right) \end{aligned}\right\} \quad n_T p = n_i^2 = N_C N_V \exp\left(-\frac{W_G}{kT}\right) \tag{3.25}$$

The intrinsic carrier concentration n_i refers to the electrons n_T (holes p) present in the CB (VB) of a pure undoped semiconductor. It depends on the bandgap W_G as well as the details of the bandedge masses, but *not* on the Fermi energy W_F; this is a relation expressing the law of mass action. At $T = 0$ the VB is completely occupied while the CB is empty, and the semiconductor has an extremely high resistance. With increasing temperature some electrons are thermally excited from the VB to the CB so that CB electrons and VB holes are generated in pairs, $n_T = p = n_i$. The presence of intrinsic carriers is detrimental to devices[28] where the current has to be modulated by some means. The Fermi level follows from Eq. (3.25) for charge neutrality $n_T = p$,

$$W_F = \frac{1}{2}(W_C + W_V) + kT\, \ln\sqrt{\frac{N_V}{N_C}} = \frac{1}{2}(W_C + W_V) + \frac{3}{4}kT\, \ln\frac{m_p}{m_n}\,. \tag{3.26}$$

At $T = 0$ the chemical potential W_F of the intrinsic semiconductor is in the centre of the forbidden band, at $T > 0$ the Fermi level shifts into the direction of the faster filling band which owns the smaller effective DOS N_B.

Pure semiconductors would have little use by themselves because of their low conductivity (carrier concentration at room temperature $\sim 10^{11}\,\mathrm{cm}^{-3}$) compared to metals ($\sim 10^{21}\,\mathrm{cm}^{-3}$). By introducing impurities the properties of semiconductors may be tailored to specific needs. When a dopant (impurity) atom is implanted into a crystal, its perfect periodicity is destroyed and additional energy levels for electrons located near the bandedges are the outcome. These levels are either near the CB edge (W_D) and can "donate" an electron to the CB (donor), or they are near the VB edge (W_A) where they can accept an electron from the VB (acceptor). The donor (acceptor) concentrations are n_D (n_A), the concentrations of the neutral donor (acceptor) atoms are $n_D^{\times}$ ($n_A^{\times}$), and the concentrations of the ionized impurities are n_D^{+} (n_A^{-}). Thus, either a quasi-free CB electron or a quasi-free VB hole is created when the impurity atoms give or take an electron by thermal excitation for $|W_{C,V} - W_{D,A}| < kT$, Fig. 3.9. For a large impurity concentration Fig. 3.9(b),(c) the impurity levels broaden to form impurity bands which may overlap with the CB or the VB, respectively. In this case, the bandgap is decreased to $W_{G_{\mathrm{eff}}}$, and the DOS $\rho(W)$ cannot be

[28] Some intrinsic carrier concentrations at room temperature $T_0 = 293\,\mathrm{K}$: $n_{i\,\mathrm{Si}} = 1.5 \times 10^{10}\,\mathrm{cm}^{-3}$, $n_{i\,\mathrm{Ge}} = 2.4 \times 10^{13}\,\mathrm{cm}^{-3}$, $n_{i\,\mathrm{GaAs}} = 1.8 \times 10^{6}\,\mathrm{cm}^{-3}$ (actually not achievable), $n_{i\,\mathrm{InP}} = 1.2 \times 10^{8}\,\mathrm{cm}^{-3}$. "The fact that n_i^2 is constant at a given temperature is often utilized to produce high resistivity (insulating) materials from impure semiconductors. Consider, for example, impure GaAs with $n_T = 10^{16}\,\mathrm{cm}^{-3}$ and $p = 10^{5}\,\mathrm{cm}^{-3}$ giving a total free carrier density $n_T + p = 10^{16}\,\mathrm{cm}^{-3}$ and $n_i^2 = n_T p = 10^{21}\,\mathrm{cm}^{-6}$ ($n_i = 3.2 \times 10^{10}\,\mathrm{cm}^{-3}$) at 180 °C. If the p-type carrier concentration is now increased by doping to $3.2 \times 10^{10}\,\mathrm{cm}^{-3}$, the sum concentration becomes $n_T + p \approx 6.4 \times 10^{10}\,\mathrm{cm}^{-3}$ since the $n_T p$ product must remain the same. The Fermi level shifts into the direction of the forbidden-band centre. This greatly reduces the material conductivity. This technique is called compensation. It must be remembered, of course, that the $n_T p$ product is constant only when the system is *in equilibrium*." Sect. 8.1 in reference Footnote 21 on Page 60

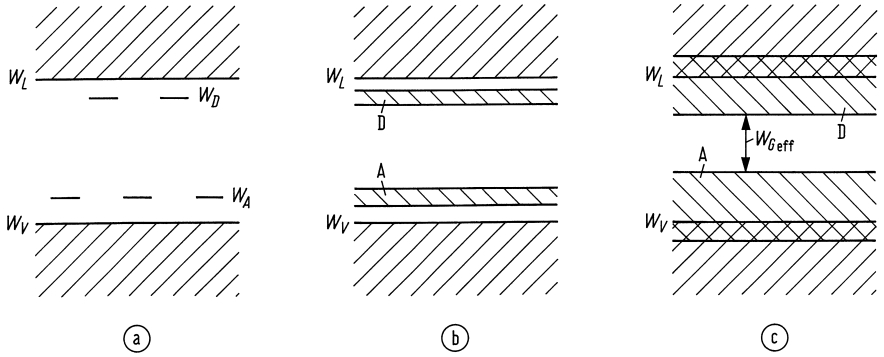


Fig. 3.9. Energy levels of impurities in a semiconductor. (a) isolated donors and acceptors. (b) impurity bands for heavier doping. (c) overlap of impurity bands with CB annd VB for heavy doping ($W_L \mathrel{\widehat{=}} W_C$, Leitungsband $\mathrel{\widehat{=}}$ conduction band)

approximated by a parabola near the bandedges in Eq. (3.20). Equation (3.25) remains valid. The Fermi level W_F can be computed from Eqs. (3.22), (3.24), (3.25) in the case of charge neutrality,

$$n_T + n_A^- = p + n_D^+ \qquad \begin{array}{ll} n_D = n_D^\times + n_D^+ & \text{donor density,} \\ n_A = n_A^\times + n_A^- & \text{acceptor density.} \end{array} \tag{3.27}$$

If for donors $(W_C - W_D)/(kT) \ll 1$ is valid (saturation, practically all donors are ionized), and if $n_D > N_C$, then the Fermi energy W_F is shifted into the CB. An analogue relation holds for acceptors.

For a non-equilibrium condition where a constant perturbation is switched on, the carriers in the CB and VB states need some time to re-arrange. This time is called the intraband relaxation time τ_{CB}, τ_{VB}. For example: At $t = 0$ we have a constant field strength $\vec{E}$ in the semiconductor. After a transit time τ the electron drift velocity is $\vec{v}_n = -\mu_n \vec{E}$ (mobility μ_n, expectation of electron velocity $\vec{v}_n$). If the relaxation follows the function $\exp(-t/\tau)$, then the time constant τ corresponds to the re-arrangement time of the electrons with respect to the CB states, and $\tau = \tau_{\mathrm{CB}}$ is called the momentum relaxation time. In terms of the intraband relaxation times, the mobilities of electrons and holes are $\mu_n = e\tau_{\mathrm{CB}}/m_n$ and $\mu_p = e\tau_{\mathrm{VB}}/m_p$. For InP we find $\mu_n = 4\,600\,\mathrm{cm}^2\,/\,\mathrm{V}$, $m_n/m_0 = 0.077$, $m_0 = 9.109\,387\,9 \times 10^{-31}\,\mathrm{kg}$, $e = 1.602\,177\,33 \times 10^{-19}\,\mathrm{C}$ from Table 3.3 on Page 60, and $\tau_{\mathrm{CB}} = 0.2\,\mathrm{ps}$ follows. After this time has passed, the occupation probability may be again described by the Fermi distribution $f(W)$ Eq. (3.22). However, because n_T, p assume different values than in the equilibrium case, the Fermi energy W_{Fn} of the electrons in the conduction band (i. e., the energy state with electron occupation probability $f_C(W_{Fn}) = 1/2$) differs from the Fermi level W_{Fp} of valence band holes (i. e., from the energy state with hole occupation probability $f_V(W_{Fp}) = 1/2$). The quantities W_{Fn} and W_{Fp} are therefore denoted as quasi Fermi levels for electrons and holes in the non-equilibrium case. The occupation probabilities for conduction band electrons and valence band holes read now

$$f_C(W) = \frac{1}{1 + \exp\left(\dfrac{W - W_{Fn}}{kT}\right)}, \quad f_V(W) = \frac{1}{1 + \exp\left(\dfrac{W - W_{Fp}}{kT}\right)}. \tag{3.28}$$

After the intraband relaxation time τ_{LB} the conduction band electrons are in a new dynamic equilibrium as it is the case for the valence band holes when the intraband relaxation time τ_{VB} has passed. However, the conduction band electrons are not in equilibrium with the valence band

holes, $W_{Fn} \neq W_{Fp}$. Analogue to Eq. (3.25) we find

$$\left.\begin{aligned} n_T &= N_C \exp\left(-\frac{W_C - W_{Fn}}{kT}\right) \\ p &= N_V \exp\left(-\frac{W_{Fp} - W_V}{kT}\right) \end{aligned}\right\} \quad \begin{aligned} n_T p &= n_i^2 \exp\left(\frac{W_{Fn} - W_{Fp}}{kT}\right) \\ n_i^2 &= N_C N_V \exp\left(-\frac{W_G}{kT}\right). \end{aligned} \tag{3.29}$$

In Sect. 3.1.3 Eq. (3.13) on Page 54 it was made plausible that laser action in a semiconductor laser requires the photon energy to be inside the bounds $W_C - W_V < hf \leq W_{Fn} - W_{Fp}$, so that either the condition $W_{Fn} \geq W_C$ or $W_{Fp} \leq W_V$ must be met. Following Eq. (3.29) the necessary pump

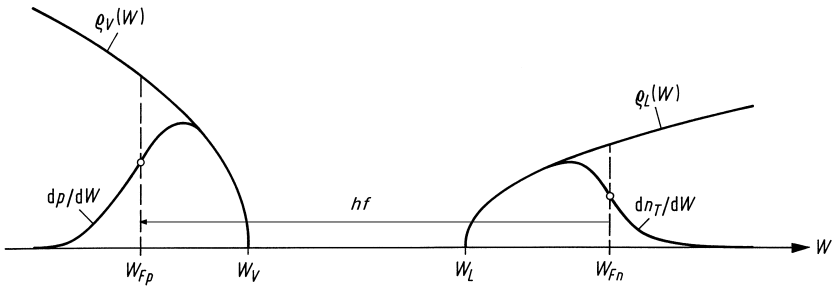


Fig. 3.10. Distribution of valence band holes and conduction band electrons under non-equilibrium conditions $n_T p \gg n_i^2$ for an inverted semiconductor at $T > 0$. The areas under the $\mathrm{d}p/\,\mathrm{d}W$ and $\mathrm{d}n_T/\,\mathrm{d}W$ curves stand for the hole and electron concentrations p and n_T in the valence and in the conduction bands, respectively. The arrow marked by hf indicates the maximum photon energy at which optical amplification is still achieved, see Eq. (3.13) on Page 54. ($\rho_L \mathrel{\widehat{=}} \rho_C$ conduction band DOS)

can be realized by carrier injection with $n_T p \gg n_i^2$. Then an increased radiative recombination rate leads to an increased emission of spontaneous photons compared to the case of true thermal equilibrium. From Eq. (3.24) we have

$$\frac{\mathrm{d}n_T}{\mathrm{d}W} = \rho_C(W)\, f_C(W), \qquad \frac{\mathrm{d}p}{\mathrm{d}W} = \rho_V(W)\,\big[1 - f_V(W)\big]. \tag{3.30}$$

Figure 3.10 shows the distribution of holes and electrons inside the bands of a population inverted semiconductor laser, and the maximum photon energy $hf = W_{Fn} - W_{Fp}$ at which an optical amplification is still possible. This will be discussed in more detail in Sect. 3.2.5.

To calculate the threshold current of a laser, we need the carrier concentrations n_T, p for shifting the quasi Fermi levels W_{Fn}, W_{Fp} into the bands. We assume a p-doped semiconductor with an equilibrium concentration of n_{T0}, p_0 and $n_{T0}p_0 = n_i^2$. By carrier injection the densities are changed to $n_T = n_{T0} + \Delta n_T$, $p = p_0 + \Delta p$. By substitution into Eq. (3.29) we find

$$W_{Fn} - W_F = kT \ln\left(1 + \frac{\Delta n_T}{n_{T0}}\right), \quad W_F - W_{Fp} = kT \ln\left(1 + \frac{\Delta p}{p_0}\right). \tag{3.31}$$

Further, we assume charge neutrality $\Delta p = \Delta n_T$. By a carrier injection the quasi Fermi level of the minority carriers shifts first (here: W_{Fn}; change of n_T by Δn_T has largest effect because n_{T0} is small). At $\Delta n_T / n_{T0} = 1$ the shift amounts to $W_{Fn} - W_F = 0.7\,kT$. Because $p_0 \gg n_{T0}$ holds in a p-semiconductor the quasi Fermi level for majority carriers starts to shift at much higher injection current levels when $\Delta p = \Delta n_T$ reaches the order of p_0.

3.2.4 Heterojunctions

Heterojunctions are composed of semiconductors with different bandgap energies W_G. They are advantageous for laser diodes and photodetectors. With $(\mathrm{Ga}_{1-x}\mathrm{Al}_x)\mathrm{As}$ of Table 3.1 and $0 \leq x \leq 0.3$

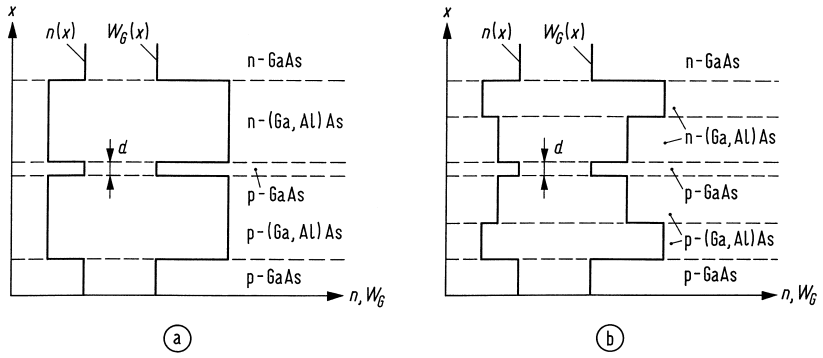


Fig. 3.11. Schematic refractive index dependence n and bandgap W_G as a function of the spatial coordinate x in a (a) 3-layer heterostructure, (b) 5-layer heterostructure

the bandgap energy W_G can be increased by 374 meV while the refractive index n decreases by nearly 6 %. Lasers are built as 3-layer or 5-layer heterostructures, Fig. 3.11. For the 3-layer heterostructure Fig. 3.11(a) the active layer (the region with induced amplification) has a thickness of $d = 0.1 \ldots 0.2\,\mu$m and consists of p-GaAs. A slight p-doping[29] decreases the electron concentration in the valence band, thereby facilitating a population inversion. The neighbouring layers are formed of (Ga,Al)As having a larger bandgap W_G and a lower refractive index n leading to the following features:

Potential walls exist for carriers n_T, p injected from both sides of the p-GaAs layer. Even from low current densities $J > 0.5\,\mathrm{kA\,/\,cm^2}$ onwards the carrier concentration n_T inside the active layer is so large that the difference of the quasi Fermi levels exceeds the bandgap, $W_{Fn} - W_{Fp} > W_G$, and laser action starts (for GaAs at about $n_T = 2 \times 10^{18}\,\mathrm{cm^{-3}}$).

Larger W_G in the (Ga,Al)As layers blocks the induced re-absorption in the non-inverted regions.

Smaller n in the (Ga,Al)As layers characterizes the cladding of a slab waveguide where the core is represented by the active layer. Because d is small, a large portion ($\approx 80\,\%$) of the electromagnetic energy propagates inside the cladding (field confinement factor $\Gamma \approx 20\,\%$).

For a 3-layer structure both the carrier and the field confinement are determined by the thickness d of the active layer, and by W_G (i. e., by the refractive index n) of the three central layers. The actual pn-junction is between the p-GaAs active layer and the neighbouring n-(Ga,Al)As layers. Devices with heterojunctions on both sides of the active zone are called double-heterostructures.

With a 5-layer heterostructure Fig. 3.11(b) the carrier confinement and the vertical field confinement in x-direction become independent. The field is guided by two (Ga,Al)As layers on each side of the active zone. Fortunately, the refractive index n in (Ga,Al)As depends only weakly on the doping ($\sim 0.1\,\%$).

Heterojunctions are called "isotype" if the semiconductors have the same conduction type, and "anisotype" if the conduction type differs. The conduction type is specified with small letters n, i, p if the semiconductor has a smaller bandgap than its neighbour, and with capital letters N, I, P if the bandgap is larger. For the structure in Fig. 3.11(a) we see the following junction types from top to bottom: nN, Np, pP and Pp. In the following, we discuss some heterojunction properties in analogy to the ordinary pn-junction.

Band diagram for heterostructures

Figure 3.12(a) explains the energy scale. Free electrons are at the vacuum energy level $W = 0$ if they move at a velocity $\vec{v} = 0$ in a region with constant potential $\varphi = 0$; in Fig. 3.12(a) we further

[29] See Footnote 28 on Page 62, Footnote 21 on Page 60

assume a potential $\varphi \neq 0$. Electrons leaving the semiconductor with $\vec{v} = 0$ are then at the vacuum

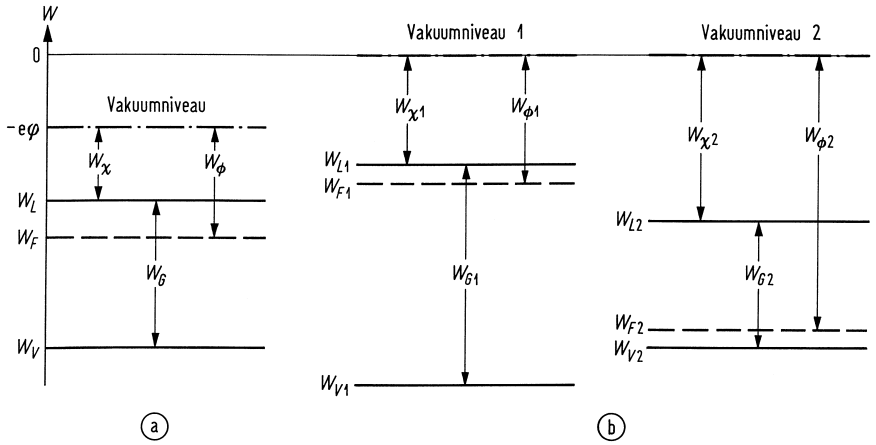


Fig. 3.12. Energy scale for electrons in a semiconductor. (a) Semiconductor at potential $\varphi \neq 0$. (b) Two independent, insulated semiconductors at potential $\varphi = 0$ with different bandgaps. W_χ electron affinity, W_ϕ work function. W_L conduction band edge ($\widehat{=} W_C$, Vacuumniveau $\widehat{=}$ vacuum level)

energy level $W = -e\varphi$. The electron affinity W_χ specifies the energetic distance from the CB edge to the vacuum level. The work function W_ϕ is defined by the energy difference of the Fermi and the vacuum level. All these quantities together with the bandgap $W_G = W_C - W_V$ are positive by definition. W_χ, W_ϕ and W_G fix the energetic distances of the CB edge, the Fermi level and the VB edge in relation to the vacuum level.

In Fig. 3.12(b) the energy-band diagrams of two semiconductors with different bandgaps are displayed. After forming the contact we have a Np-junction as in Fig. 3.11(a). Assuming $\varphi = 0$ for both separated semiconductors we define the quantities

$$\begin{aligned} \Delta W_G &= W_{G2} - W_{G1}\,, \\ \Delta W_C &= W_{C2} - W_{C1} = W_{\chi 1} - W_{\chi 2}\,, \\ \Delta W_V &= W_{V2} - W_{V1} = \Delta W_C - \Delta W_G\,. \end{aligned} \tag{3.32}$$

In the case of Fig. 3.12(b). the relations ΔW_G, $\Delta W_C < 0$ and $\Delta W_V > 0$ hold. For $(Ga_{1-x}Al_x)As$ with $0 \leq x \leq 0.3$ we find nearly independently of x

$$\frac{\Delta W_C}{\Delta W_G} = 0.65\,, \qquad \frac{\Delta W_V}{\Delta W_G} = -0.35\,. \tag{3.33}$$

For each of the (non-degenerate) semiconductors in Fig. 3.12(b), Eqs. (3.25)–(3.27) are valid. When the contact is formed all states of equal energy are occupied with the same probability in the case of thermal equilibrium. If we fix the potential of semiconductor 2 at $\varphi_2 = 0$, the potential of semiconductor 1 rises until $e\varphi_1 + W_{\phi 1} = W_{\phi 2}$, i. e., the potential is given by the so-called built-in potential $U_D = \varphi_1 = (W_{F1} - W_{F2})/e > 0$ (*German* Diffusionspannung). From Eq. (3.25) we calculate

$$e\varphi_1 = W_{F1} - W_{F2} = W_{C1} - W_{V2} + kT\,\ln\frac{n_{T1}p_2}{N_{C1}N_{V2}} = W_{G1} - \Delta W_V + kT\,\ln\frac{n_{T1}p_2}{N_{C1}N_{V2}}\,. \tag{3.34}$$

Under the assumption of shallow saturated impurities with $n_{T1} = n_D$, $p_2 = n_A$, we find from Eq. (3.25) for a homojunction the diffusion voltage or built-in potential U_D of the pn-junction (U_T is the thermal voltage)

$$\varphi_1 = U_D = U_T\,\ln\frac{n_D n_A}{n_i^2}, \qquad U_T = \frac{kT}{e}, \qquad W_G = -kT\,\ln\frac{n_i^2}{N_C N_V}\,. \tag{3.35}$$

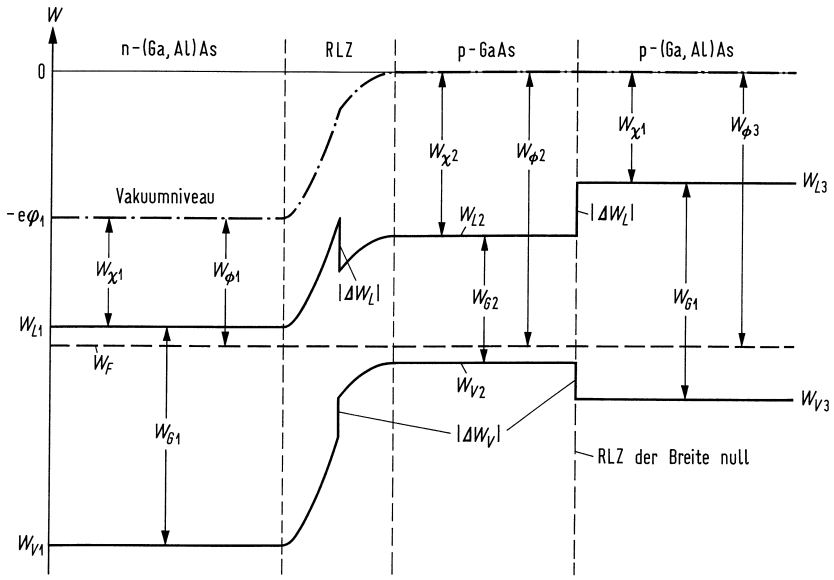


Fig. 3.13. Energy-band diagram of a double-heterostructure with anisotype Np-junction and a special isotype pP-junction with diffusion voltage zero ($W_L \mathrel{\widehat{=}} W_C$, Leitungsband $\mathrel{\widehat{=}}$ conduction band, Vakuumniveau $\mathrel{\widehat{=}}$ vacuum level, Raumladungszone RLZ der Breite null $\mathrel{\widehat{=}}$ space-charge region of zero width)

At room temperature $T = 293\,\mathrm{K}$ the thermal voltage is $U_T = 25\,\mathrm{mV}$. Figure 3.13 (not drawn to scale) shows the band-energy diagram of the NpP-heterojunction of Fig. 3.11(a) in the case of thermal equilibrium; we used the semiconductors of Fig. 3.12(b), supplemented by a p-semiconductor with electron affinity $W_{\chi 1}$ and bandgap W_{G1}. The doping of the p-(Ga,Al)As layer was chosen such that the diffusion voltage of the isotype pP-junction is zero, $U_{D\,\mathrm{pP}} = 0$. The bandedge energies (and therefore the carrier concentrations) are not continuous but exhibit steps by $|\Delta W_C|, |\Delta W_V|$, see Eq. (3.32). The component of the dielectric displacement vector $\vec{D} = \epsilon_0 \epsilon_r \vec{E}$ perpendicularly to the boundary plane is continuous while the refractive index $n = \sqrt{\epsilon_r}$ is discontinuous. Therefore, the normal component of the electric field vector $\vec{E} = -\operatorname{grad} \varphi$ is also discontinuous. This leads to a kink of the vacuum level at the semiconductor boundary; the slope inside the semiconductor with the larger W_G (smaller n) is larger than inside the semiconductor with the smaller W_G.

Figure 3.14 displays the energy-band diagram from Fig. 3.13 assuming the flat-band case for simplicity. However, because of unavoidable series resistances it is practically impossible to adjust an external forward voltage such that the junction voltage U compensates the diffusion voltage U_D from Fig. 3.13. Far away from the junction the quasi Fermi levels of electrons and holes are practically identical. However, inside the thin p-GaAs layer and inside the diffusion zones we have $W_{Fn} > W_{Fp}$ due to the carrier injection. Because of the longer diffusion length of electrons compared to the diffusion length of holes, $L_n > L_p$, the diffusion zone of the p-semiconductor is larger than inside the n-semiconductor (for GaAs: $L_n/L_p = 5$).

Electrons and holes are confined to a potential film[30] (quantum film) inside the p-GaAs layer. In Fig. 3.14 the quasi Fermi level for electrons W_{Fn} was moved into the CB. By appropriate injection (pump) currents, the semiconductor may be population inverted, see Eq. (3.13).

[30]This quantum film, a thin layer between layers with larger bandgaps, confines electrons in one direction only (the growth direction). Nonetheless, the structure is usually called a potential or quantum "well". However, according to common understanding, a well has a two-dimensional structure and confines water (or electrons in our case) in two orthogonal directions.

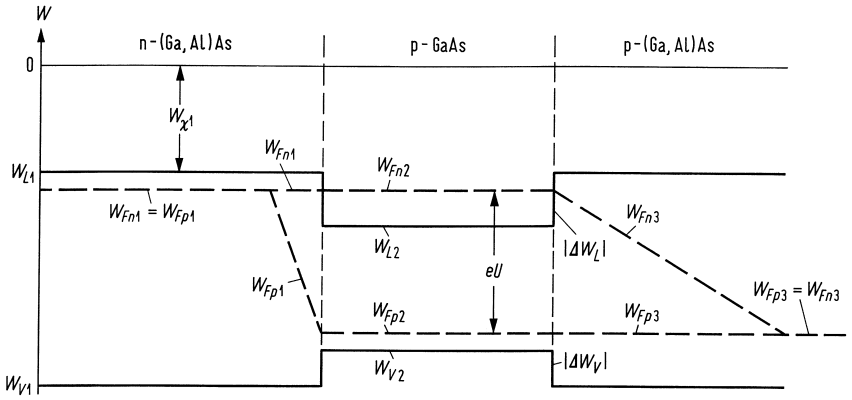


Fig. 3.14. Energy-band diagram of a NpP-heterojunction of Fig. 3.13 with a forward bias voltage. $L_p/L_n \approx 0.2$ for GaAs ($W_L \mathrel{\widehat{=}} W_C$)

3.2.5 Emission and absorption of light in a semiconductor

General considerations

Let us assume a microsystem according to Fig. 3.3(a) with energy levels W_2, W_1 and $W_2 = W_1 + hf$. The electromagnetic field in the active optical volume V is given as an expansion of orthonormal modes with total number $M_{\rm tot}$, Eq. (3.4).

A certain mode with frequency f (photon energy hf) contains N_P photons. As a result of a first-order perturbation theory[31] (dubbed "Fermi's Golden Rule") we find the probabilities for the induced ($w^{(\rm e)}_{\rm ind}$) and spontaneous ($w^{(\rm e)}_{\rm sp}$) photon emission or absorption of a microsystem, which is initially in the excited state or in the ground state, respectively,

$$w^{(\rm e)} = w^{(\rm e)}_{\rm ind} + w^{(\rm e)}_{\rm sp} \sim (N_P + 1)\, t\,, \qquad w^{(\rm a)} \sim N_P t \tag{3.36}$$

The relations are asymptotically valid for $t \to \infty$, because only then the excited state has a well defined energy. Spontaneous transitions which are independent of the photon number N_P are only possible for emission, see the discussion on Page 51.

The transition probability $w = 1$ specifies the lifetime τ of the energy state. The energy is conserved for the transition if $t, \tau \to \infty$ holds, $w^{(\rm e)}, w^{(\rm a)} \sim \delta(W_2 - W_1 - hf)$. If the interaction time t is finite, the δ-function of the Golden Rule changes over to a broadened function $\rho(f)$ so that in this case the energy conservation is no longer strictly satisfied.

Note that according to Eq. (3.8) the energy uncertainty for $\tau_2 = 0.1\,\rm ps$ is about $\Delta W_2 = 5\,\rm meV \mathrel{\widehat{=}} 6\,\rm THz$. Spontaneously emitted photons occupy at random all electromagnetic modes with frequency f in V. With Eqs. (3.4), (3.7) for the total number $M_{\rm tot}$ of modes in V for frequencies $0 \ldots f$ and for the spectral mode density $\varrho_{\rm tot}(f)\, V = {\rm d}M_{\rm tot}/\,{\rm d}f$, the probability of a spontaneous emission, $w^{(\rm e)}_{\rm sp}$, in any mode of frequency f is $w_{\rm sp} = \int w^{(\rm e)}_{\rm sp}\, {\rm d}M_{\rm tot}$.

Because of the finite lifetime the emission is not monochromatic but has a lineshape $\rho(f)$ with a half-maximum bandwidth $\Delta f_H \ll f_0$ centred at f_0. A detailed calculation leads to a Lorentzian,

$$\rho(f) = \frac{2}{\pi \Delta f_H} \frac{1}{1 + \left(\frac{f - f_0}{\Delta f_H/2}\right)^2}\,, \qquad \int\limits_{-\infty}^{+\infty} \rho(f)\, {\rm d}f = 1\,, \qquad 2\pi\, \tau_{\rm sp}\, \Delta f_H = 1\,. \tag{3.37}$$

The probability of an induced emission into a certain electromagnetic mode is by N_P larger than the probability of a spontaneous emission (representing noise) into the same mode. The

[31] See App. I.1 Page 813 in reference Footnote 21 on Page 60

induced emission of photons from a microsystem in state W_2 happens with the same probability as absorption from the ground state W_1.

To achieve amplification the emission rate must be larger than the absorption rate, i. e., the number N_2 of microsystems in the excited state W_2 must be larger than the number N_1 of microsystems in the ground state W_1.

Purcell[32] suggested more than half a century ago to tailor the spontaneous emission probability of radiating dipoles into a specific mode of frequency f by using a cavity to modify the dipole-field coupling and the density of available photon modes.[33]

Induced and spontaneous transitions

Conduction and valence band states of a semiconductor volume V under non-equilibrium conditions are described by the respective Fermi functions $f_C(W), f_V(W)$ Eq. (3.28), and by the density of states $\rho_C(W), \rho_V(W)$ Eq. (3.20). For the carrier concentrations n_T and p we write Eq. (3.24), substituting $f(W)$ by $f_C(W)$ and $f_V(W)$, respectively. We are interested in a certain electromagnetic mode with photon energy $hf = W_2 - W_1$. The total emission probability into this mode results from the product of the probabilities for

- emission $w^{(\mathrm{e})}$,
- occupation of a CB state $f_C(W_2)$, and
- for the event that the corresponding state in the VB is unoccupied, $[1 - f_V(W_1)]$.

Further, the number of states per energy, i. e., the density of states $\rho_C(W_2)$ and $\rho_V(W_1)$, have to be taken into account, Eqs. (3.19), (3.20). Finally, we have to sum over all possible transitions.

The gain is determined by the difference in induced emission $r_{\mathrm{ind}}^{(\mathrm{eM})}$ and absorption rates $r_{\mathrm{ind}}^{(\mathrm{aM})}$ (unit $\mathrm{m}^{-3}\,\mathrm{s}^{-1}$), i. e., by the net number of photons $r_{\mathrm{ind}}^{(\mathrm{M})}$ *e*mitted or *a*bsorbed per volume and time into a fixed *m*ode with frequency f. The spontaneous emission rate $r_{\mathrm{sp}}^{(\mathrm{eM})}$ into mode f does not depend on the photon number. The quantum mechanical properties of the transition (e. g., the transition matrix element $|\mu_{21}|^2$ specifying the interaction with the electromagnetic field, averaged over all possible spatial orientations of the microsystem) are combined in a quantity K_0 (unit $\mathrm{W}^2\,\mathrm{s}$). For a modulus-$\vec{k}_\mu$ selection rule, i. e., a k_μ-selection rule[35,36], we write (without giving a detailed

[32] Purcell, E. M.: Spontaneous emission probabilities at radio frequencies. Phys. Rev. 69 (1946) 681

[33] Let a narrow-line cavity with a linewidth of Δf_H centred at a frequency f_0 be described by the line shape function $\rho(f)$ of Eq. (3.37). Such a resonator with a quality factor $Q_R = f_0/\Delta f_H = \omega_0\,\tau_P$ presents exactly one mode having a photon lifetime τ_P, provided that the dipole radiation is spectrally narrower than Δf_H. The number of modes (= 1) per frequency interval $\pi\Delta f_H/2 \approx 1.6 \times \Delta f_H$ is therefore $1/\left(\pi\Delta f_H/2\right) = 2/(\pi\Delta f_H)$, which represents the equivalent density of cavity modes. Compared with the density of free-space radiating modes $\varrho_{\mathrm{tot}}(f_0)\,V$ in a volume V, the mode density is increased by the so-called Purcell figure of merit[34] F_P,

$$F_P = \frac{\rho(f_0)}{\varrho_{\mathrm{tot}}(f_0)\,V} = \frac{2/(\pi\Delta f_H)}{8\pi V(f_0 n)^2 n_g/c^3} = \frac{Q_R}{4\pi^2 V n^2 n_g (f_0/c)^3}\,, \qquad Q_R = \frac{f_0}{\Delta f_H} = \omega_0\,\tau_P\,. \tag{3.38}$$

Enhancing the spontaneous emission probability (Purcell effect) of a solid-state emitter by making $F_P \gg 1$ would allow in particular the fabrication of high-efficiency light-emitting diodes, see Sect. 3.3.1. This can be achieved with high-Q_R microcavities, and by exploiting the properties of photonic crystals.

[34] Gerard, J.-M.; Gayral, B.: Strong Purcell effect for InAs quantum boxes in three-dimensional solid-state microcavities. J. Lightwave Technol. 17 (1999) 2089–2095. — Here, the Purcell factor definition (see reference Footnote 32 on Page 69) is larger by 3 which stems from a 1/3 averaging factor accounting for the random polarization of free-space modes with respect to the spontaneously radiating dipole. — Various expressions can be found in the literature for F_P, which differ by a numerical factor as large as ten. Therefore, care must be taken in comparing various experimental outcomes.

[35] Adams, M. J.; Landsberg, P. T.: The theory of the injection laser. In: Gooch, C. H. (Ed.): Gallium arsenide lasers. London: Wiley-Interscience 1969. Page 38

[36] On Page 59 of this text the conservation of momentum $\vec{k}_\mu$ for a transition was discussed. The assumption of a $\vec{k}_\mu$-selection rule implies a reasonably pure semiconductor. For an injection laser the impurity concentration has an order of magnitude such that impurity scattering will modify the momentum matrix elements involved in interband transitions, see Ref. 35. The result of such scattering is to effectively relax the strict vectorial $\vec{k}_\mu$-selection rule.

derivation)

$$\begin{aligned} r_{\mathrm{ind}}^{(\mathrm{M})} &= r_{\mathrm{ind}}^{(\mathrm{eM})} - r_{\mathrm{ind}}^{(\mathrm{aM})} \\ &= \tfrac{1}{2} N_P\, V K_0\, \rho_C(W_0)\, \rho_V(W_0 - hf) \big[f_C(W_0) - f_V(W_0 - hf) \big], \\ r_{\mathrm{sp}}^{(\mathrm{eM})} &= \quad \tfrac{1}{2} V K_0\, \rho_C(W_0)\, \rho_V(W_0 - hf)\, f_C(W_0) \big[1 - f_V(W_0 - hf) \big], \\ W_0 &= W_C + \tfrac{\hbar^2 k_{\mu 0}^2}{2 m_n} = W_C + \tfrac{hf - W_G}{1 + m_n/m_p} \quad \text{for a } k_\mu\text{-selection rule.}^{36} \end{aligned} \tag{3.39}$$

For photon energies exceeding the bandgap energy the DOS product[37] is positive,

$$\rho_C(W_0) \rho_V(W_0 - hf) \sim \left(\frac{hf - W_G}{kT_0} \right) kT_0 \,. \tag{3.40}$$

The difference of the Fermi functions in Eq. (3.39) reads

$$\begin{aligned} f_C(W_0) - f_V(W_0 - hf) &= \frac{1}{1 + \exp\Big[\Big(W_C + \frac{hf - W_G}{1 + m_n/m_p} - W_{Fn} \Big) \Big/ kT \Big]} \\ &\quad - \frac{1}{1 + \exp\Big[\Big(W_C - hf + \frac{hf - W_G}{1 + m_n/m_p} - W_{Fp} \Big) \Big/ kT \Big]} \,, \\ f_C(W_0) - f_V(W_0 - hf) &> 0 \quad \text{for} \\ W_C + \frac{hf - W_G}{1 + m_n/m_p} - W_{Fn} &< W_C - hf + \frac{hf - W_G}{1 + m_n/m_p} - W_{Fp} \\ &\quad \text{or} \quad hf < W_{Fn} - W_{Fp} \,. \end{aligned} \tag{3.41}$$

Optical amplification

From Eqs. (3.39), (3.41) it follows that for an optical amplification, i. e., for a net induced emission rate $r_{\mathrm{ind}}^{(\mathrm{M})} > 0$ at $T > 0$, at least *one* quasi Fermi level must be inside the CB or the VB, but (accepting a reduced gain) not necessarily both, see Fig. 3.14. However, at $T = 0$ (not very practical, because the electrons and holes are trapped at the donor and acceptor levels, freeze-out range) *both* quasi Fermi levels need be inside the CB and the VB, respectively. The general inversion condition for an amplification of an electromagnetic wave by a semiconductor reads (see Eq. (3.13))

$$W_G < hf \le W_{Fn} - W_{Fp} \quad \text{for} \quad T \ge 0 \quad \text{and} \quad r_{\mathrm{ind}}^{(\mathrm{M})} = r_{\mathrm{ind}}^{(\mathrm{eM})} - r_{\mathrm{ind}}^{(\mathrm{aM})} \ge 0 \,. \tag{3.42}$$

This includes the transparency point where $r_{\mathrm{ind}}^{(\mathrm{M})} = r_{\mathrm{ind}}^{(\mathrm{eM})} - r_{\mathrm{ind}}^{(\mathrm{aM})} = 0$. From the definition of $r_{\mathrm{ind}}^{(\mathrm{M})}$ (Eq. (3.39)) and from the gain rate G being defined as the temporal increase of the photon number by stimulated transitions, a relation may be established between both quantities,

$$\left.\begin{aligned} r_{\mathrm{ind}}^{(\mathrm{M})} &= \tfrac{1}{V} \tfrac{\mathrm{d}N_P}{\mathrm{d}t} \\ G &= \tfrac{1}{N_P} \tfrac{\mathrm{d}N_P}{\mathrm{d}t} \end{aligned}\right\} \qquad G = \frac{r_{\mathrm{ind}}^{(\mathrm{M})}}{N_P / V}. \tag{3.43}$$

The spontaneous emission rate into the mode f and its net gain rate are connected by the inversion factor n_{sp}, which is determined by the ratio of the number $N_2 \sim r_{\mathrm{sp}}^{(\mathrm{eM})}$ of excited microsystems to

Instead, we require the conservation of the *modulus* $|\vec{k}_{\mu\, C,V}| = k_{\mu\, C,V}$, $k_{\mu\, C} = k_{\mu\, V} = k_{\mu 0}$ of the crystal momentum (k_μ-selection rule, not $\vec{k}_\mu$-selection!). Therefore, only the transitions at energies $W_{2,1} = W_{C,V} \pm (\hbar k_{\mu 0})^2 / (2 m_{n,p})$ in Fig. 3.7(b) are allowed, where $W_2 - W_1 = hf$ holds.

[37] As a consequence of the relaxed k_μ-selection rule, the DOS product is linear in frequency. For a strict $\vec{k}_\mu$-selection rule the stimulated $r_{\mathrm{ind}}^{(\mathrm{M})}$ and spontaneous emission rates $r_{\mathrm{sp}}^{(\mathrm{eM})}$ would vary according to $\sqrt{hf - W_G}$, see reference in Footnote 35 on Page 69

the net number $N_2 - N_1 \sim r_{\rm ind}^{\rm (eM)} - r_{\rm ind}^{\rm (aM)}$ of emitting microsystems (total number $N = N_1 + N_2$),

$$\frac{r_{\rm sp}^{\rm (eM)}}{r_{\rm ind}^{\rm (M)}/N_P} = \frac{r_{\rm sp}^{\rm (eM)}}{G/V} = \frac{f_C(W_0)[1 - f_V(W_0 - hf)]}{f_C(W_0) - f_V(W_0 - hf)} = n_{\rm sp}\,, \tag{3.44}$$

$$n_{\rm sp} = \frac{1}{1 - \exp\left(\dfrac{hf - (W_{Fn} - W_{Fp})}{kT}\right)} = \frac{N_2}{N_2 - N_1} = \frac{1}{1 - N_1/N_2}\,. \tag{3.45}$$

Maximum gain G is reached for complete inversion $n_{\rm sp} = 1$, i. e., for $N_1 = 0$ when all microsystems are excited and the VB is empty, $f_V = 0$. For practical operating points we have $n_{\rm sp} = 1.5 \ldots 2.5$. If the gain rate G is kept fixed for a certain device, the noise caused by the incoherent, spontaneous emission of photons is in proportion to the inversion factor $n_{\rm sp}$. For low-noise optical amplification the inversion factor should be as closely to 1 as possible.

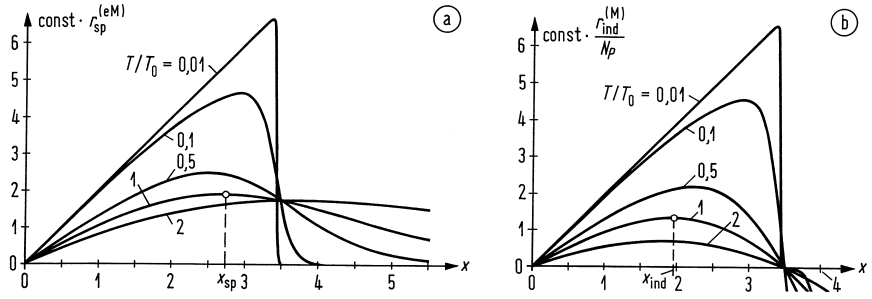


Fig. 3.15. Frequency dependence of spontaneous and induced emission for various temperatures $T/T_0 = 0.01$, 0.1, 0.5, 1, 2 (T_0 reference temperature; $W_{Fn} - W_C$ and $W_V - W_{Fp}$ are kept constant to $3\,kT_0$ and $0.5\,kT_0$, respectively; $m_n/m_p = 0.14$ as in GaAs). Normalized frequency $x = (hf - W_G)/(kT_0)$. (a) Spontaneous emission and (b) induced emission per photon, Eq. (3.46). The multiplicative constant is identical in both diagrams.

Figure 3.15 displays the spontaneous and induced emission spectra Eqs. (3.39), (3.41) for fixed quasi Fermi levels $W_{Fn} - W_C = 3kT_0$, $W_V - W_{Fp} = 0.5\,kT_0$, $(W_{Fn} - W_{Fp}) - W_G = 3.5\,kT_0$ and varying ratios T/T_0 (T_0 is a fixed reference temperature). The carrier masses $m_n/m_p = 0.14$ are that of GaAs, Table 3.3. The photon energy is expressed by the normalized quantity $x = (hf - W_G)/(kT_0)$. The DOS product and the Fermi functions are

$$f_C(W_0) = \frac{1}{1 + \exp\Big[\ \Big(\underbrace{\dfrac{1}{1 + m_n/m_p}}_{0.877}\ \underbrace{\dfrac{hf - W_G}{kT_0}}_{x} - \underbrace{\dfrac{W_{Fn} - W_C}{kT_0}}_{3}\Big)\Big/\dfrac{T}{T_0}\Big]}\,, \tag{3.46}$$

$$f_V(W_0 - hf) = \frac{1}{1 + \exp\Big[-\Big(\underbrace{\dfrac{1}{1 + m_p/m_n}}_{0.123}\ \underbrace{\dfrac{hf - W_G}{kT_0}}_{x} - \underbrace{\dfrac{W_V - W_{Fp}}{kT_0}}_{0.5}\Big)\Big/\dfrac{T}{T_0}\Big]}\,.$$

The photon energy is expressed by the normalized quantity x,

$$x = \frac{hf - W_G}{kT_0}, \qquad x_0 = \frac{(W_{Fn} - W_{Fp}) - W_G}{kT_0} = 3.5\,. \tag{3.47}$$

The maximum spontaneous emission (see Fig. 3.15(a)) at $T \approx 0$ is located at a normalized frequency $x = x_0$, which corresponds to the difference of the quasi Fermi levels. It shifts for higher temperatures to lower frequencies (down to $x = 2.515$ at $T/T_0 = 0.52$ with the special assumptions of Eq. (3.46)). With a further temperature increase $T/T_0 > 0.52$ the maximum moves continuously

to larger frequencies. The spontaneous emission maximum for the reference temperature $T/T_0 = 1$ is at $x_{\mathrm{sp}} = 2.73$.

The induced net emission rate per photon $r_{\mathrm{ind}}^{(\mathrm{M})}/N_P = G/V$ and therefore the optical gain rate G is positive only for $0 < x < x_0$. It is zero at $x = x_0$ and becomes negative for $x > x_0 = 3.5$ because photons with energies $hf > W_{Fn} - W_{Fp}$ are absorbed, Fig. 3.15(b). At $T = 0$ the spectra of spontaneous emission and the gain rate are identical for $x < 3.5$ with an emission maximum at $hf = W_{Fn} - W_{Fp}$. However, with increasing T the maximum gain shifts to lower frequencies. For a fixed temperature the maximum gain is always at a lower frequency than the maximum spontaneous emission. At $T/T_0 = 1$ the maximum emissions are at $x_{\mathrm{sp}} = 2.73$ and $x_{\mathrm{ind}} = 1.95$, respectively. For GaAs at $T_0 = 293\,\mathrm{K}$ with a gain maximum at $\lambda = 0.842\,\mu\mathrm{m}$ the difference $x_{\mathrm{sp}} - x_{\mathrm{ind}} = 0.78$ at $T/T_0 = 1$ corresponds to a wavelength shift of $\Delta\lambda = 11.2\,\mathrm{nm}$.

For the chosen model of a k_μ-selection rule the spontaneous and stimulated emission spectra Fig. 3.15 exhibit a non-zero slope at $x = 0$ (this is also true for strict $\vec{k}_\mu$-selection). For the k_μ-selection rule[38] the ratio of the first derivatives $\mathrm{d}\, r_{\mathrm{sp}}^{(\mathrm{eM})}/\,\mathrm{d}\, x$ and $\mathrm{d}(r_{\mathrm{ind}}^{(\mathrm{M})}/N_P)/\,\mathrm{d}\, x$ amounts to n_{sp}. Without k-selection these slopes are zero, and the ratio of the second derivatives would be n_{sp}.

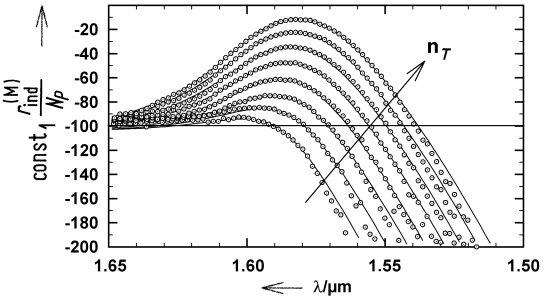


Fig. 3.16. Measured wavelength dependence of induced emission per photon for various carrier densities n_T. The multiplicative constant is different from Fig. 3.15 (after reference in Footnote 39).

Figure 3.16 displays a measured gain curve[39] of an InAlGaAs/InP semiconductor laser. The zero-slope at the bandgap wavelength $\lambda_G = 1.65\,\mu\mathrm{m}$ (see Table 3.2 on Page 56 for the InGaAsP/InP compound), and the shift of $W_{Fn} - W_{Fp}$ to smaller emission wavelengths (larger frequencies) when increasing the carrier concentration n_T are clearly visible.

The ratio of the spontaneous emission rate $r_{\mathrm{sp}}^{(\mathrm{eM})}(f)$ into one mode of frequency f and the total spontaneous emission rate r_{sp} into all possible modes corresponds to the quotient of the probability density $\rho(f) \le 2/(\pi \Delta f_H)$ that a microsystem with the required energy state is available, and the spectral density of all possible electromagnetic modes $\varrho_{\mathrm{tot}}(f)\, V$ inside the active volume V. The corresponding ratio for the induced emission rate $r_{\mathrm{ind}}^{(\mathrm{eM})}(f)$ into one mode of frequency f is larger by the photon number N_P (Eq. (3.39)),

$$\frac{r_{\mathrm{sp}}^{(\mathrm{eM})}(f)}{r_{\mathrm{sp}}} = \frac{\rho(f)}{\varrho_{\mathrm{tot}}(f)\, V}\,, \qquad \frac{r_{\mathrm{ind}}^{(\mathrm{eM})}(f)}{r_{\mathrm{sp}}} = \frac{r_{\mathrm{ind}}^{(\mathrm{eM})}(f)}{r_{\mathrm{sp}}^{(\mathrm{eM})}(f)}\, \frac{r_{\mathrm{sp}}^{(\mathrm{eM})}(f)}{r_{\mathrm{sp}}} = \frac{N_P}{1}\, \frac{\rho(f)}{\varrho_{\mathrm{tot}}(f)\, V}\,. \tag{3.48}$$

The left-hand side of Eq. (3.48) is an expression similar to Purcell's figure of merit[40] F_P. However, in Eq. (3.48) the relation $\rho(f) \ll \varrho_{\mathrm{tot}}(f)\, V$ holds normally, i. e., the emission linewidth "sees" a large number of modes where to emit. If the quotient $N_P/1$ of the induced and of the spontaneous emission rate into a mode f is reduced by the same factor $\rho(f)/\left(\varrho_{\mathrm{tot}}(f)\, V\right)$, the ratio of the

[38] See Footnote 36 on Page 69

[39] Wüst, F.: Optischer Gewinn und Alpha-Faktor in InAlGaAs/InP Quantenfilmlasern. PhD Thesis Karlsruhe 1999. — The material system offers a higher characteristic temperature T_0, a larger differential gain $\mathrm{d}G/\,\mathrm{d}n_T$ and a smaller linewidth enhancement factor α. This leads to a faster direct modulation capability than is possible for a InGaAsP/InP laser.

[40] See Eq. (3.38) in Footnote 33 on Page 69

induced emission rate into the mode f and the total emission probability into all modes results, right-hand side of Eq. (3.48).

By integrating the emission rate $r_{\mathrm{sp}}^{(\mathrm{eM})}(f)$ into one mode over all relevant modes the total spontaneous emission rate r_{sp} (unit $\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$) can be calculated ($r_{\mathrm{sp}}^{(\mathrm{eM})} \geq 0$ holds for $f \geq W_G/h$ only),

$$r_{\mathrm{sp}} \equiv \int_{-\infty}^{\infty} \underbrace{r_{\mathrm{sp}}\,\rho(f)}_{r_{\mathrm{sp}}(f)}\,\mathrm{d}f = \int_{-\infty}^{\infty} r_{\mathrm{sp}}^{(\mathrm{eM})}(f)\,\varrho_{\mathrm{tot}}(f)V\,\mathrm{d}f = \int_{W_G/h}^{\infty} r_{\mathrm{sp}}^{(\mathrm{eM})}(f)\,\varrho_{\mathrm{tot}}(f)V\,\mathrm{d}f\,. \tag{3.49}$$

The left-most equation in the chain exploits the identity $1 \equiv \int_{-\infty}^{\infty} \rho(f)\,\mathrm{d}f$, Eq. (3.37).

Radiative and nonradiative transitions

For a radiative transition, both an electron and a hole participate. Therefore it is plausible that the total spontaneous emission rate can be written as:

$$r_{\mathrm{sp}} = Bn_Tp, \quad B = \begin{cases} 1\times10^{-10}\ldots7\times10^{-10}\,\mathrm{cm}^3\,\mathrm{s}^{-1} & \text{(Ga,Al)As} \\ 8.6\times10^{-11}\,\mathrm{cm}^3\,\mathrm{s}^{-1} & \text{(In,Ga)(As,P)} \end{cases} \tag{3.50}$$

The smaller recombination coefficients B for (Ga,Al)As are valid for band-band transitions, the larger ones for transitions from the CB into non-ionized well localized (small spatial uncertainty Δx) shallow acceptor states which can provide a large difference momentum $\Delta(\hbar k) \geq (\hbar/2)/\Delta x$. For (In,Ga)(As,P) the temperature dependence is $B \sim 1/T^{\kappa}$ with $1 \leq \kappa \leq 1.5$. In thermal equilibrium the carrier concentrations Eq. (3.25) follow, i. e., $r_{\mathrm{sp}} = Bn_i^2$. In the non-equilibrium case the spontaneous emission rate is $r_{\mathrm{sp}} = B(n_Tp - n_i^2)$, but because normally $n_Tp \gg n_i^2$ holds, Eq. (3.50) is a good approximation to the actual case. Figure 3.17(a) shows schematically

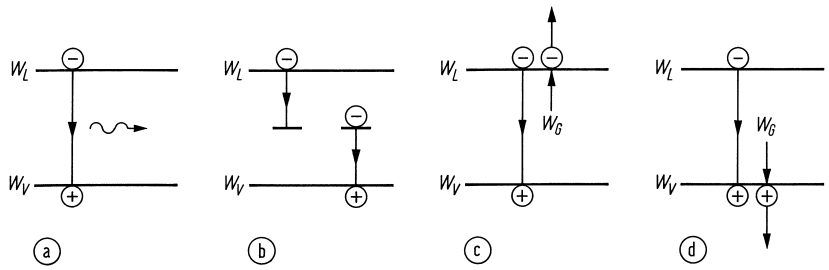


Fig. 3.17. Radiative and nonradiative transitions. (a) Radiative band-band transition. (b) Nonradiative transition via localized states in the forbidden band. (c) (d) Nonradiative Auger recombinations (recombination energy excites an electron in the CB or in the VB)

a radiative recombination. Figure 3.17(b) displays a nonradiative recombination (rate r_{ns}, unit $\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$) via localized impurities in the forbidden band (rate $r_{\ell\mathrm{S}}$; such impurities can help in shortening the lifetime, see the discussion in Sect. 3.1.2 on Page 52). Eventually, Fig. 3.17(c),(d) presents Auger processes (rate: r_{Au}), which are nonradiative. For (In,Ga)(As,P) the process Fig. 3.17(d) is important, while in (Ga,Al)As Auger processes are of no consequence. In summary we have:

$$r_{\mathrm{ns}} = r_{\ell\mathrm{S}} + r_{\mathrm{Au}} \qquad \begin{matrix} r_{\ell\mathrm{S}} = An_T \\ r_{\mathrm{Au}} = Cn_Tp^2 \end{matrix} \tag{3.51}$$

Measurements in (In,Ga)(As,P) result in coefficients $A = 1/(10\,\mathrm{ns})$ (undoped samples) up to $A = 1/(0.1\,\mathrm{ns})$ ($n_A = 2\times10^{18}\,\mathrm{cm}^{-3}$), and in $C = 4\times10^{-29}\,\mathrm{cm}^6\,\mathrm{s}^{-1}$ (calculated: $C = 10^{-27}\ldots10^{-31}\times\mathrm{cm}^6\,\mathrm{s}^{-1}$). The Auger coefficient C increases with temperature; also A increases slightly. The structure of $r_{\ell\mathrm{S}}$ follows from the proportionality $r_{\ell\mathrm{S}} \sim n_T$ because only one carrier type is involved. For

Auger recombination we have $r_{\mathrm{Au}} \sim n_T p \cdot p$ because two carrier types recombine, and additionally a hole (case of Fig. 3.17(d)) must be available to take over the excitation energy.[41] It is useful to define an effective recombination rate

$$r_{\mathrm{eff}} = r_{\mathrm{sp}} + r_{\mathrm{ns}} = r_{\mathrm{sp}} + r_{\ell\mathrm{S}} + r_{\mathrm{Au}}\,. \tag{3.52}$$

Inside the recombination zone of a diode (layer height d, cross-section area F) the carrier density changes if the injected carrier rate (injection current density J, elementary charge e) deviates from the recombination rate,

$$\frac{\mathrm{d}n_T}{\mathrm{d}t} = \frac{J}{ed} - r_{\mathrm{eff}}(n_T). \tag{3.53}$$

Strictly speaking, $r_{\mathrm{eff}}(n_T)$ in Eq. (3.53) should be replaced by $r_{\mathrm{eff}}(n_T) - r_{\mathrm{eff}}(n_{T\,\mathrm{equil}})$ for the correct solution at a concentration $n_{T\,\mathrm{equil}}$ for thermal equilibrium $J = 0$. With a step perturbation of the current density from J_0 to $J_0 + J_1$, a perturbation ansatz $n_T(t) = n_{T0} + n_{T1}(t)$ together with a series expansion of $r_{\mathrm{eff}} = r_{\mathrm{eff}\,0} + (\partial r_{\mathrm{eff}}/\partial n_T) n_{T1}$ at n_{T0} results in

$$n_{T1}(t) = \frac{J_1 \tau_{\mathrm{eff}}}{ed}\left(1 - \mathrm{e}^{-t/\tau_{\mathrm{eff}}}\right), \quad \text{with} \quad \tau_{\mathrm{eff}}^{-1} = \frac{\partial r_{\mathrm{eff}}}{\partial n_T}\,. \tag{3.54}$$

In an analogous form the (carrier concentration dependent) lifetimes for the other recombination processes may be defined. With Eqs. (3.50), (3.51) on calculates:

$$\left.\begin{aligned} \tau_{\mathrm{sp}}^{-1} &= \frac{\partial r_{\mathrm{sp}}}{\partial n_T} = B\left(p + n_T \frac{\partial p}{\partial n_T}\right), \\ \tau_{\ell\mathrm{S}}^{-1} &= \frac{\partial r_{\ell\mathrm{S}}}{\partial n_T} = A, \\ \tau_{\mathrm{Au}}^{-1} &= \frac{\partial r_{\mathrm{Au}}}{\partial n_T} = C\left(p^2 + 2 n_T p \frac{\partial p}{\partial n_T}\right), \end{aligned}\right\} \quad \begin{aligned} \tau_{\mathrm{eff}}^{-1} &= \tau_{\mathrm{sp}}^{-1} + \tau_{\mathrm{ns}}^{-1}\,, \\ \tau_{\mathrm{ns}}^{-1} &= \tau_{\ell\mathrm{S}}^{-1} + \tau_{\mathrm{Au}}^{-1}\,. \end{aligned} \tag{3.55}$$

The internal quantum efficiency η_{int} of radiative recombination is defined by

$$\frac{1}{\eta_{\mathrm{int}}} = \frac{\tau_{\mathrm{sp}}}{\tau_{\mathrm{eff}}} = 1 + \frac{\tau_{\mathrm{sp}}}{\tau_{\mathrm{ns}}} = 1 + \tau_{\mathrm{sp}}\left(\frac{1}{\tau_{\ell\mathrm{S}}} + \frac{1}{\tau_{\mathrm{Au}}}\right). \tag{3.56}$$

The smaller the effective lifetime τ_{eff} is, the faster the spontaneously emitted light can follow, see the discussion in Sect. 3.1.2 on Page 52. However, if nonradiative processes determine the lifetime, the internal quantum efficiency η_{int} deteriorates.

For moderate modulation frequencies, for which the modulation period is much smaller than τ_{eff} so that $\left(1 - \mathrm{e}^{-t/\tau_{\mathrm{eff}}}\right) \approx 1$ is a valid approximation, the generated radiation power amplitude $P_1 = (n_{T1} F d)\, hf/\tau_{\mathrm{sp}}$ follows the instantaneous current amplitude I_1, $P_1 \sim n_{T1} \sim J_1 F = I_1$. The quantity $n_{T1} F d$ is the number of spontaneously emitted photons,

$$P_1 = \frac{n_{T1} F d \cdot hf}{\tau_{\mathrm{sp}}} = \frac{J_1(t)\tau_{\mathrm{eff}}}{ed}\frac{hfFd}{\tau_{\mathrm{sp}}} = \eta_{\mathrm{int}} hf\,\frac{I_1}{e}, \qquad \eta_{\mathrm{int}} = \frac{P_1/(hf)}{I_1/e}\,. \tag{3.57}$$

For high current densities the current dependence of τ_{sp} and η_{int} (see Eq. (3.55)) leads to nonlinear distortions. The internal quantum efficiency η_{int} may be interpreted as the average number of photons per injected electron.

[41] For high electron injection $p \approx n_T$ the recombination rate follows the law $r_{\mathrm{Au}} \sim n_T^3$, i. e., it increases faster than the radiative emission rate Eq. (3.50). Therefore, (In,Ga)(As,P) lasers or LED must not be highly p-doped or operated at high current densities.

3.3 Light-emitting diode

3.3.1 Output power and modulation properties

Light-emitting diodes (LED) operate without end mirrors in a mode where the spontaneous emission rate $r_{\rm sp}^{\rm (eM)}$ dominates, Eq. (3.39). For communication purposes LED with double-heterostructures are common, Fig. 3.11. The generated light power P is given by Eq. (3.57),

$$P = \frac{n_T F d \cdot hf}{\tau_{\rm sp}} = \eta_{\rm int} hf \, \frac{I}{e}\,, \qquad \eta_{\rm int} = \frac{\tau_{\rm eff}}{\tau_{\rm sp}} = \frac{P/(hf)}{I/e}\,. \tag{3.58}$$

The mean photon energy hf of the emission line is slightly larger than the bandgap energy W_G. For a flat-diode configuration the power reflection factor at the boundary according to Eq. (2.35) is R_P assuming nearly perpendicular incidence and neglecting polarization effects. The radiation

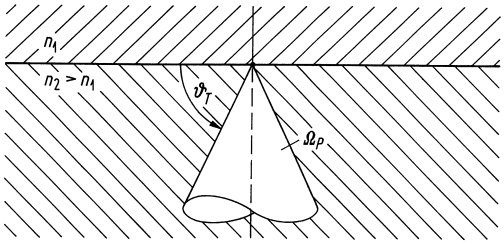


Fig. 3.18. Plane boundary between two media ($n_1, n_2 > n_1$ refractive indices, ϑ_T critical angle of total reflection, R_P power reflection factor). Only the fraction $(1 - R_P)$ of the radiation from the solid angle Ω_P is transmitted into the medium n_1.

is isotropically emitted into the full solid angle 4π, but only a fraction $(1 - R_P)\,\Omega_P/(4\pi)$ given by the critical solid angle Ω_P for total internal reflection (cone semi-angle $\pi/2 - \vartheta_T$) is transmitted into the medium with lower refractive index $n_1 < n_2$, Fig. 3.18. The optical efficiency $\eta_{\rm opt}$ describes the amount of usable light. The following numerical values apply for the radiation from a semiconductor into air (or fused silica):

$$\eta_{\rm opt} = \frac{\Omega_P}{4\pi}\,(1 - R_P) = 1.5\,\%\ (3.5\,\%) \qquad \begin{cases} \Omega_P = 2\pi(1 - \sin\vartheta_T) = 0.27\,{\rm sr}\ (0.54\,{\rm sr}), \\ \cos\vartheta_T = n_1/n_2 \quad\ = 73\,^\circ \quad (66\,^\circ), \\ R_P = \left(\dfrac{n_1 - n_2}{n_1 + n_2}\right)^2 \ = 32\ \% \quad (18\ \%). \end{cases} \tag{3.59}$$

The radiated output power P_a of the LED is

$$P_a = \eta_{\rm opt} P = \eta_{\rm ext} hf \, \frac{I}{e}, \quad \rightarrow \quad \eta_{\rm ext} = \frac{P_a/(hf)}{I/e} = \eta_{\rm opt}\, \eta_{\rm int}. \tag{3.60}$$

The external quantum efficiency is denoted as $\eta_{\rm ext}$. The voltage drop U over the pn-junction when an injection current I is flowing corresponds to the energetic distance of the quasi Fermi levels (Fig. 3.13), which is in the order of the photon energy, $eU = W_{Fn} - W_{Fp} \approx hf$, Eq. (3.13). With the electrical input power $P_{\rm el} = UI$ a total power-conversion or wall-plug efficiency $\eta_{\rm tot}$ may be defined by

$$P_a = \eta_{\rm tot} P_{\rm el} = \eta_{\rm ext} hf \, \frac{I}{e} \quad \rightarrow \quad \eta_{\rm tot} = \eta_{\rm ext} \frac{hf}{eU} \approx \eta_{\rm ext} = \frac{P_a/(hf)}{P_{\rm el}/(eU)} \quad \text{for } P_{\rm el} = UI\,. \tag{3.61}$$

For direct semiconductors we have typically $0.5 \le \eta_{\rm int} \le 0.9$. The larger values are found with moderate p-doping of the active layer, because then the electron number in the CB is reduced and inversion is more simply to achieve, see Footnote 28 on Page 62. The total conversion efficiency

for plane surface emitters is in the order $\eta_{\text{tot}} = \eta_{\text{opt}}\,\eta_{\text{int}} = 0.75 \ldots 3.1\,\%$. With operating currents up to 200 mA the emitted power reaches the 10 mW range.

For larger injection currents I the optical power P_a tends to saturate and even to diminish. For (In,Ga)(AsP) this would be true even at a constant junction temperature because of Auger recombination, but actually the increased heat enforces this effect by a thermally reduced B, by an increase of C (not for GaAs), and by the reduced carrier confinement in the double-heterostructure at elevated temperatures (less pronounced in GaAs because of larger barriers than in (In,Ga)(AsP)). The temperature coefficient $c_X = (1/X)(\mathrm{d}X/\,\mathrm{d}T)$ of the power amounts to $c_{Pa} = -1.4 \times 10^{-2}\,\mathrm{K}^{-1}$ for GaAs and $c_{Pa} = -2 \times 10^{-2}\,\mathrm{K}^{-1}$ for (In,Ga)(As,P).

A small signal (perturbation) ansatz $g = g_0 + g_1(\omega)\,\mathrm{e}^{\mathrm{j}\,\omega t}$ for n_T, J in Eq. (3.53) and for P, I, P_a in Eq. (3.58), (3.60) together with the effective lifetime τ_{eff} Eq. (3.54) leads to the spectral relation

$$P_{a1}(\omega) = \eta_{\text{ext}} h f\,\frac{I_1(\omega)}{e}\,\frac{1}{1 + \mathrm{j}\,\omega\tau_{\text{eff}}}\,. \tag{3.62}$$

For a constant modulation current amplitude $|I_1(\omega)|$ we find the current-power transfer function

$$\left|\frac{P_{a1}(\omega)}{P_{a1}(0)}\right| = \frac{1}{\sqrt{1 + \omega^2 {\tau_{\text{eff}}}^2}} \qquad \rightarrow \quad \omega_c = \frac{1}{\tau_{\text{eff}}}\,. \tag{3.63}$$

The angular 3-dB cutoff frequency at $|P_{a1}(\omega_c)/P_{a1}(0)| = 1/\sqrt{2}$ is denoted as ω_c. The photodetector current is proportional to the received optical power, $i_1(\omega) \sim P_{a1}(\omega)$. The angular frequency ω_c corresponds to the half-power point of the received signal power $i_1^2(\omega) \sim |P_{a1}(\omega)|^2$ at the angular modulation frequency ω. The signal amplitude increases with the efficiencies and thus decreases with the cutoff frequency, Eqs. (3.62), (3.60), (3.56),

$$i_1(\omega) \sim P_{a1}(0) \sim \eta_{\text{ext}} \sim \eta_{\text{int}} \sim \tau_{\text{eff}} \sim \frac{1}{\omega_c}. \tag{3.64}$$

For a given material the cutoff angular frequency ω_c may be increased only by forcing nonradiative recombination at the cost of efficiency. For high-speed LED $f_c = 1\,\mathrm{GHz}$ is possible. A further decrease of τ_{eff} is counterproductive because eventually the LED junction capacitance C in combination with the source resistance R fixes the time constant $\tau = RC$.

3.3.2 Devices

The surface emitter (LED) and the edge emitter (ELED) are the two basic device configurations to couple the LED light output into a small-diameter glass fibre.

Surface emitter

For the surface emitter Fig. 3.19 the emitting area of the junction is confined by oxide isolation, and the contact is usually $15 \ldots 100\,\mu\mathrm{m}$ in diameter. The active p-GaAs layer is part of a 3-layer heterostructure. The device is known as a Burrus diode.[42] The n-GaAs substrate is thinned by etching to reduce the absorption. This is not required in an (In,Ga)(As,P) system because the InP substrate is transparent having a wider bandgap than the active (In,Ga)(As,P) layer. The smaller the junction area F is, the better the heat can be removed, the higher the current density can be chosen, and the brighter the emitted light will be. Depending on the angle γ measured from an axis perpendicular to the emitting surface, the apparent radiating area changes according to $F\cos\gamma$. The radiance L is defined as the differential power $\mathrm{d}P$ radiated from a differential apparent area $\mathrm{d}F\cos\gamma$ into a differential solid angle $\mathrm{d}\Omega$ centred at an angle γ,

$$L = \frac{\mathrm{d}^2 P}{\mathrm{d}F\cos\gamma\,\mathrm{d}\Omega}\,, \qquad \frac{\mathrm{d}P}{\mathrm{d}\Omega} = \mathrm{d}F\cos\gamma\,. \tag{3.65}$$

[42] Burrus, C. A.; Miller, B. I.: Small-area DH Al-Ga-As electroluminescent diode sources for optical fiber transmission lines. Opt. Commun. 4 (1971) 307–309

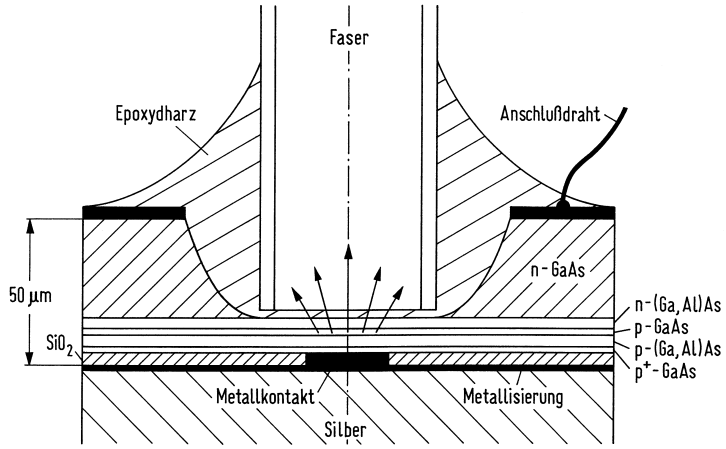


Fig. 3.19. Small-area high-radiance (Ga,Al)As double-heterostructure surface-emitting LED with attached fibre (Burrus diode). Epoxydharz = epoxy resin, Anschlußdraht = bond wire, Metallkontakt = metal contact, Metallisierung = metallization

An emitter with a constant radiance L and a $\cos\gamma$-dependent far-field power distribution P is called a Lambertian source. Its half-power width is $\Delta\gamma_H = 120\,°$. Because of the large emission angle, LED light may be coupled efficiently only to multimode waveguides.

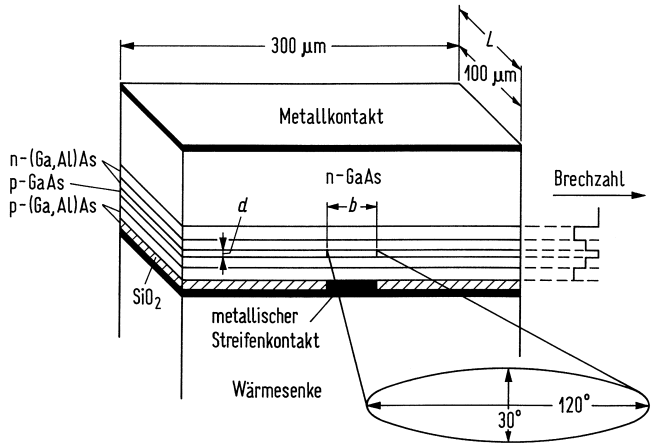


Fig. 3.20. Edge-emitting double-heterostructure LED. L, b, d are length, width and thickness of the active zone. Metallkontakt = metal contact, metallischer Streifenkontakt = metallic contact strip, Wärmesenke = heat sink, Brechzahl = refractive index

Edge emitter

The edge emitter is shown in Fig. 3.20. The active layer is very thin, $d = 0.05 \ldots 0.1\,\mu$m. Together with the adjacent layers of reduced refractive index this 5-layer double-heterostructure forms a vertically singlemoded strip waveguide. The lateral width of $b = 10 \ldots 50\,\mu$m is effectively defined by the carrier injection from the contact stripe. Because of the vertical waveguiding the far-field half-power width *in vacuum* is less than 120 °, typically $\Delta\gamma_H = 0.52 \mathrel{\widehat{=}} 30\,°$. For a single-mode slab waveguide with a normalized frequency $V = (d/2)k_0 A_N$ and $V \le V_{1G} = \pi/2$ (Eqs. (2.38), (2.52)) a vacuum far-field angle $\gamma_N \approx \sin\gamma_N = A_N = (V_{1G}/\pi)(\lambda/d) = \frac{1}{2}\lambda/d$ would be expected. With $\Delta\gamma_H = 2\gamma_N \approx \lambda/d_{\text{eff}}$ and $\lambda = 1.3\,\mu$m an associated effective vertical field extension d_{eff}

would be in the order of the vacuum wavelength, $d_{\text{eff}} = \lambda/\Delta\gamma_H = 2.5\,\mu$m. This is larger than the actual significant field extension *in the waveguide* $x_M = 1.8\,d = 0.18\,\mu$m (Eq. (2.53) on Page 26) for $h \mathrel{\widehat{=}} d = 0.1\,\mu$m. While immediately at the semiconductor-vacuum interface the transverse field extension is x_M, a few wavelength away from the interface into the vacuum the effective field extension is of the order of λ, because of diffraction.

Laterally the ELED behaves as a Lambertian source. The device radiates perpendicular to the main current flow direction through the cleaved end faces. To increase the efficiency, one endface may be coated to yield a very high reflection factor, while the other end face may be antireflection-coated. Typical device lengths are $L = 100\,\mu$m.

The narrow contact stripe allows high current densities. Therefore, and because of the spatially coherent field in the vertical waveguide, the radiance of edge-emitters is larger (up to $L = 1\,000\,\text{W}\,\text{cm}^{-2}\,\text{sr}^{-1}$) than for surface emitters, and a more efficient power coupling into single-mode fibres becomes possible.

Superluminescent diode

With higher current densities and larger lengths up to about $L = 500\,\mu$m induced amplification of the spontaneously emitted light can become important. Such a device is called a superluminescent LED (SLED). Because the effective carrier lifetime τ_{eff} is reduced by stimulated emission, Eq. (3.8), the maximum modulation frequency $f_c = 1/(2\pi\tau_{\text{eff}})$ increases without paying an output power penalty, Eqs. (3.63), (3.64). Simultaneously, the emission linewidth becomes smaller because of the frequency-selective amplification, i. e., the temporal coherence becomes larger. The SLED light may be coupled to external waveguides as well as it this the case for an ELED.

3.3.3 LED spectrum

The detailed distribution of the carriers into states of the CB and VB depends on the Fermi level, i. e., on the impurities in the active layer and on the injection current. With increasing temperature the Fermi function changes, and the spontaneous emission maximum shifts first to lower, then to higher frequencies, Fig. 3.15 on Page 71. The bandgap energy decreases with increasing temperature. Both effects together result in a shift of the emission maximum to lower frequencies (larger wavelengths) by a rate of $0.2\,\text{nm}\,\text{K}^{-1}$ for GaAs and $0.4\,\text{nm}\,\text{K}^{-1}$ for (In,Ga)(As,P) near $\lambda = 1.3\,\mu$m.

Figure 3.7(a) on Page 59 shows the bandstructure of a direct semiconductor. The spectral width of the emission is determined mainly by the energetic distribution of the carriers in the bands. The occupation probabilities are determined by the Fermi function, Fig. 3.8 on Page 61. The quasi Fermi levels $f(W_{Fn,p}) = 1/2$ are usually near the bandedge, $W_{Fn,p} \approx W_{C,V}$. The Fermi function changes significantly from $f(W_{Fn,p} - 2kT_0) = 0.88$ to $f(W_{Fn,p} + 2kT_0) = 0.12$ in an energetic interval $4kT_0$ around the Fermi energy. For both GaAs and InP, the curvatures of the CB and VB in Fig. 3.7(b) on Page 59, i. e., the reciprocal effective carrier masses in Eq. (3.17) and in Table 3.3 on Page 60, are significantly different, $m_n \ll m_p$. Compared to the CB, the VB is virtually flat. Therefore, photons are emitted in the spectral range $W_G \le hf \le (W_C + 2kT_0) - W_V$. The total spectral width of the emission essentially amounts to $h\,\Delta f_H = 2kT_0$,

$$h\,\Delta f_H = 2\,kT_0 = 50\,\text{meV}, \quad \Delta f_H = 12.1\,\text{THz} \quad \text{at room temperature } T_0 = 293\,\text{K}\,. \tag{3.66}$$

For GaAs we have $\Delta\lambda_H = 30\,\text{nm}$, for (In,Ga)(As,P) $\Delta\lambda_H = 70\,\text{nm}$ at $\lambda = 1.3\,\mu$m. The quantity Δf_{gain} is also an estimate for the amplification bandwidth of semiconductor laser devices. Basically, it corresponds to the width Δf_H of the lineshape $\rho(f)$, Eq. (3.37) on Page 68.

3.4 Laser diode

3.4.1 Basic relations

A basic laser diode (LD) that has a rectangular cavity is equivalent to a Fabry-Perot (FP) resonator, Fig. 3.1 on Page 47 and Fig. 3.6 on Page 58, and is thus called a Fabry-Perot laser diode (FP LD). The structure is similar to the one of the edge-emitter Fig. 3.20 having a laser-active volume $V = dbL$ with dimensions $d = 0.1 \ldots 0.2\,\mu$m (vertical, x-direction), $b = 2 \ldots 5\,\mu$m (lateral, y-direction) and $L = 300 \ldots 1\,200\,\mu$m, (longitudinal, z-direction).

Waveguiding properties and resonances The transverse waveguiding mechanism is described by an effective refractive index $n_e < n$ (Eq. (2.37)) which is smaller than n in Fig. 3.1. To avoid complicated subscripts we drop the index e for convenience, and implicitly regard the propagation quantities as effective waveguide quantities. This applies, e. g., to the (complex) propagation constant, to the modal loss α_V in Eq. (3.78), and to the longitudinal mode spacing Δf_z.

For plane waves (Eq. (2.16) on Page 10) propagating along the z-axis ($k_{x,y} = 0$, $k_z = k$) the longitudinal resonances are given by

$$k \cdot 2L = k_0 n \cdot 2L = \omega n \cdot 2L/c = m_z \cdot 2\pi\,, \quad m_z = 1, 2, 3, \ldots \tag{3.67}$$

Regarding m_z for the moment as a continuous variable, we differentiate $fn = m_z c/(2L)$ with respect to m_z resulting in

$$\frac{\mathrm{d}(fn)}{\mathrm{d}m_z} = \frac{\mathrm{d}f}{\mathrm{d}m_z}\, n + f\, \frac{\mathrm{d}n}{\mathrm{d}f}\, \frac{\mathrm{d}f}{\mathrm{d}m_z} = \frac{\mathrm{d}f}{\mathrm{d}m_z} \left(n + f\, \frac{\mathrm{d}n}{\mathrm{d}f} \right) = \frac{c}{2L}\,.$$

Recalling the discrete nature of m_z, the replacements $\mathrm{d}m_z \to 1$ and $\mathrm{d}f \to \Delta f_z$ are appropriate. Introducing the group index n_g, the group velocity v_g (Eq. (2.21)) and the photon round-trip time τ_U, this leads to the equidistant longitudinal mode spacing (free spectral range FSR, see also Eq. (3.2) on Page 48),

$$\Delta f_z = \frac{c}{2 n_g L} = \frac{v_g}{2L} = \frac{1}{\tau_U}\,. \tag{3.68}$$

The typical comb structure of the spectrum is displayed in Fig. 3.26 on Page 93.

Field confinement factor The field energy is not concentrated in the active volume V alone, because the transversely evanescent field extends into the cladding of the waveguide Fig. 3.1. For the vertical slab waveguide Fig. 2.7 the field dependence is depicted in Fig. 2.8 ($m = 0$). The extent of the field concentration is given by the field confinement factor Γ, Eq. (2.73). For the TE_0 mode of the slab waveguide it is defined as

$$\Gamma_{\mathrm{TE}} = \int_{-d/2}^{+d/2} |E_y(x)|^2\, \mathrm{d}x \Big/ \int_{-\infty}^{+\infty} |E_y(x)|^2\, \mathrm{d}x\,. \tag{3.69}$$

An approximation valid for all V (Eq. (2.38)) with a maximum error of 1.5 % is[43]

$$\Gamma_{\mathrm{TE}} = \frac{2V^2}{1 + 2V^2}, \qquad V = \frac{d}{2}\, k_0 \sqrt{n_1^2 - n_2^2}. \tag{3.70}$$

The field confinement Γ for the TE_0 mode is slightly larger than the one for the TM_0 mode, $\Gamma_{\mathrm{TE}} > \Gamma_{\mathrm{TM}}$, see Sect. 2.3.2 on Page 23. An example for a laser is $\Gamma_{\mathrm{TE}} = 0.184$ and $\Gamma_{\mathrm{TM}} = 0.145$, for a laser amplifier it is $\Gamma_{\mathrm{TE}} = 0.3$ and $\Gamma_{\mathrm{TM}} = 0.25$. For $d = 0.1 \ldots 0.2\,\mu$m the values $\Gamma = 0.2 \ldots 0.6$ are typical. This and the larger endface reflection factor for TE polarization, Fig. 2.6, is the reason why diode lasers usually oscillate in TE polarization ($\vec{E}$ parallel to y–axis).

[43] Botez, D.: Analytical approximation of the radiation confinement factor for the TE_0 mode of a double heterojunction laser. IEEE J. Quantum Electron. QE-14 (1978) 230–232

Emission and absorption rates With respect to Γ, the equations for induced and spontaneous emission as well as the total mode number and density M_{tot} and ϱ_{tot} have to be modified. The oscillating laser mode having a photon number N_P fills effectively a volume V/Γ which is larger than the active volume V. This increases the total mode number M_{tot} and mode density ϱ_{tot}, because in Eqs. (3.4), (3.7) the active volume V has to be replaced by the effective mode volume V/Γ. The number of photons interacting with the active medium is reduced to ΓN_P, and the equivalent energy density for spontaneous and induced transitions ($(N_P+1)\,hf\,/V$ (see Eq. (3.36), (3.39)) should be replaced by $\Gamma(N_P+1)\,hf\,/V$. The transverse waveguiding mechanism is described by an effective refractive index $n_e < n$ (Eq. (2.37)) which is smaller than n in Fig. 3.1. In summary, we have to substitute in Eqs. (3.43), (3.44), (3.4), (3.7)

$$\begin{aligned} r_{\mathrm{ind}}^{(\mathrm{M})} &\longrightarrow r_{\mathrm{ind}\,e}^{(\mathrm{M})} = \Gamma r_{\mathrm{ind}}^{(\mathrm{M})}, \qquad M_{\mathrm{tot}} \longrightarrow M_{\mathrm{tot}\,e} = M_{\mathrm{tot}}/\Gamma\,, \\ r_{\mathrm{sp}}^{(\mathrm{eM})} &\longrightarrow r_{\mathrm{sp}\,e}^{(\mathrm{eM})} = \Gamma r_{\mathrm{sp}}^{(\mathrm{eM})}, \qquad \varrho_{\mathrm{tot}} \longrightarrow \varrho_{\mathrm{tot}\,e} = \varrho_{\mathrm{tot}}/\Gamma\,. \end{aligned} \tag{3.71}$$

The inversion factor n_{sp} Eq. (3.45) remains unchanged because of Eq. (3.44). The same is true for the total spontaneous emission rate r_{sp} Eq. (3.49), and as a consequence also for the effective recombination rate $r_{\mathrm{eff}} = r_{\mathrm{sp}} + r_{\mathrm{ns}} = r_{\mathrm{sp}} + r_{\ell\mathrm{S}} + r_{\mathrm{Au}}$ Eq. (3.52),

$$\begin{aligned} n_{\mathrm{sp}} \longrightarrow n_{\mathrm{sp}\,e} = n_{\mathrm{sp}}\,, \quad & r_{\mathrm{sp}} \longrightarrow r_{\mathrm{sp}\,e} = r_{\mathrm{sp}}\,, \\ & r_{\mathrm{eff}} \longrightarrow r_{\mathrm{eff}\,e} = r_{\mathrm{eff}}\,. \end{aligned} \tag{3.72}$$

Gain and loss Actually, the resonator is longitudinally multimoded, Eq. (3.68). We describe the modes by plane waves with effective propagation properties and a complex (effective) refractive index $\bar{n} = n - \mathrm{j}\,n_i$ with real part n and imaginary part $-n_i$ (Eq. (2.6), dropping the subscript e as discussed in the beginning of Sect. 3.4.1 on Page 79),

$$\exp(-\mathrm{j}\,\bar{k}z), \qquad \left\{ \begin{array}{l} \bar{k} = k_0\bar{n} = k + \frac{1}{2}\,\mathrm{j}\,(g - \alpha_V), \\ \bar{n} = n - \mathrm{j}\,n_i, \\ k_0 = \omega/c, \end{array} \right\}, \qquad g - \alpha_V = -2\,k_0 n_i. \tag{3.73}$$

The quantities g, α_V are the modal power gain and loss constants corresponding to the net effective gain rate ΓG due to band-band transitions, and a power loss time constant $1/\tau_V$ to be discussed in the following which does not include band-band transitions.

According to Eq. (3.73) the wave experiences a net power gain of $\exp\left[(g - \alpha_V)2L\right]$ for a round-trip of length L between the resonator mirrors with power reflection coefficients $R_{1,2}$. Because of the mixture of length-distributed gain and localized mirror losses one is usually not interested in keeping track of how many times the light goes back and forth for amplification. Instead of using the gain per length it is then a more practical approach to define an equivalent gain per time.

At each partially transparent mirror the localized losses are equivalently described by a power "gain" $R_1R_2 = \exp\left(-\alpha_{R1}2L\right)\exp\left(-\alpha_{R2}2L\right) = \exp\left(-\alpha_R 2L\right)$ distributed over a round-trip through the resonator. The gain rate G (see Eq. (3.43) on Page 70) specifies the number of photons generated per second. The total losses of photons per second $1/\tau_P$ are described by the photon lifetime τ_P. The round-trip time $\tau_U = 2L/v_g$ for a photon can be computed from its group velocity v_g (see Eq. (2.21) and the following section). Assuming a constant gain rate per round-trip time τ_U, the net increase in photon number per time including all losses is

$$G - \frac{1}{\tau_P} = \frac{1}{N_P}\frac{\mathrm{d}N_P}{\mathrm{d}t}\,, \qquad \frac{N_P(\tau_U)}{N_P(0)} = \exp\left[\left(G - \frac{1}{\tau_P}\right)\tau_U\right], \qquad \tau_U = \frac{2L}{v_g}\,. \tag{3.74}$$

Taking into regard the loss mechanisms discussed above, the following relations hold between the gain rate Eq. (3.74) and the modal power gain Eq. (3.73),

$$\begin{aligned} \exp\big[(G - 1/\tau_P)\tau_U\big] &= \exp\big[(G - 1/\tau_V - 1/\tau_R)\tau_U\big] \\ &= R_1R_2\exp\left[(G - 1/\tau_V)\tau_U\right] = R_1R_2\exp\left[(G - 1/\tau_V)2L/v_g\right] \\ &= R_1R_2\exp\big[(g - \alpha_V)2L\big] \\ &= \exp\big[(g - \alpha_V - \alpha_R)2L\big] = \exp\big[(g - \alpha_V - \alpha_{R1} - \alpha_{R2})2L\big]. \end{aligned} \tag{3.75}$$

Comparing the various forms in Eq. (3.75) we find

$$\begin{aligned} G &= v_g g, \\ 1/\tau_V &= v_g \alpha_V, \\ 1/\tau_{R1,2} &= v_g \alpha_{R1,2} = -v_g \ln R_{1,2}/(2L), \\ 1/\tau_R &= v_g \alpha_R \ = -v_g \ln(R_1 R_2)/(2L), \\ 1/\tau_P &= v_g(\alpha_V + \alpha_R) = v_g[\alpha_V - \ln(R_1 R_2)/(2L)]. \end{aligned} \tag{3.76}$$

With the same reasoning as above, the net material gain rate $G = v_g g$ must be replaced by the net modal gain rate $\Gamma G = v_g \Gamma g$,.

$$G \longrightarrow G_e = \Gamma G\,, \quad g \longrightarrow g_e = \Gamma g\,. \tag{3.77}$$

The resonant mode is attenuated mainly in the active zone (a background material power loss constant α_V *not* including band-band transitions) and by the adjacent heterolayers (power loss constant $\alpha_{\rm het}$). Additionally, interface scattering and substrate losses could have some influence (power loss constant $\alpha_{\rm add}$). So the material loss has to be replaced by the modal loss,

$$\alpha_V \longrightarrow \alpha_{Ve} = \Gamma\alpha_V + (1-\Gamma)\alpha_{\rm het} + \alpha_{\rm add}\,. \tag{3.78}$$

Gain model The (effective) modal loss constant Eq. (3.78) ($\alpha_V \approx \alpha_{\rm het}$, $\alpha_{\rm add} = 0$, $\alpha_{Ve} \approx \alpha_V$) is typically in the order of $\alpha_V = 20 \ldots 50\,{\rm cm}^{-1}$. The necessary threshold gain constant $\Gamma g = \alpha_V + \alpha_R$ is in the region $\Gamma g = 25 \ldots 90\,{\rm cm}^{-1}$. For GaAs and at the spectral gain maximum (see Fig. 3.15(b)) the approximate n_T-dependency is

$$g = g_0 \times (n_T/n_t - 1), \quad g_0 = 330\,{\rm cm}^{-1}, \quad n_t = 1.1 \times 10^{18}\,{\rm cm}^{-3}\,. \tag{3.79}$$

The carrier concentration for a zero net gain constant $g = 0$ is called the transparency carrier concentration n_t. A typical threshold carrier density amounts to $n_{TS} = 1.2 \times 10^{18} \ldots 1.4 \times 10^{18}\,{\rm cm}^{-3}$. The refractive index of the active layer is about $n = 3.5$, Table 3.2, the group refractive index is in the region $n_g = 3.75 \ldots 5$.

The laser oscillates near the frequency f_0 of the maximum spectral gain. However, a larger injected carrier number leads to a nonlinear gain compression, because the energy states near the laser resonance energy hf_0 deplete due to hot carrier effects and spectral hole burning.[44] To fill the depleted states it needs the intraband relaxation time $\tau_{\rm CB}$ (Page 63), and this represents a "bottleneck" for the number of carriers available in energy states near hf_0. Phenomenologically, the nonlinear gain compression is modeled by a photon-number dependent decrease of the gain described by a gain compression factor ε_G. With the differential gain G_d and the transparency concentration n_t the optical gain is

$$G(n_T, N_P) = \frac{G(n_T)}{1 + \varepsilon_G \frac{\Gamma N_P}{V}} = G_d \frac{n_T - n_t}{1 + \varepsilon_G \frac{\Gamma N_P}{V}}\,. \tag{3.80}$$

[44] Schuster, S.; Haug, H.: Calculation of the gain saturation in cw semiconductor lasers with Boltzmann kinetics for Coulomb and LO phonon scattering. Semicond. Sci. Technol. 10 (1995) 281–289

„Spectral hole burning means the formation of a dip in the carrier distribution function around the laser resonance due to the finite intraband scattering time.

The stimulated emission heats the CB carriers by removing cool particles [between the bandedge and the quasi Fermi energy W_{Fn}], because the energy of the recombining electron-hole pairs is smaller than the average pair energy which is roughly the difference of the quasi Fermi energies $W_{Fn} - W_{Fp}$.

The phenomenological gain saturation coefficient ε_G stems from the finite intraband scattering time which yields an increase of the CB carrier density with increasing pump current. An injected hot electron-hole pair needs a finite time for the scattering into the laser resonance [energy hf] where stimulated recombination can occur. Therefore a higher pump current does not only result in a higher light intensity but also in a more pronounced non-equilibrium carrier distribution function with particle depletion around resonance (spectral hole burning), an increasing average carrier energy (carrier heating) and a growing density due to this kinetic 'bottleneck'.

For smaller pump rates the gain saturation calculated within the microscopic model [treating the carrier distributions in terms of the Boltzmann collision integral] decreases, which indicates that in this regime the kinetic limitation of the pump efficiency due to the carrier scattering into the laser resonance is not yet fully established."

The gain compression factor is in the order of $\varepsilon_G = 2.5 \times 10^{-17} \ldots 3.1 \times 10^{-17}\,\mathrm{cm}^3$. In the case of an empty VB (very high and constant hole concentration $p = n_A$, i. e., $f_V \approx 0$ in the range of interest and complete inversion $n_{\mathrm{sp}} = 1$), the transparency concentration becomes zero, $n_t = 0$, because the slightest electron concentration in the CB already establishes some gain. This can be also seen from Eq. (3.39), (3.41). If we further neglect gain compression, $\varepsilon_G = 0$, a linear gain dependency follows,

$$G(n_T) = G_d\, n_T\,. \tag{3.81}$$

3.4.2 Rate equations

The operating characteristics of a semiconductor laser are well described by a set of rate equations that govern the interaction of photons and electrons inside the active region. A rigorous derivation starts from Maxwell's equations and includes in a semi-classical approach the quantum mechanical calculation of light-matter interaction. If spontaneous emission is to be included rigorously, quantum electrodynamics becomes involved where the optical field is quantized, too.

The rate equation can also be written heuristically by considering the phenomena through which the number N_P of photons and electrons $n_T V$ change with time inside the active volume V. We assume that the valence band is practically empty of electrons. This is true if — as with GaAs and InP — the curvature of both the CB and VB, i. e., the reciprocal effective carrier masses in Eq. (3.17) and the effective DOS $N_{C,V}$ in Table 3.3 on Page 60, are significantly different, $m_n \ll m_p$ and $N_C \ll N_V$, see the discussion of Eq. (3.66) on Page 78. Therefore the hole concentration in the VB is large and virtually invariant, $\partial p/\partial n_T \approx 0$, so that $\tau_{\mathrm{sp}}^{-1} = \partial r_{\mathrm{sp}}/\partial n_T \approx Bp$ in Eq. (3.55) holds, and the spontaneous recombination rate $r_{\mathrm{sp}} = Bn_T p$ of Eq. (3.50) may be approximated by $r_{\mathrm{sp}} \approx n_T/\tau_{\mathrm{sp}}$. An equivalent procedure approximates the effective recombination rate r_{eff} from Eq. (3.52) by $r_{\mathrm{eff}} \approx n_T/\tau_{\mathrm{eff}}$. The result is

$$\begin{aligned} r_{\mathrm{sp}} &= Bn_T p \approx n_T/\tau_{\mathrm{sp}}\,, \\ r_{\mathrm{eff}} &= r_{\mathrm{sp}} + r_{\ell\mathrm{S}} + r_{\mathrm{Au}} \approx n_T/\tau_{\mathrm{eff}} \\ &\quad \text{if } N_C \ll N_V \text{ and therefore} \end{aligned} \qquad \begin{aligned} \tau_{\mathrm{sp}}^{-1} &= \partial r_{\mathrm{sp}}/\partial n_T \approx Bp\,, \\ \tau_{\ell\mathrm{S}}^{-1} &= \partial r_{\ell\mathrm{S}}/\partial n_T = A\,, \\ \tau_{\mathrm{Au}}^{-1} &= \partial r_{\mathrm{Au}}/\partial n_T \approx Cp^2, \\ \tau_{\mathrm{eff}}^{-1} &= \tau_{\mathrm{sp}}^{-1} + \tau_{\ell\mathrm{S}}^{-1} + \tau_{\mathrm{Au}}^{-1}\,. \end{aligned} \tag{3.82}$$

The lifetimes τ_{sp} and τ_{eff} depend on electron and hole concentrations, if the hole concentration p changes noticeably with the electron concentration n_T. For a longitudinally and laterally single-moded laser, the rate equations take the form[45].

$$\begin{aligned} \underbrace{\frac{\mathrm{d}N_P}{\mathrm{d}t}}_{\substack{\text{change of photon}\\ \text{number per time}}} &= + \underbrace{N_P \Gamma G(n_T, N_P)}_{\substack{\text{stimulatedly generated}\\ \text{photons per time}}} + \underbrace{Q\,\frac{n_T V}{\tau_{\mathrm{eff}}}}_{\substack{\text{spontaneously generated}\\ \text{photons per mode and time}}} - \underbrace{\frac{N_P}{\tau_P}}_{\substack{\text{stimulatedly depleted}\\ \text{photons per time}}}, \\ \underbrace{\frac{\mathrm{d}(n_T V)}{\mathrm{d}t}}_{\substack{\text{change of electron}\\ \text{number per time}}} &= - \underbrace{N_P \Gamma G(n_T, N_P)}_{\substack{\text{stimulatedly depleted}\\ \text{electrons per time}}} - \underbrace{\frac{n_T V}{\tau_{\mathrm{eff}}}}_{\substack{\text{spontaneoulsy depleted}\\ \text{electrons per time}}} + \underbrace{\frac{I}{e}}_{\substack{\text{injected electrons}\\ \text{per time}}}. \end{aligned} \tag{3.83}$$

The first equation (3.83) means in words: The number of photons N_P increases through photons which are generated by stimulated emissions with a net gain rate ΓG, and it increases through photons generated by spontaneous recombinations of electrons at a rate $1/\tau_{\mathrm{eff}}$, where only a fraction Q leads to spontaneous emissions into the mode under consideration. Further, the photon number decreases with a rate $1/\tau_P$ determined by the photon lifetime τ_P from Eq. (3.76).

The second equation (3.83) has to be read as follows: The number of electrons $n_T V$ in the active volume V decreases through electrons which recombine when stimulated by photons existing in

[45] See Sect. 3.5.4 Eq. (3.116) on Page 183 in reference Footnote 9 on Page 49

the mode under consideration, therefore a stimulated increase of the photon number corresponds to an equivalent decrease of the electron count. The number of carriers is further depleted with a rate $1/\tau_{\text{eff}}$ determined by the effective electron lifetime τ_{eff}, which takes into regard radiative (τ_{sp}) and nonradiative recombinations ($\tau_{\ell\text{S}}^{-1}$, τ_{Au}^{-1}) according to Eq. (3.55). Finally, the charge carrier number increases at a rate I/e fixed by the injection current I.

Because of stimulated ($N_P \Gamma G$) and spontaneous recombinations ($Q\, n_T V/\tau_{\text{eff}}$) the rate equations for the photon number N_P and for the CB carrier number $n_T V$ are nonlinear and coupled.

Electrons deplete spontaneously at a rate $1/\tau_{\text{eff}}$. The spontaneous emission factor Q tells how many spontaneous recombinations actually lead to photons being emitted into the mode under consideration. Taking into account the field confinement factor Γ from Eq. (3.69), (3.70), and looking at the translation relations Eq. (3.71), (3.72), the ratio of the spontaneous radiative recombination rate $\Gamma r_{\text{sp}}^{(\text{eM})}$ into one mode and the effective recombination rate r_{eff} defines Q. Using Eq. (3.48), (3.82) one finds[46]

$$Q = \frac{\Gamma r_{\text{sp}}^{(\text{eM})}}{r_{\text{eff}}} = \frac{\Gamma r_{\text{sp}}}{r_{\text{eff}}} \frac{\rho(f)}{\varrho_{\text{tot}}(f) V} = \Gamma \frac{\tau_{\text{eff}}}{\tau_{\text{sp}}} \frac{\rho(f)}{\varrho_{\text{tot}}(f) V}\,. \tag{3.84}$$

Lasing threshold Neglecting spontaneous emission ($Q = 0$) and assuming $N_P G \ll n_T V/\tau_{\text{eff}}$ we define the *lasing threshold* (subscript S) using Eq. (3.83) for the case $\mathrm{d}\,/\,\mathrm{d}t = 0$ with τ_P from Eq. (3.76); above threshold the device starts oscillating as described in Sect. 3.1.2 on Page 51,

$$\begin{aligned} \Gamma G(n_{TS}, 0) = \Gamma G_S &= \frac{1}{\tau_P} = v_g \Big[\alpha_V - \frac{\ln(R_1 R_2)}{2L}\Big], \\ \frac{I_S}{e} &= \frac{n_{TS} V}{\tau_{\text{eff}}} = r_{\text{eff}} V. \end{aligned} \tag{3.85}$$

It is for the threshold carrier concentration $n_T = n_{TS}$ that the net gain rate ΓG_S just compensates the loss rate, represented by the reciprocal photon lifetime $1/\tau_P$ of Eq. (3.76). Obviously, ΓG_S is larger than the net gain rate $G(n_T, N_P) = G(n_t, N_P) = 0$ where the material becomes transparent. Only above threshold the number of photons generated per time becomes larger than the number of photons annihilated. Excluding other loss mechanisms, the maximum photon lifetime τ_P is determined by the minimum mirror transmission losses for the given configuration. For $Q \neq 0$ the photon number N_P becomes already significant for $\Gamma G < \Gamma G_S = 1/\tau_P$, so that for $\mathrm{d}\,/\,\mathrm{d}t = 0$ the gain rate ΓG is always smaller than the idealized threshold gain ΓG_S as *defined* in Eq. (3.85). From Eq. (3.83) we see that

$$N_P = \frac{(Q/\tau_{\text{eff}}) n_T V}{1/\tau_P - \Gamma G} = \frac{\Gamma r_{\text{sp}}^{(\text{eM})} V}{1/\tau_P - \Gamma G} = \frac{r_{\text{sp}}^{(\text{eM})} V}{G_S - G}\,. \tag{3.86}$$

Threshold current

The threshold current density $J_S = I_S/(bL)$ for the 5-layer structure Fig. 3.11(b) becomes minimum for a certain height d of the active layer, because the field confinement factor depends on d, $\Gamma = \Gamma(d)$. If d is small, only a small portion of the field interacts with the amplifying medium and the carrier concentration must be high. If d is large, the field is well confined inside the active layer, but only the region with the maximum field strength interacts efficiently with the population-inverted semiconductor, and again the carrier concentration must be high.

With Eq. (3.85) and g from Eq. (3.79), (3.80) we calculate the threshold current by eliminating n_{TS},

$$J_S = \frac{I_S}{bL} = \frac{e n_t}{\tau_{\text{eff}}} \left[d + \frac{\Gamma \alpha_V + \alpha_R}{g_0} \frac{d}{\Gamma(d)} \right]. \tag{3.87}$$

[46] Usually, the spontaneous emission factor is very small, $Q = 10^{-5} \ldots 10^{-4}$. This is true if the wave is guided by a difference of the real part of the refractive index (index guided laser, Fig. 3.6 on Page 58 and Fig. 3.27(b) on Page 94). However, if the waveguiding is dominated by the gain mechanism itself (gain guided laser, Fig. 3.27(a) on Page 94), the Q-factor is increased to $Q_e = K_e Q$, $K_e = 10 \ldots 20$. More details on these structures are given in Sect. 3.4.5 on Page 93 (see also Sect. 3.6.2 Page 191 in reference Footnote 9 on Page 49).

From the approximation $\Gamma(d) \approx 2V^2/(1+2V^2)$ in Eq. (3.70) and $V \sim d$ we see that $\Gamma \sim d^2$ for d small ($2V^2 \ll 1$), and $\Gamma \to 1$ for d, V large. Equation (3.87) has the structure ($c_{1,2} = \text{const}_d$)

$$J_S = c_1 d + \frac{c_2}{d} \qquad \text{for} \quad d < d_c = \sqrt{10}\,\frac{\lambda}{\sqrt{2}\,\pi}\Big/\sqrt{n^2 - n_2^2} = 0.71 \times \lambda \Big/ \sqrt{n^2 - n_2^2}\,, \quad (3.88)$$

so a minimum J_S is obvious. The term $c_1 d = (e n_t/\tau_{\text{eff}})d$ is the transparency current density for disappearing losses c_2/d. For a specific parameter set of a GaAs-(Ga,Al)As laser (g_0, n_t from Eq. (3.79), $\tau_{\text{eff}} = 1\,\text{ns}$, $\alpha_V + \alpha_R = 50\,\text{cm}^{-1}$, $\lambda = 0.87\,\mu\text{m}$, $n = 3.59$, $n_2 = 3.45$, $d_c = 0.62\,\mu\text{m}$) the minimum threshold current density is $J_S = 2.88\,\text{kA}\,\text{cm}^{-2}$ for an optimum layer thickness of $d = 0.07\,\mu\text{m}$. With $L = 300\,\mu\text{m}$, $b = 5\,\mu\text{m}$ the threshold current is $I_S = 43\,\text{mA}$.

The threshold current depends on temperature. The dependence on a temperature increase ΔT is well described by the empirical function

$$I_S(\Delta T) = I_S(0)\,\mathrm{e}^{\Delta T/T_0}, \quad \frac{1}{T_0} = \frac{1}{I_S(0)}\,\frac{\mathrm{d}I_S(\Delta T)}{\mathrm{d}\Delta T}\bigg|_{\Delta T = 0} = \alpha_{IS}\,. \quad (3.89)$$

The characteristic temperature T_0 has to be found from measurements. The temperature dependence comes from the strongly temperature-dependent distribution of carriers in the CB and VB, see Fig. 3.15(b). It is seen that with increasing T the net gain $r_{\text{ind}}^{(\text{M})}/N_P$ decreases. Further, the carrier confinement by the potential walls of the heterostructure becomes worse with increasing T. Because these walls are lower for the (In,Ga)(As,P) system, and because Auger recombinations become increasingly important with rising temperature, the temperature coefficient α_{IS} for InP is larger than for GaAs (InP: $T_0 = 40 \ldots 80\,\text{K}$. GaAs: $T_0 = 120 \ldots 230\,\text{K}$). Therefore the characteristic $P_a = P_a(I)$ Eq. (3.99) shifts with the temperature-dependent threshold.

Normalized rate equations

The rate equations Eq. (3.83) can be normalized with the help of Eq. (3.85),

$$\begin{aligned} \tau_P\,\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{N_P/\tau_P}{I_S/e}\Big) &= \frac{N_P/\tau_P}{I_S/e}\Big[\frac{\Gamma G(n_T, N_P)}{1/\tau_P} - 1\Big] + Q\,\frac{n_T V}{n_{TS} V}\,, \\ \tau_{\text{eff}}(n_{TS})\,\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{n_T V}{n_{TS} V}\Big) &= \frac{I/e}{I_S/e} - \frac{n_T V}{n_{TS} V} - \frac{N_P/\tau_P}{I_S/e}\,\frac{\Gamma G(n_T, N_P)}{\Gamma G_S}\,. \end{aligned} \quad (3.90)$$

We define normalized quantities for the photon and carrier number, for the pump current and for the gain rate,

$$\begin{aligned} N_P^\times &= \frac{N_P/\tau_P}{I_S/e}\,, \qquad N_T^\times(n_T) = \frac{n_T}{n_{TS}}\,, \quad N_t^\times = \frac{n_t}{n_{TS}}\,, \quad \varepsilon_G^\times = \frac{\Gamma}{V}\,\frac{I_S}{e}\,\tau_P\,\varepsilon_G\,, \\ I^\times &= \frac{I}{I_S}\,, \qquad G^\times(N_T^\times, N_P^\times) = \frac{\Gamma G(n_T, N_P)}{1/\tau_P} = \frac{N_T^\times - N_t^\times}{1 - N_t^\times}\,\frac{1}{1 + \varepsilon_G^\times N_P^\times}\,. \end{aligned} \quad (3.91)$$

With the spontaneous emission factor Eq. (3.84), the normalized rate equations are written as

$$\begin{aligned} \tau_P\,\frac{\mathrm{d}N_P^\times}{\mathrm{d}t} &= N_P^\times(G^\times - 1) + Q N_T^\times\,, \\ \tau_{\text{eff}}\,\frac{\mathrm{d}N_T^\times}{\mathrm{d}t} &= I^\times - N_T^\times - N_P^\times G^\times. \end{aligned} \quad (3.92)$$

Characteristic curves

For $\mathrm{d}/\mathrm{d}t = 0$ the rate equations (3.92) may be easily solved if spontaneous emission and gain compression is neglected, $Q = 0$ and $\varepsilon_G^\times = 0$. Below and at threshold $I^\times \le 1$ the photon number is zero, $N_P = 0$, and the first line of Eq. (3.92) is fulfilled for any $G^\times$. The carrier number increases

with the current, $N_T^\times = I^\times$. Above threshold for $N_P^\times > 0$ the normalized gain is clamped to $G^\times = 1$, and so is the carrier number, $N_T^\times = 1$, as may be seen from Eq. (3.91). Therefore the photon number increases according to $N_P^\times = I^\times - 1$. The clamped carrier density has the consequence that any residual dependency of $\tau_{\rm sp}$, $\tau_{\rm eff}$ and Q on the carrier concentration becomes essentially unimportant. In summary we have

$$\begin{aligned} &I^\times \le 1\text{:}\quad N_T^\times = I^\times, \quad N_P^\times = 0\,, \qquad\qquad Q = 0\,, \\ &I^\times > 1\text{:}\quad N_T^\times = 1\,, \quad\; N_P^\times = I^\times - 1\,, \quad G^\times = 1\,, \end{aligned} \tag{3.93}$$

The normalized light output and the CB carrier density versus the normalized injection current, $N_P^\times = f(I^\times)$ and $N_T^\times = g(I^\times)$ are displayed in Fig. 3.21. If spontaneous emission into the lasing mode is important, $Q \neq 0$, the straight lines are "softened". This is to be seen in Fig. 3.21 where a simplified gain dependence $G^\times = N_T^\times$ according to Eq. (3.81) was assumed.

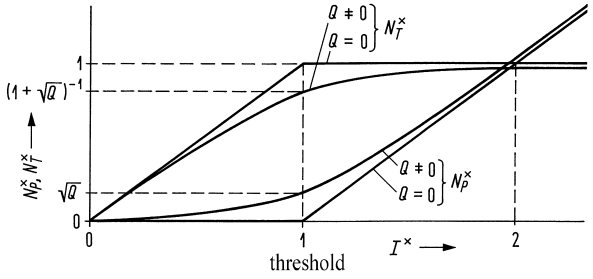


Fig. 3.21. Normalized photon number $N_P^\times$ and normalized CB carrier density $N_T^\times$ as a function of the normalized injection current $I^\times$. For $Q \neq 0$ a simplified gain dependence $G^\times = N_T^\times$ according to Eq. (3.81) is assumed.

Powers and Efficiencies

The totally generated and the total output power, respectively, are denoted as P and P_a. They are given by the total photon energy $N_P hf$ in the resonator volume per lifetime $\tau_{P,R}$ of the photons with respect to total losses (τ_P) or mirror transmittivity (τ_R),[47]

$$P = \frac{N_P hf}{\tau_P}\,, \qquad P_a = \frac{N_P hf}{\tau_R}\,. \tag{3.94}$$

Analogous to Eqs. (3.58), (3.60) one defines an internal and an external quantum efficiency for the laser diode by

$$\eta_{\rm int}^{\rm LD} = \frac{P/(hf)}{I/e}\,, \qquad \eta_{\rm ext}^{\rm LD} = \frac{P_a/(hf)}{I/e}\,, \tag{3.95}$$

From these relations and with Eq. (3.76) we write

$$\frac{\eta_{\rm ext}^{\rm LD}}{\eta_{\rm int}^{\rm LD}} = \frac{\tau_P}{\tau_R} = \frac{1}{1 - \frac{2\alpha_V L}{\ln(R_1 R_2)}}\,. \tag{3.96}$$

Through the slope of the characteristic curve $P_a = P_a(I)$ above threshold one defines a differential quantum efficiency

$$\eta_d = \mathrm{d}\left(\frac{P_a}{hf}\right) \Big/ \mathrm{d}\left(\frac{I}{e}\right)\,. \tag{3.97}$$

[47] If an electric heater consumes an energy of 2 kW h within a time of 1 h, then the power consumption of the device is obviously $P = 2\,\mathrm{kW\,h}\,/1\,\mathrm{h} = 2\,\mathrm{kW}$. The same is true for the laser resonator: We know the energy $N_P hf$ stored inside, and we know the time τ_R after which this energy has been lost through the mirrors (disregarding other loss mechnisms). Therefore, the power output from both mirrors is $P_a = N_P hf/\tau_R$, Eq. (3.94).

For an ideal laser diode with $Q = 0$, $\varepsilon_G^\times = 0$ and with the solution $N_P^\times = I^\times - 1$ Eq. (3.93) we find

$$\left.\begin{aligned} P &= \frac{hf}{e}(I - I_S), \\ P_a &= \frac{hf}{e}\frac{\tau_P}{\tau_R}(I - I_S), \end{aligned}\right\} \qquad \eta_d = \frac{\tau_P}{\tau_R} = \frac{\eta_{\text{ext}}^{\text{LD}}}{\eta_{\text{int}}^{\text{LD}}}\,. \tag{3.98}$$

For an actual laser one approximates the stimulated total power and the output power dependence, respectively, by

$$P = \eta_{\text{ind}}\frac{hf}{e}(I - I_S), \quad P_a = \eta_{\text{ind}}\frac{hf}{e}\frac{\tau_P}{\tau_R}(I - I_S). \tag{3.99}$$

The quantity η_{ind} is the efficiency for induced emission, indicating which percentage of the totally generated photons originated from stimulated emission acts. Because spontaneous emission goes into each of the resonator modes contained inside the linewidth Δf_H of the lineshape $\rho(f)$ Eq. (3.37), η_{ind} is of the order $N_P/(N_P + \varrho_{\text{tot}}(f)V\Delta f_H)$, where $\varrho_{\text{tot}}(f)V\Delta f_H$ is the number of relevant laser modes Eq. (3.7). From Eq. (3.99) the following interrelations may be derived,

$$\eta_d = \eta_{\text{ind}}\frac{\tau_P}{\tau_R} = \frac{\eta_{\text{ind}}}{1 - \frac{2\alpha_V L}{\ln(R_1 R_2)}} = \eta_{\text{ind}}\frac{\eta_{\text{ext}}^{\text{LD}}}{\eta_{\text{int}}^{\text{LD}}}\,. \tag{3.100}$$

By measuring the differential laser quantum efficiency with $R_1 = R_2 = R$ for varying resonator lengths L, the differential quantum efficiency η_d may be extrapolated for $L = 0$ and the efficiency η_{ind} Eq. (3.99) for induced emission may be determined. Typical differential efficiencies are in the range $\eta_d = 0.5 \ldots 0.8$, efficiencies for induced emission are found to be $\eta_{\text{ind}} = 0.65 \ldots 0.9$. The differential quantum efficiency is small for small reflection factors $R_{1,2}$ or for a small mirror lifetime τ_R Eq. (3.98). The more power is coupled out of the resonator, the higher the intensity modulation sensitivity $\mathrm{d}P_a/\,\mathrm{d}\,I = (hf/e)(\tau_P/\tau_R)$ will be. Typical output powers are in the range $P_a = 1 \ldots 10\,\text{mW}$ for communication lasers with current modulation, and $P_a = 100 \ldots 500\,\text{mW}$ for continuous wave applications.

The current-voltage characteristic of a laser diode is written as (saturation current I_{S0}, threshold current I_S)

$$\begin{aligned} I &= I_{S0}\Big[\exp[\beta(U - R_S I)] - 1\Big] \quad && I \le I_S, \\ U &= W_G/e + R_S I && I \ge I_S. \end{aligned} \tag{3.101}$$

The CB carrier density is assumed to be clamped to the threshold value for $I > I_S$ thereby fixing the energetic difference of the quasi Fermi levels,

$$e(U - R_S I) = W_{Fn} - W_{Fp} \approx W_G\,. \tag{3.102}$$

The series resistance is in the order $R_S = 1 \ldots 10\,\Omega$. The quantity $\beta = 1/(\kappa U_T) = 15 \ldots 30\,\text{V}^{-1}$ is connected to the thermal voltage $U_T = kT_{\text{RT}}/e$ ($U_T = 25\,\text{mV}$ at room temperature $T_{\text{RT}} = 293\,\text{K}$). Typical values for the ideality factor κ are $\kappa = 1.3 \ldots 2.7$.

Small-signal intensity modulation

To encode information into the laser beam, the optical output of the laser must be modulated. One of the unique attractions of a laser diode is the possibility of directly modulating the output light power by modulating the injection current. Because of amplitude-phase coupling, Sect. 3.4.3 on Page 91, this current modulation leads to a change in the laser mode frequency (chirping). For very high speed communications above bit rates of 10 Gbit/s the chirping of optical pulses can be avoided by employing a continuous wave laser diode and using an external modulator with a much better chirp behaviour, $|\alpha| \le 1$.

This section discusses analytical small-signal approximations of the highly nonlinear rate equations Eq. (3.83), and in the following section we demonstrate some large-signal properties by numerical solutions.

Perturbation ansatz We assume a static operation point above threshold given by the time-independent quantities N_{P0}, n_{T0}, $G_0 = G(n_{T0}, N_{P0})$, τ_P, τ_{eff}, ε_G in Eqs. (3.83), (3.93), and small time-dependent perturbations $N_{P1}(t)$, $n_{T1}(t)$, $I_1(t)$,

$$\begin{aligned} N_P(t) &= N_{P0} + N_{P1}(t), \quad G(t) = G_0 + \frac{\partial G_0}{\partial n_{T0}}\, n_{T1}(t) + \frac{\partial G_0}{\partial N_{P0}}\, N_{P1}(t), \\ n_T(t) &= n_{T0} + n_{T1}(t), \qquad G_0 = G(n_{T0}, N_{P0}), \qquad I(t) = I_0 + I_1(t). \end{aligned} \tag{3.103}$$

The differential gain rate $\partial G_0/\partial n_{T0}$ has typical values of $1.8 \times 10^{-6} \ldots 2.9 \times 10^{-6}\,\mathrm{cm^3\,s^{-1}}$. Substituting Eq. (3.103) into Eq. (3.83) and neglecting products of perturbation quantities, we solve the linearized rate equations with a Fourier ansatz $X_1(t) = X_1(\omega)\exp(\mathrm{j}\,\omega t)$, where $X_1(\omega)$ is the complex amplitude at the modulation frequency $f = \omega/(2\pi)$,

$$\begin{aligned} N_{P1}(\omega)\Big(\mathrm{j}\,\omega + \frac{1}{\tau_P} - \frac{\Gamma G_0}{1+\varepsilon_G \frac{\Gamma N_{P0}}{V}}\Big) &= \Big(\frac{Q}{\tau_{\text{eff}}} + \frac{N_{P0}}{V}\,\frac{\partial \Gamma G_0}{\partial n_{T0}}\Big) n_{T1}(\omega) V, \\ n_{T1}(\omega) V \Big(\mathrm{j}\,\omega + \frac{1}{\tau_{\text{eff}}} + \frac{N_{P0}}{V}\,\frac{\partial \Gamma G_0}{\partial n_{T0}}\Big) &= \frac{I_1(\omega)}{e} - \frac{\Gamma G_0}{1+\varepsilon_G \frac{\Gamma N_{P0}}{V}}\, N_{P1}(\omega)\,. \end{aligned} \tag{3.104}$$

Elimination of $n_{T1}(\omega)$ leads to the modulation transfer function

$$\frac{N_{P1}(\omega)}{I_1(\omega)} = \underbrace{\frac{\left(\frac{N_{P0}}{V}\,\frac{\partial \Gamma G_0}{\partial n_{T0}} + \frac{Q}{\tau_{\text{eff}}}\right)}{\omega_r^2 \tau_P}}_{\approx 1} \frac{\omega_r^2}{(\mathrm{j}\,\omega)^2 + 2\gamma_r(\mathrm{j}\,\omega) + \omega_r^2} \tag{3.105}$$

with the angular relaxation frequency ω_r and the damping constant γ_r,

$$\begin{aligned} \omega_r^2 \tau_P &= \frac{N_{P0}}{V}\,\frac{\partial \Gamma G_0}{\partial n_{T0}} + \underbrace{\overbrace{\frac{\tau_P}{\tau_{\text{eff}}}}^{\ll 1}\Big(\underbrace{\frac{1}{\tau_P} - \Gamma G_0}_{\approx 0}\, \overbrace{\frac{1-Q}{1+\varepsilon_G \frac{\Gamma N_{P0}}{V}}}^{\approx 1}\Big)}_{\approx 0} \approx \frac{N_{P0}}{V}\,\frac{\partial \Gamma G_0}{\partial n_{T0}}\,, \\ 2\gamma_r &= \frac{1}{\tau_{\text{eff}}} + \frac{N_{P0}}{V}\,\frac{\partial \Gamma G_0}{\partial n_{T0}} + \frac{1}{\tau_P}\Big(1 - \overbrace{\Gamma G_0 \tau_P}^{\approx 1}\, \overbrace{\frac{1}{1+\varepsilon_G \frac{\Gamma N_{P0}}{V}}}^{\approx 1-\varepsilon_G \frac{\Gamma N_{P0}}{V}}\Big) \\ &\approx \frac{1}{\tau_{\text{eff}}} + \underbrace{\frac{N_{P0}}{V}\,\frac{\partial \Gamma G_0}{\partial n_{T0}} + \Gamma G_0\, \varepsilon_G \frac{\Gamma N_{P0}}{V}}_{=\omega_r^2 K_r}\,. \end{aligned} \tag{3.106}$$

For large photon numbers N_{P0} the relaxation frequency and the damping constant are determined by the differential gain $\partial \Gamma G_0/\partial n_{T0}$. For low photons numbers the effective electron lifetime τ_{eff} controls the damping. Inside the laser two reservoirs exchange their energy: the electromagnetic energy stored in the resonator, and the electronic energy states. An increase in photon number reduces the carrier number, and vice versa.

This relaxation oscillation dies out with a damping constant γ_r, because photons and electrons have finite lifetimes τ_P and τ_{eff}, respectively. Such an energy exchange may be compared to the behaviour of a damped resonance circuit, where the inductor as a magnetic energy store and the capacitor storing the electric field energy are interacting. By a sudden perturbation (e. g., by a injection current step) damped oscillations $\exp(\mathrm{j}\,\omega t)$ are excited. Values of $\mathrm{j}\,\omega$ are defined by the zeros of the denominator in Eq. (3.105):

$$\mathrm{j}\,\omega = \begin{cases} -\gamma_r \pm \mathrm{j}\,\sqrt{\omega_r^2 - \gamma_r^2} & \text{damped oscillation,} \quad \gamma_r/\omega_r < 1\,, \\ -\gamma_r & \text{aperiodic limit,} \quad \gamma_r/\omega_r = 1\,, \\ -\gamma_r \pm \sqrt{\gamma_r^2 - \omega_r^2} & \text{aperiodic behaviour,} \quad \gamma_r/\omega_r > 1\,. \end{cases} \tag{3.107}$$

In the aperiodic region the slower decaying term describes the characteristic time dependence. The angular frequency of the free oscillation is $\sqrt{\omega_r^2 - \gamma_r^2}$. The aperiodic limiting case can be reached for very large photon numbers N_{P0}. For a constant current amplitude $|I_1(\omega)| = I_1 = \text{const}_\omega$ the modulation transfer function Eq. (3.105) shows a maximum at the small-signal resonance angular frequency ω_R,

$$\left| \frac{N_{P1}(\omega)/\tau_P}{I_1/e} \right| \;\rightarrow\; \max: \qquad \omega_R = \sqrt{\omega_r^2 - 2\gamma_r^2}. \tag{3.108}$$

A resonance is only possible for $\gamma_r/\omega_r < 1/\sqrt{2}$. The 3-dB bandwidth (no ripple nor resonance wanted!) is defined by

$$\left| \frac{N_{P1}(\omega_{3\,\mathrm{dB}})}{N_{P1}(0)} \right| = \frac{1}{\sqrt{2}}\,, \quad \omega_{3\,\mathrm{dB}}^2 = (\omega_r^2 - 2\gamma_r^2) + \sqrt{(\omega_r^2 - 2\gamma_r^2)^2 + \omega_r^4}\,. \tag{3.109}$$

Figure 3.22 displays the small-signal current-light transfer function as a function of the normalized

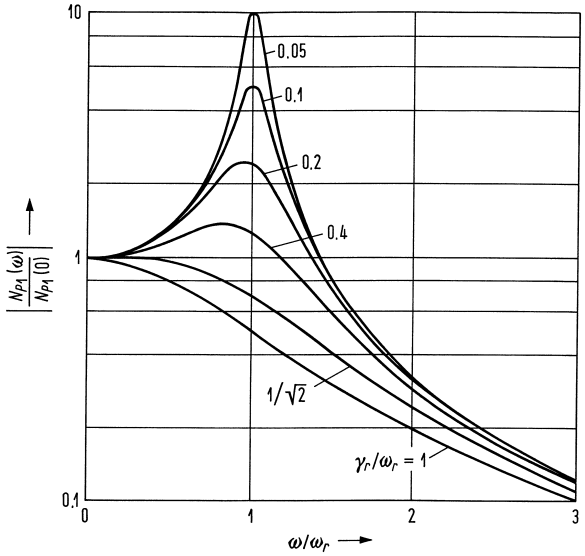


Fig. 3.22. Modulus of current-light modulation transfer function as a function of normalized current modulation frequency for various values of γ_r/ω_r

current modulation frequency. The resonance overshoot disappears for $\gamma_r/\omega_r > 1/\sqrt{2}$ (overcritical damping). Because of $\omega_r^2 \sim N_{P0}$, $\gamma_r^2 \sim N_{P0}^2$ the resonance overshoot becomes smaller with increasing operating current. The 3-dB bandwidth becomes larger because ω_r increases. However, γ_r increases faster than ω_r, and for critical damping $\gamma_r/\omega_r = 1/\sqrt{2}$ and large photon numbers the 3-dB modulation bandwidth becomes maximum,

$$\begin{aligned} &\omega_{3\,\mathrm{dB}}^{\max} = \omega_r\,, \qquad \omega_r = \gamma_r\sqrt{2} \approx \frac{\sqrt{2}}{2}\,\omega_r^2\Big(\tau_P + \frac{\varepsilon_G}{\partial G_0/\partial n_{T0}}\Big), \\ &\omega_{3\,\mathrm{dB}}^{\max} = \frac{\sqrt{2}}{K_r}\,, \qquad\qquad K_r = \tau_P + \frac{\varepsilon_G}{\partial G_0/\partial n_{T0}}\,. \end{aligned} \tag{3.110}$$

This is the intrinsic maximum modulation bandwidth. Often it is not possible to operate the laser at this operating point because of temperature limitations or because of the onset of multimode operation. In the region of overcritical damping $\gamma_r/\omega_r > 1/\sqrt{2}$ the modulation bandwidth starts to decrease, Eq. (3.109). The relaxation frequency could be enlarged by decreasing the photon

lifetime τ_P (e. g., by reducing the resonator length L), but this broadens the spectrum, so that only the differential gain $\partial G_0/\partial n_{T0}$ may be manipulated for highest $\omega_{3\,\mathrm{dB}}^{\max}$. This can be achieved by a p-doping of the active layer,[48] see also Page 65. The differential gain $\partial G_0/\partial n_{T0}$ becomes larger by a factor 5 if n_A is increased from $5 \times 10^{16}\,\mathrm{cm}^{-3}$ to $5 \times 10^{18}\,\mathrm{cm}^{-3}$. By using quantum film or quantum wire lasers the differential gain becomes even larger. Relaxation frequencies of 38 GHz could be achieved.

Figure 3.23 shows the electric equivalent circuit if the laser diode. The "internal laser diode" may be regarded as a short circuit because of the voltage clamping, Eq. (3.102). R_S is the series resistance, C_P a parasitic parallel capacitance (in Fig. 3.27(b) parasitic pn-junctions in parallel to the active layer), L_S is the bond wire inductance, typically 1 nH / mm, R_G the generator impedance. The cutoff frequencies defined by $\omega R_S C_P = 1$ and $\omega L_S = R_G$ should not impair the intrinsic modulation capability of the device. For a cutoff frequency of 20 GHz at $R_S = 10\,\Omega$, $R_G = 50\,\Omega$ the values $C_P < 0.8\,\mathrm{pF}$, $L_S < 0.4\,\mathrm{nH}$ should be chosen.

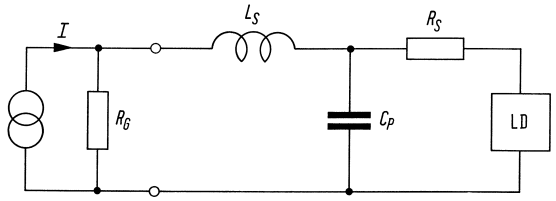


Fig. 3.23. Electric equivalent circuit of the laser diode. The internal laser diode may be regarded as a short circuit.

Large-signal intensity modulation

A general analytic solution of the nonlinear rate equations (3.83) is not known, so a specific numerical solution of the simplified normalized rate equations will be discussed, Fig. 3.24. During

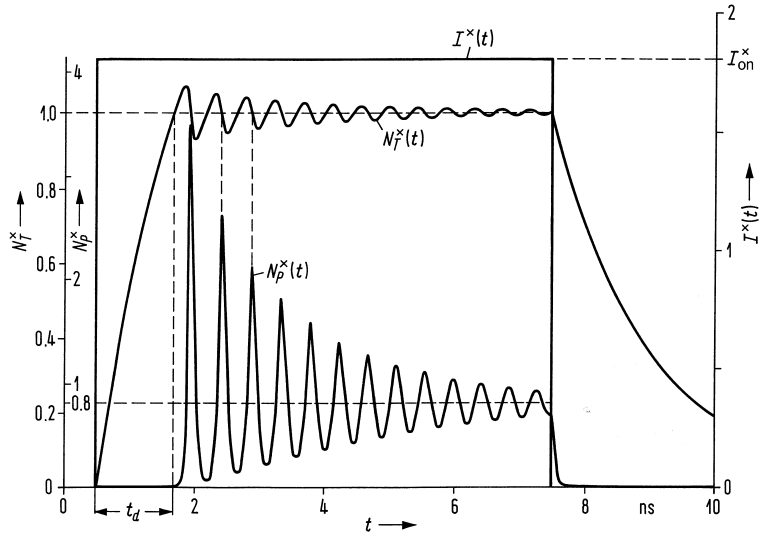


Fig. 3.24. Relaxation oscillation for a current step $I^\times = 1.8$. Parameters are $\tau_P = 2.5\,\mathrm{ps}$, $\tau_{\mathrm{eff}} = 1.5\,\mathrm{ns}$, $Q = 5 \times 10^{-4}$

the delay time t_d the normalized carrier density rises to $N_T^\times = 1$. When the threshold is reached the photon number $N_P^\times$ starts rising. At first, the carrier density increases further, but when

[48] See Footnote 28 on Page 62, Footnote 21 on Page 60

the photon number has become large enough and the stimulated emissions are numerous, the carrier density decreases. As long as $N_T^\times > 1$ holds, the photon number increases further. When the condition $N_T^\times < 1$ is met, the gain rate becomes negative and the photon number decreases rapidly. Therefore the number of stimulated emissions decreases and the carrier number can be re-established by the injection current.. When the threshold $N_T^\times = 1$ is exceeded, the photon number increases again. A weakly damped relaxation oscillation results for the parameters chosen in Fig. 3.24, which has not yet died out when the current impulse is switched off. A small variation of the carrier density $N_T^\times$ causes very large changes in the photon number $N_P^\times$. When the current is switched on the carrier density starts with a finite slope while the photon number has with a zero slope.

To calculate the delay time t_d when switching from $I_{\text{off}} < I_S$ to $I_{\text{on}} > I_S$ the induced emission term in the second rate equation (3.83) may be neglected. The solution of the resulting equation

$$\frac{\mathrm{d}n_T}{\mathrm{d}t} = \frac{I}{eV} - \frac{n_T}{\tau_{\text{eff}}} \tag{3.111}$$

for the initial condition $I_{\text{off}} < I_S$ for $t < 0$, $I_{\text{on}} > I_S$ for $t > 0$ is

$$n_T(t) = \frac{\tau_{\text{eff}}}{eV}\Big[I_{\text{off}} + (I_{\text{on}} - I_{\text{off}})\left(1 - \mathrm{e}^{-t/\tau_{\text{eff}}}\right)\Big]. \tag{3.112}$$

After the delay time t_d the threshold carrier density is reached,

$$n_T(t_d) = n_{TS} = \frac{\tau_{\text{eff}} I_S}{eV}\,, \tag{3.113}$$

from which the delay time t_d results,

$$t_d = \tau_{\text{eff}} \ln \frac{I_{\text{on}} - I_{\text{off}}}{I_{\text{on}} - I_S}\,, \qquad I_S \sim \frac{1}{\tau_{\text{eff}}}\,, \quad \text{see Eq. (3.87).} \tag{3.114}$$

Thus, the effective carrier lifetime may be measured from the switch-on delay time t_d.

For actually modulating the laser, a bias current I_{off} at or slightly above threshold I_S is used. If $I_{\text{off}} = I_S$, $I_{\text{on}} > I_S$ is chosen the solution Eq. (3.112) is still approximately valid. For small times and for calculating the steep slope of the first photon impulse in Fig. 3.24 we apply the approximation

$$n_T(t) - n_{TS} = \frac{1}{eV}\,(I_{\text{on}} - I_S)\,t\,. \tag{3.115}$$

In the first of the rate equations (3.83) the spontaneous emission term is neglected. The first term is approximated near threshold by $\Gamma G - 1/\tau_P = \dfrac{\partial \Gamma G_S}{\partial n_{TS}}(n_T - n_{TS})$,

$$\frac{\mathrm{d}N_P}{\mathrm{d}t} = N_P \frac{\partial \Gamma G}{\partial n_T}(n_T - n_{TS}). \tag{3.116}$$

Substituting $n_T - n_{TS}$ from Eq. (3.115) we integrate and have

$$\frac{N_P(t)}{N_P(0)} = \exp\Big(\frac{I_{\text{on}} - I_S}{2eV}\,\frac{\partial \Gamma G}{\partial n_T}\,t^2\Big). \tag{3.117}$$

This is an impulse starting with a slope of zero. From Eqs. (3.94), (3.98) we write

$$\frac{I_{\text{on}} - I_S}{e} = \frac{N_{P\text{on}}}{\tau_P}\,. \tag{3.118}$$

With the help of Eq. (3.106) we find the final result

$$\frac{N_P(t)}{N_P(0)} = \exp\Big(\frac{\Gamma N_{P\text{on}}}{2\tau_P V}\,\frac{\partial G}{\partial n_T}\,t^2\Big) = \exp\Big(\frac{1}{2}\omega_r^2 t^2\Big). \tag{3.119}$$

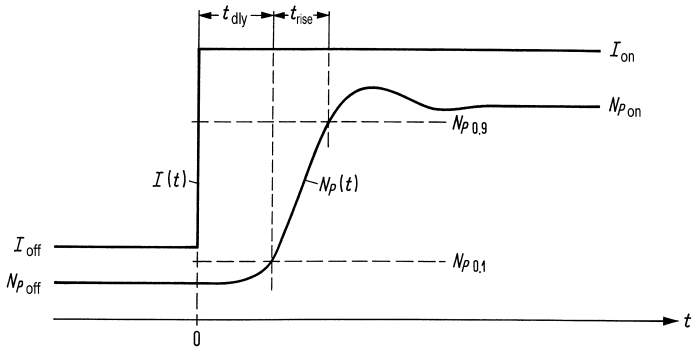


Fig. 3.25. Current step for a laser with a bias above threshold ($I_{\rm off}, I_{\rm on} > I_S$). Impulse delay time $t_{\rm dly}$, risetime $t_{\rm rise}$. $N_{P\,0.1}$ and $N_{P\,0.9}$ are the photon numbers at the 10-% and 90-% points of the impulse maximum

The relaxation frequency ω_r is attributed to a bias current $I_{\rm on}$. — An analogous procedure leads to the current-off response. Thus, the large-signal property is also determined by the relaxation frequency. In Fig. 3.25, a schematic of the rising portion of the impulse is displayed. At $t = 0$ the current is abruptly switched on. The impulse delay time $t_{\rm dly}$ and the risetime $t_{\rm rise}$ are defined at the 10-% and 90-% points of the photon number,

$$\begin{aligned} N_{P\,0.1} &= N_{P\rm off} + 0.1\,(N_{P\rm on} - N_{P\rm off})\,, \\ N_{P\,0.9} &= N_{P\rm off} + 0.9\,(N_{P\rm on} - N_{P\rm off})\,. \end{aligned} \tag{3.120}$$

If we apply Eq. (3.119) for calculating $t_{\rm dly}$, $t_{\rm rise}$ we find the result

$$\begin{aligned} t_{\rm dly} &= \frac{1}{f_r}\frac{1}{\pi\sqrt{2}}\sqrt{\ln\Big(0.9 + \frac{0.1\,N_{P\rm on}}{N_{P\rm off}}\Big)}\,, \\ t_{\rm rise} &= \frac{1}{f_r}\frac{1}{\pi\sqrt{2}}\sqrt{\ln\Big(0.1 + \frac{0.9\,N_{P\rm on}}{N_{P\rm off}}\Big)} - t_{\rm dly}. \end{aligned} \tag{3.121}$$

For an on-off ratio $N_{P\rm on}/N_{P\rm off} = 10$ we have the delay $t_{\rm dly} = 0.18/f_r$ and the risetime $t_{\rm rise} = 0.15/f_r$. Falltimes are defined in an analogous way. Obviously, the frequency limit of large-signal intensity modulation is also determined by the relaxation frequency f_r as it is the case for small-signal modulation. The clock period of a binary modulation sequence must equal at least the sum of rise and falltimes of an impulse. If this sum is approximated by $0.8/f_r$, the maximum bit rate is $f_r/0.8 = 1.25\,f_r$.

3.4.3 Amplitude-phase coupling

Because of Eq. (3.44) the gain rate G is a function of the frequency f and (via the quasi Fermi levels) a function of the carrier concentration n_T, Eq. (3.24). This is also true for the modal power gain g and consequently for the imaginary part of the refractive index $-n_i$, Eq. (3.73). The real part n of the complex refractive index $\bar{n}$ depends on f, n_T because of three reasons (see Eq. (2.7) on Page 8):

Band filling With the carrier injection the band-band absorption decreases because of the filling of lower CB states, and the absorption energy $hf_1 = W_G + \Delta W_1$ increases (ΔW_1 increasing with n_T). Therefore, $\Delta n < 0$ for $f < f_1$ and $f > f_1$ according to the Kramers-Kronig relations depicted in Fig. 2.1(b), where f is *outside* the region of anomalous dispersion $\mathrm{d}n/\,\mathrm{d}f < 0$. For InP typical data are $\Delta n = -7.7 \times 10^{-21} n_T/\,\mathrm{cm}^{-3}$ at $\lambda = 1.24\,\mu$m, $\Delta n = -5.6 \times 10^{-21} n_T/\,\mathrm{cm}^{-3}$ at $\lambda = 1.55\,\mu$m.

Coulomb interaction The interaction of carriers by coulomb forces reduces the bandgap, and the absorption increases, especially in the vicinity of the bandgap energy $hf_2 = W_G$. Therefore, $\Delta n > 0$ for $f < f_2$ and for $f > f_2$ result.

Free carrier The free-carrier absorption causes always $\Delta n < 0$ at optical frequencies.

$$\Delta n = -\frac{e^2\mu_0\lambda^2}{8\pi^2 n}\Big(\frac{n_T}{m_n} + \frac{p}{m_p}\Big) = -\frac{4.485\times10^{-22}}{n}\Big(\frac{\lambda}{\mu\mathrm{m}}\Big)^2\Big[\frac{n_T/\,\mathrm{cm}^{-3}}{m_n/m_0} + \frac{p/\,\mathrm{cm}^{-3}}{m_p/m_0}\Big]. \quad (3.122)$$

The ratios $m_n/m_0, m_p/m_0$ are given in Table 3.3 on Page 60.

At the oscillation frequency of a laser diode the combination of these effects leads to a reduced refractive index n ($\Delta n < 0$) for increasing CB carrier density n_T, and (via the Kramers-Kronig relations Eq. (2.7)) to an increased gain constant g. With the help of Eq. (3.73) one defines a quantity α (α-factor, line broadening factor, Henry factor)

$$\alpha = \frac{\partial n/\partial n_T}{\partial n_i/\partial n_T} = -2k_0\frac{\partial n/\partial n_T}{\partial(\Gamma g - \alpha_V)/\partial n_T} = -2k_0\frac{\partial n/\partial n_T}{\partial(\Gamma g)/\partial n_T} > 0. \quad (3.123)$$

The last form of Eq. (3.123) assumes the loss constant α_V to be independent of the carrier density n_T. Typical values for laser diodes are in the range $\alpha = 2\ldots8$. Therefore a correlation exists between amplitude and phase of the laser diode oscillator. Spontaneous emissions cause amplitude and phase changes. Because of Eq. (3.123) such an amplitude change gives rise to a secondary phase change, so that a broadening of the emission line is to be expected.

For a stationary laser oscillation the operating point (subscript 0) is given by $G(n_{T0}) = 1/\tau_P$. This is to be seen from Eq. (3.93) and (3.91) where $G^\times = \Gamma G(n_{T0}, N_{P0})\,\tau_P = 1$ may be deduced. The angular optical frequency is ω_0, Eq. (3.67). When changing the carrier density differentially, the gain rate G varies, and the "instantaneous" (on the scale of an optical period $1/f_0$ slowly varying) optical frequency ω deviates from its unperturbed value by a small amount $\mathrm{d}\omega$. This frequency difference $\Delta\omega$ defines the time derivative of the optical phase, $\mathrm{d}\varphi/\,\mathrm{d}t = \Delta\omega$. From Eq. (3.67) we find

$$\begin{aligned}
\mathrm{d}(\omega n) &= \frac{\partial(\omega n)}{\partial\omega}\,\mathrm{d}\omega + \frac{\partial(\omega n)}{\partial n}\,\mathrm{d}n = \Big(\overbrace{n + \omega\frac{\partial n}{\partial\omega}}^{n_g}\Big)\,\mathrm{d}\omega + \omega\,\mathrm{d}n \stackrel{!}{=} 0\,,\\
\mathrm{d}\omega &= -\frac{\omega}{n_g}\,\mathrm{d}n = -\frac{\omega}{n_g}\frac{\partial n}{\partial n_T}\,\mathrm{d}n_T = -\frac{\alpha\omega}{2k_0 n_g}\frac{\partial(\Gamma g)}{\partial n_T}\,\mathrm{d}n_T \approx \Delta\omega = \omega - \omega_0 = \frac{\mathrm{d}\varphi}{\mathrm{d}t}\,, \qquad (3.124)\\
\frac{\mathrm{d}\varphi}{\mathrm{d}t} &= \frac{\alpha}{2}v_g\frac{\partial(\Gamma g)}{\partial n_T}\Delta n_T \approx \frac{\alpha}{2}\frac{\partial(\Gamma G)}{\partial n_T}\Delta n_T\,.
\end{aligned}$$

However, with $\mathrm{d}v_g/\,\mathrm{d}n_T, \mathrm{d}n_g/\,\mathrm{d}n_T \neq 0$ the last form of Eq. (3.124) is only approximately valid,

$$\frac{\partial(\Gamma G)}{\partial n_T} = \frac{\partial(\Gamma v_g g)}{\partial n_T} = v_g\frac{\partial(\Gamma g)}{\partial n_T}\left(1 - \frac{\frac{1}{n_g}\frac{\partial n_g}{\partial n_T}}{\frac{1}{\Gamma g}\frac{\partial(\Gamma g)}{\partial n_T}}\right) \approx v_g\frac{\partial(\Gamma g)}{\partial n_T}\,. \quad (3.125)$$

Because of Eqs. (3.79), (3.123) ($\alpha = 2\ldots8$) the ratios of the relative change of n_g, g with n_T are smaller than 10^{-2}. Therefore, Eq. (3.124) may be written with reference to Eqs. (3.74), (3.77) and neglecting spontaneous emission as

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{\alpha}{2}\Big(\underbrace{\frac{\partial(\Gamma G)}{\partial n_T}\Delta n_T + \overbrace{\Gamma G(n_{T0}) - \frac{1}{\tau_P}}^{=0}}_{\Gamma G(n_T)}\Big) = \frac{\alpha}{2}\Big(\Gamma G - \frac{1}{\tau_P}\Big) = \frac{\alpha}{2}\frac{1}{N_P}\frac{\mathrm{d}N_P}{\mathrm{d}t}\,. \quad (3.126)$$

For the unperturbed stationary oscillation we have $\omega = \omega_0$, $\mathrm{d}\omega = \mathrm{d}\varphi/\,\mathrm{d}t = 0$. A perturbation in the carrier density $\mathrm{d}n_T \neq 0$ leads to a change in optical angular frequency, Eq. (3.124). The phase change by amplitude-phase coupling may be incorporated into the basic equations (3.83). The relation for the photon number is supplemented by an equation for the phase. Neglecting again spontaneously emitted photons we find for the analytic electric field $\underline{E}(t)$ at the laser mirrors

$$\frac{\mathrm{d}N_P}{\mathrm{d}t} = N_P\Big(\Gamma G - \frac{1}{\tau_P}\Big), \quad \frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{\alpha}{2}\Big(\Gamma G - \frac{1}{\tau_P}\Big), \quad \underline{E}(t) \sim \sqrt{N_P(t)}\;\mathrm{e}^{\mathrm{j}[\omega_0 t + \varphi(t)]}\,. \quad (3.127)$$

For a measurement of α we derive the oscillation mode frequency dependence from Eq. (3.124),

$$\frac{\mathrm{d}\omega_0}{\mathrm{d}n_T} = -\frac{\omega_0}{n_g}\frac{\partial n}{\partial n_T} \approx \frac{\alpha}{2}\frac{\partial(\Gamma G)}{\partial n_T}\,, \qquad \alpha = 2\,\frac{\mathrm{d}\omega_0}{\mathrm{d}n_T}\Big/\frac{\partial(\Gamma G)}{\partial n_T}\,. \tag{3.128}$$

By a measuring of the (effective) gain rate change and of the shift in angular resonance frequency ω_0 with the carrier density n_T the a-factor may be calculated.

3.4.4 LD spectrum

When the injection current is below threshold, the laser diode behaves like an LED and the output is mainly due to spontaneous emission and, hence, the spectrum is broad. As the current increases beyond threshold, the longitudinal modes having a larger gain and smaller resonator loss

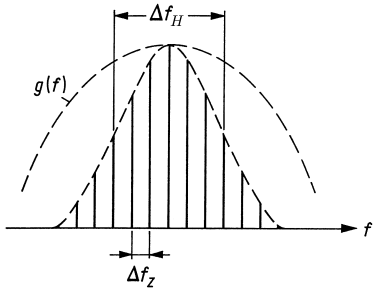


Fig. 3.26. Longitudinal mode spectrum of a gain-guided Fabry-Perot laser diode (see Page 94). Frequency-dependent net gain constant $g(f) \sim r_{\mathrm{ind}}^{(\mathrm{M})}(f)/N_P$ Fig. 3.15(b), longitudinal mode distance Δf_z Eq. (3.68), half-power bandwidth of spectral envelope Δf_H

begin to oscillate, and the spectrum changes significantly. As the current is further increased, the output spectrum becomes spectrally more concentrated, Fig. 3.26 for a gain-guided Fabry-Perot laser diode (see Footnote 46 on Page 83 and Sect. 3.4.5 on Page 94). The half-power width of the spectral envelope Δf_H is then given by (α_V is the modal or effective loss)

$$\Delta f_H\, P_a = \mathrm{const} \cdot n_{\mathrm{sp}}(1+\alpha^2) h f v_g^2 (\alpha_V + \alpha_R)\alpha_R\,. \tag{3.129}$$

The quantity Δf_H of the laser oscillator cannot be compared to the spontaneous linewidths Eq. (3.37), (3.66) on Pages 68, 78. It is proportional to the reciprocal of the total output power P_a through both mirrors. Typical values for GaAs gain-guided lasers are $\Delta f_H\, P_a = 3\,000\,\mathrm{GHz\,mW}$, i. e., with $P_a = 1\,\mathrm{mW}$ and $\Delta f_z = 100\,\mathrm{GHz}$ about 30 modes oscillate simultaneously, while at $P_a = 10\,\mathrm{mW}$ there remain only 3 modes. For index-guided lasers even single-mode oscillation may be achieved.

3.4.5 Devices

As illustrated in Figs. 3.1, 3.6, a simple laser resonator (Fabry-Perot resonator) is a rectangular cavity of six walls, all of which should provide good photon and carrier confinement to reduce the cavity loss. Among the six walls, two are at the longitudinal ends of the cavity ($z = 0, L$) which need to couple light out, and two are the heterojunctions of a 3 or 5-layer heterostructure ($x = \pm d/2$) which achieve both carrier and photon confinement from the energy bandgap and refractive index differences, respectively, Fig. 3.11. To provide the confinement at the two transverse sides ($y = \pm b/2$), two basically different structures have been used, namely *gain-guided* and *index-guided* lasers, Fig. 3.27.

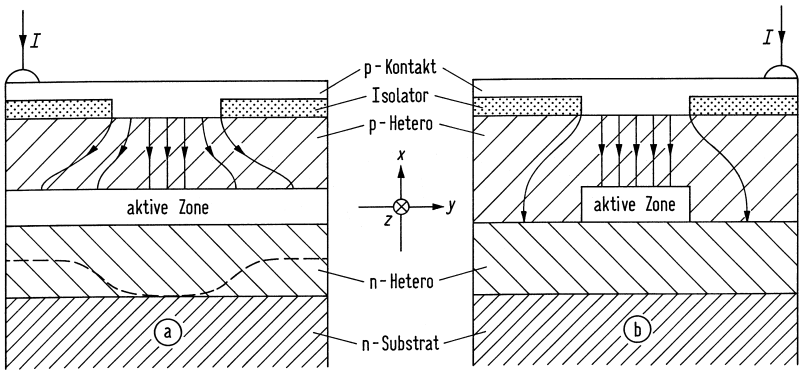


Fig. 3.27. Basic laser diode structures. (a) Gain-guided laser (b) Index-guided laser. The origin of the coordinate system is located in the centre of the active zones (p-Kontakt = p-contact, Isolator = insulator, aktive Zone = active zone).

Gain-guided lasers

A gain-guided laser, Fig. 3.27(a), has a structure that confines the transverse current flow. There is no physical confinement for photons on the two sides, but the field is concentrated near the z-axis, because $g - \alpha_V = -2k_0 n_i$ (see Eq. (3.73)) has an on-axis maximum and decreases with increasing $|y|$. This profile is due to a lateral decrease of the current density as in Fig. 3.27(a) resulting in a corresponding reduction of g, but it is also possible to increase the lateral loss α_V by moving absorbing regions nearer to the active zone, dashed line. Naturally, the (effective) real part n of the complex refractive index depends on y, too. The high-current region has a lower refractive index causing even an antiguiding effect. For a gain-guided laser the waveguiding by the lateral decrease of n_i dominates. Because gain-guided lasers have no strong transverse photon confinement, they have a relatively large threshold current in the order of $I_S = 100\,\mathrm{mA}$.

As mentioned[49] in Sect. 3.4.2, a spontaneous-emission correction factor $K_e = 10 \ldots 20$ has to be introduced for gain-guided lasers. This comes from the non-orthogonality of the gain-guided modes.

Index-guided laser

To provide a better photon confinement, index difference must be introduced on the two transverse sides $y = \pm b/2$. Laser diodes that add the index difference are called index-guided lasers, Fig. 3.27(b). Even without a current the strip waveguide cavity is defined, buried into the material with larger bandgap, i. e., lower refractive index. Such a buried-heterostructure laser introduces another two heterojunctions on the lateral sides (total four heterojunctions) to provide both carrier and photon confinement. Because of the excellent confinement, the threshold current can be as low as $I_S = 10\,\mathrm{mA}$.

Lasers with distributed feedback

Fabry-Perot lasers tend to oscillate multimoded because all resonator modes have approximately the same resonance linewidth or photon lifetime τ_P Eq. (3.76). Therefore, modes are discriminated mainly by the spectral dependence of the gain, see Fig. 3.26 on Page 93. For a better spectral control τ_P could be made strongly dispersive by choosing frequency-dependent reflection factors R_1, R_2 for each mode.

Figure 3.28 depicts a waveguide having a propagation constant $\beta = k_0 n_e$. Because of the cladding corrugation the effective refractive index n_e becomes weakly modulated along the z-axis. A test wave with power amplitude s_{2a} starting at $z = 0$ experiences reflections with a phase jump

[49] See Footnote 46 on Page 83

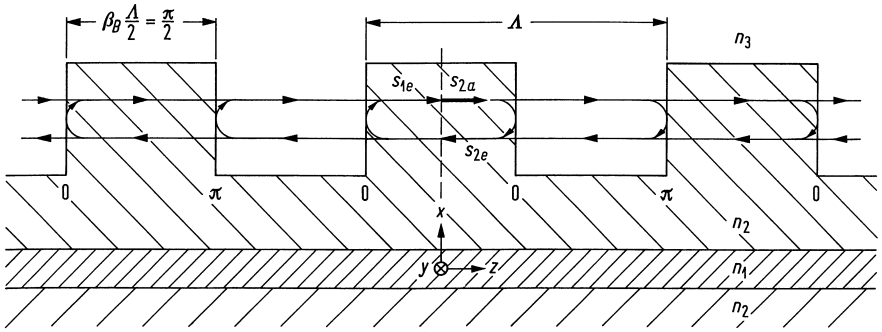


Fig. 3.28. Distributed feedback waveguide with cladding corrugation, period Λ, and refractive indices $n_1 > n_2 > n_3$. Power amplitude s_{2a} of a test wave, and total power amplitudes s_{2e}, s_{1e} of waves returning from $z > 0$ and $z < 0$ to $z = 0$

0 if n_e becomes locally smaller ("open circuit)" or with a phase π if n_e increases ("short circuit"). For the transmitted wave the phase does not change. All partial waves returning from $z > 0$ to the plane $z = 0$ are in phase, and the magnitude of the amplitude reflection coefficient $r(\omega) = s_{2e}/s_{2a}$ becomes maximum if the Bragg[50] condition $\beta = \beta_B$ is fulfilled,

$$\beta_B \cdot \Lambda = p\pi, \quad p = 1, 2, \ldots \qquad \left(\beta_B = \frac{2\pi f_B}{c} n_e\right). \tag{3.130}$$

Gratings with $p = 1, 2, \ldots$ are denoted as gratings of 1st, 2nd, ... order, and f_B is the Bragg frequency. The strength of the refractive index modulation is given by the coupling factor κ (usually $\kappa < 160\,\mathrm{cm}^{-1}$). For the fundamental spatial frequency of the rectangular grating depicted in Fig. 3.28 the refractive index variation may be written as

$$n_{e\,z}(z) = n_e + n'_e \cos(2\beta_B z), \qquad \kappa = \frac{\pi n'_e}{\lambda_B}, \quad \lambda_B = \frac{2\pi n_e}{\beta_B}. \tag{3.131}$$

For a grating with length L the maximum power reflection factor R and the frequencies $f_{R\nu}$ around f_B where $R(f) = 0$ are[51]

$$\begin{aligned} R &= \tanh^2(\kappa L) && \text{at } f = f_B, \\ f_{R\nu} - f_B &= \pm\frac{c}{2\pi n_e}\sqrt{\kappa^2 + (\nu\pi/L)^2} && \text{zeros of } R \text{ numbered by } \nu = 1, 2, 3, \ldots \end{aligned} \tag{3.132}$$

For $R > 0.32$ ($R \approx 0.32$ for the cleaved facets of a Fabry-Perot laser with $n_e = 3.6$) we require $\kappa L > 0.64$. This corresponds to $\kappa = 32\,\mathrm{cm}^{-1}$ at $L = 200\,\mu$m or $\kappa = 64\,\mathrm{cm}^{-1}$ at $L = 100\,\mu$m, and it leads to spectral widths of 213 GHz or 425 GHz between the first reflection zeros f_{R1}.

Distributed Bragg reflector laser If a grating is used as a mirror instead of a cleaved endface the device is dubbed *d*istributed *B*ragg *r*eflector laser (DBR laser), Fig. 3.29(a). If the reflection grating remains unpumped its loss has to be taken into account. The phase dependency of the complex amplitude reflection coefficient $r(\omega) = |r(\omega)| \exp[\mathrm{j}\,\varphi(\omega)]$ in the vicinity of its maximum $|r(\omega)|$ is described by an equivalent length L_{eff} (penetration depth) that is smaller than the physical length L,

$$\varphi(\omega) = k_0 n_e \cdot 2L_{\mathrm{eff}}, \qquad k_0 = \omega/c. \tag{3.133}$$

[50] Bragg, Sir (since 1920) William Henry, British physicist, ⋆ Westward (near Wigton, County Cumbria) 2.7.1862, † London 12.3.1942. Partially in cooperation with his son Sir (since 1941) William Lawrence Bragg (⋆ 1890, † 1971) he developed the turning-crystal method, and thereby founded the technique of X-ray structure analysis and X-ray spectroscopy. They shared the Nobel prize in physics in 1915.

[51] Ebeling, K. J.: Integrierte Optoelektronik, 2. Ed. Berlin: Springer-Verlag 1992. In German. English translation also available. Sect. 5.1.3 Eq. (5.14) Page 113

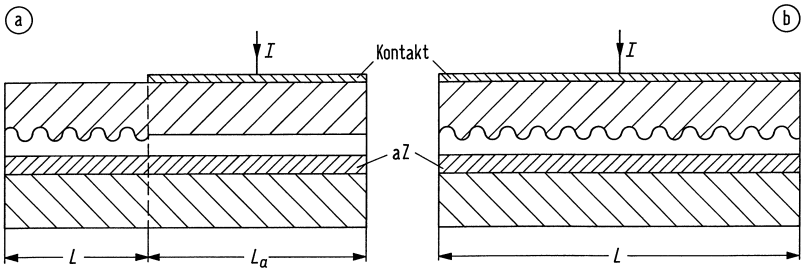


Fig. 3.29. Schematic (aZ: active zone) of a (a) DBR laser. Length of active zone L_a, length of DBR section L (b) DFB laser. Length of DFB region L

For a Fabry-Perot laser where the wave propagates with propagation constant β the resonance condition is described equivalently to Eq. (3.67) on Page 79,

$$\begin{aligned} \beta \cdot 2L = k_0 n_e \cdot 2L = 2\pi q, \qquad & q = \pm 1, \pm 2, \ldots \\ \Delta f_q = \frac{c}{2n_{eg}L} = \frac{v_g}{2L} = \frac{1}{\tau_U}, \qquad & \left(n_{eg} = n_e + \omega \frac{\partial n_e}{\partial \omega};\ v_g = \frac{c}{n_{eg}} \right), \end{aligned} \tag{3.134}$$

The effective refractive index in the active zone of length L_a is n_{ae}. For a DBR laser the resonance condition reads

$$2k_0(n_e L_{\text{eff}} + n_{ae} L_a) = 2\pi q, \qquad q = \pm 1, \pm 2, \ldots \tag{3.135}$$

A change in gain influences n_{ae} only and does not affect the resonance frequency because n_e remains constant. The DBR laser therefore has a smaller shift in emission frequency than the Fabry-Perot laser with the same length L_a of the active zone, and therefore the Henry factor (see definition Eq. (3.128)) becomes smaller.

Distributed feedback laser A grating distributed over all the active region leads to the concept of a *d*istributed *f*eed*b*ack laser (DFB laser), Fig. 3.29(b). From Fig. 3.28 it may be concluded that this laser does not oscillate at the Bragg frequency f_B: Partial waves excited by the test wave s_{2a} and contributing to the returning waves s_{1e} left of $z = 0$ are in antiphase to s_{2a} at $f = f_B$ thereby preventing an oscillation at f_B. A stopband $|f - f_B|$ results in which wave propagation becomes impossible,[52] see Eq. (3.132) on Page 95 for $\kappa L \gg \nu\pi$ (strong coupling),

$$|f - f_B| \le \frac{c\kappa}{2\pi n_e}, \quad |\beta - \beta_B| \le \kappa, \qquad \beta = \frac{2\pi f}{c} n_e . \tag{3.136}$$

In the limit of large gain $g_e - \alpha_{Ve} \gg 2\kappa$, see Eq. (3.73) on Page 80, the resonances are symmetrically arranged with respect to f_B. The longitudinal mode spacing is approximately the same as for a Fabry-Perot laser, Eq. (3.134) or Eq. (3.68) on Page 79,

$$f_q = f_B + \left(\frac{1}{2} + q \right) \frac{c}{2n_e L}, \quad q = 0, \pm 1, \pm 2, \ldots \tag{3.137}$$

A numerical evaluation of the dispersion relation shows that the resonances are slightly farther apart from f_B than suggested by Eq. (3.137); this "repulsion" of the resonances increases with κ. For $\kappa L > \pi$ the stopband becomes larger than the approximate spacing Eq. (3.137) of modes adjacent to f_B. The resonances are displayed in Fig. 3.30 where the height of the lines represents

[52] This is a very general property of periodic structures: Photonic (electromagnetic) waves in a photonic "crystal" like the one schematically depicted in Fig. 3.29(b) experience stop bands and transmission bands much the same as electron waves in a semiconductor crystal where the concept of forbidden bands, valence and conduction bands is well known.

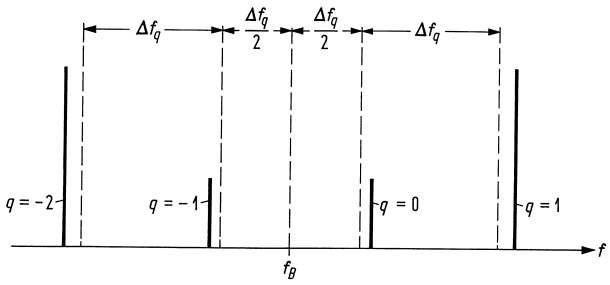


Fig. 3.30. Resonance frequencies of a DFB laser. The height of the lines represents the gain that is required for oscillation

the gain required for oscillation. For a large gain $g_e - \alpha_{Ve} \gg 2\kappa$ we may approximate (see Eq. (3.73) for the definitions of g_e, α_{Ve}, β)

$$4(\beta - \beta_B)^2 = \kappa^2 \exp[(g_e - \alpha_{Ve})L] - (g_e - \alpha_{Ve})^2. \tag{3.138}$$

For the limiting case of small gain $g_e - \alpha_{Ve} \ll 2\kappa$ the first two resonances are just at the stopband edges,

$$\beta = \beta_B \pm \kappa, \qquad g_e - \alpha_{Ve} = \frac{2\pi^2}{\kappa^2 L^3}\,. \tag{3.139}$$

At threshold the gain compensates losses, $g = \alpha_V + \alpha_R$ from Eq. (3.75), (3.85). From Eq. (3.138) or (3.139) we may define an equivalent mirror attenuation constant by $\alpha_R = g_e - \alpha_{Ve}$ that may be inferred from Eq. (3.139) if $g_e - \alpha_{Ve} \ll 2\kappa$. Measured DFB laser minimum linewidths are 900 kHz at 4 mW output power. Typical linewidths specified for off-the-shelf DFB communication lasers are about 40 MHz.

Real lasers show structural asymmetries so that only one of the resonances $q = 0$ or $q = -1$ starts oscillating. When modulating the laser, however, mode hopping can occur, and the output power changes from one to the other mode which leads to frequency jumps in the order of 100 GHz. In view of a WDM grid spacing of 100 GHz this would be a catastrophic failure. To increase the dynamical stability the asymmetry can be increased by an anti-reflection coating at only one end facet.

Phase-shifted grating Phase-shifted DFB lasers have two identical gratings on both sides of $z = 0$ that are shifted apart by L_1 equal to a quarter of the effective wavelength λ_e. This may be viewed as introducing a defect (an "impurity") into the otherwise periodic structure, leading to an additional phase shift of $\lambda_e/4 \mathrel{\widehat{=}} \pi/2$. According to the explanations near Eq. (3.136), an oscillation at the Bragg frequency $f = f_B$ of a strictly periodic grating is impossible because s_{1e} and s_{2a} are in anti-phase, Fig. 3.28. An additional shift of the round-trip phase by π would move the resonance frequency to $f = f_B$ thereby preventing the oscillation of adjacent modes because of their larger threshold gain. Equation (3.137) must be modified. The region L_1 has an effective propagation constant $\beta_1 > \beta$ with an associated phase shift $\Delta\varphi$, and the resonance condition becomes

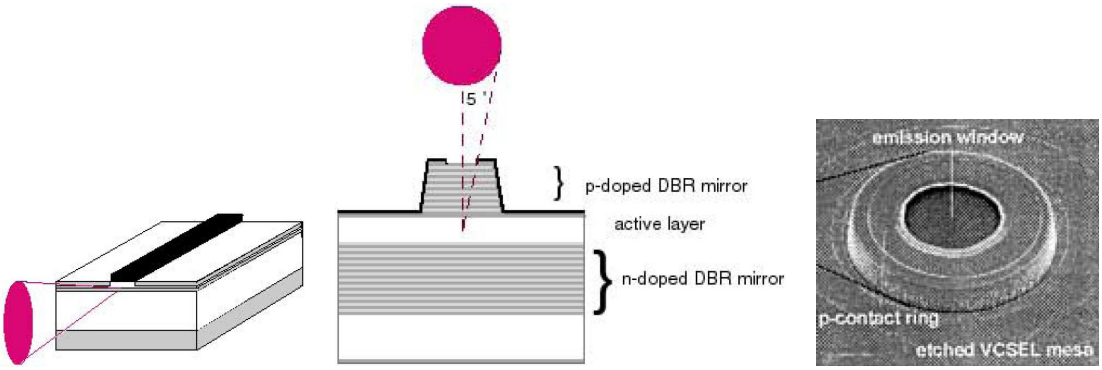
$$(\beta - \beta_B) \cdot 2L + \Delta\varphi = \left(q + \frac{1}{2}\right) \cdot 2\pi\,, \qquad \Delta\varphi = (\beta_1 - \beta) \cdot 2L_1\,. \tag{3.140}$$

If L_1 is chosen such that $\Delta\varphi = \pi$, the resonances are

$$(\beta - \beta_B) \cdot 2L = 2\pi q, \qquad q = 0, \pm 1, \pm 2, \ldots \tag{3.141}$$

The resonance with the lowest threshold gain is found for $q = 0$ at the Bragg frequency where $\beta = \beta_B$.

Line broadening factor For a DFB laser without $\lambda_e/4$ phase adjustment the Henry factor α has to be modified. In Eq. (3.129) on Page. 93 this can be accounted for by a multiplicative correction



(a) (b) (c)

Fig. 3.31. Edge-emitting and vertically-emitting laser diodes (a) edge-emitting laser diode and far-field radiation characteristic (b) VCSEL layer structure. p-doped DBR mirror: 25 layers $Al_{0.3}Ga_{0.7}As/AlAs$; active zone: 220 nm $Al_{0.3}Ga_{0.7}As$ with 3 $Al_{0.12}Ga_{0.88}As$ quantum films, height about 7 nm each; n-doped DBR mirror: 40 layers $Al_{0.3}Ga_{0.7}As/AlAs$ (c) microscopic image of VCSEL (all after reference Footnote 55 on Page 98)

factor. For $\kappa L > 1.6$ the linewidth of mode $q = 0$ ($f > f_B$, see Fig. 3.30) is up to five times larger than the linewidth of mode $q = -1$. For a standard DFB laser the frequency f_{ind} of the maximum gain should be smaller than the Bragg frequency f_B to enforce oscillation of the mode with the smaller linewidth.

Vertical cavity surface emitting laser

*V*ertical *c*avity *s*urface *e*mitting *l*asers (VCSEL, pronounced ['vɪksel]) are semiconductor lasers which emit perpendicularly to their p-n junction plane in a manner analogous to that of a surface-emitting LED, and feature circular, low-divergence beams. This new class of lasers emerged during the 1990s.[53,54,55] VCSEL operate in a single longitudinal mode due to an extremely small cavity length $L = \lambda_e/2 \approx 1\,\mu$m so that $q = 1$ in Eq. (3.134) on Page 96, see Fig. 3.31(b). The mode spacing $\Delta f_z = c/(n_g\lambda_e) = c/\lambda \approx 300\,$THz for $\lambda = 1\,\mu$m (Eq. (3.68) on Page 79) exceeds the gain bandwidth $\Delta f_H \approx 12\,$THz by far (Eq. (3.66) on Page 3.66).

The properties of VCSEL are attractive for many purposes. Traditional edge-emitting diode lasers only partially achieve these requirements. Such lasers have elliptical, divergent beams which must be optically corrected in order to collimate the beam even over short distances. Furthermore it is often necessary to isolate the laser from back-reflection of the emitted light into the resonator, which can lead to changes in the laser's output characteristics. The divergent, elliptical beam is a result of diffraction at the rectangular emission area of a conventional diode laser resonator. The divergence varies inversely with the size of this emission area. Consequently the beam divergence perpendicular to the junction is significantly greater than parallel to it, see Fig. 3.31(a).

In order to fully control the emitted beam profile it is necessary to define the geometry of the emission area. This is only possible with a vertical resonator, perpendicular to the p-n junction plane, since the active layer is parallel to the emission area, Fig. 3.31(b). To obtain a round, low-divergence beam, a circular output complex may be applied to the entire emitting area. In the case of a vertical resonator the length of the gain medium is defined by the thickness of the p-n junction. In order to achieve laser operation, this must be compensated by the use of highly efficient distributed Bragg reflectors as resonator mirrors. The lower resonator mirror is made up of 40 alternating layers of $Al_xGa_{1-x}As$ and AlAs, each layer $\lambda_e/4$ thick, which give a power reflection factor in excess of $R_1 = 99.99\,\%$. The output top mirror, with twenty-five $\lambda_e/4$ layers,

[53] See Sect. 5.2.4 Page 191–192 in reference Footnote 3 on Page 47

[54] Li, H.; Iga, K.: Vertical-cavity surface-emitting laser devices. New York: Springer 2001

[55] LASER COMPONENTS (UK) Ltd. Goldlay House, 114 Parkway, Chelmsford, Essex. CM2 7PR UK. `http://www.lasercomponents.co.uk`

has a reflectivity of $R_2 = 99.9\,\%$. The high resonator efficiency and small gain medium volume combine to give threshold currents of only a few mA, which means that the low operating currents make VCSEL suitable for any application.

A further advantage of this mirror design is that any light reflected back towards the laser from any other part of the optical system can not re-enter the resonator. VCSEL are effectively isolated against such reflections due to their DBR output coupler structure. The resonator design also means that the laser emission is singlemode. The desired wavelength is achieved by using the correct layer structure.

VCSEL may feature a special mesa structure which ensures that all additional transverse modes are suppressed. The result of this is predominantly singlemode emission with a correspondingly narrow emission linewidth ($< 50\,\mathrm{MHz}$ at $\lambda = 780\,\mathrm{nm}$, for example). This mode will tune within the gain profile as the laser temperature is varied, which allows VCSEL to be temperature-tuned over a range of several nanometers, without mode hopping. The typical tuning rate, about $0.4\,\mathrm{nm}\,/\,\mathrm{mA}$, is significantly higher than in an edge emitter. The vertical structure has a further significant advantage: the laser may be tested before bonding and whilst still on the wafer. Furthermore, VCSEL may be easily incorporated into electronic integrated circuits and are also highly suitable for use in 2-dimensional arrays.

Popular VCSEL are offered[56] with wavelengths in the range $760 \ldots 960\,\mathrm{nm}$. Singlemode output powers are typically $0.3 \ldots 0.5\,\mathrm{mW}$, with $5\,\mathrm{mW}$ multimode versions also available.

Comparison between edge-emitting lasers and VCSEL The technology of VCSEL is showing great promise as a low cost alternative in new applications such as *v*ery *s*hort *r*each (VSR) transceivers, tunable and high power pump lasers. Consequently, the adoption of VCSEL-based optoelectronics could be the ultimate cost reduction vehicle carriers are searching for.[57]

Although advances continue in edge emitter based technology, it seems evident that it just may be a case of diminishing returns. Edge emitter technology was a key enabler of optical communication as we know it today. First used in short reach, single channel applications, the edge emitter evolved to support DWDM and long haul transmission. With the introduction of EDFA, edge emitters adapted to pump laser applications first at 980 nm and later at 1 480 nm. Edge emitters also filled the vital role of high power continuous wave (CW) lasers for lithium niobate modulators and then morphed once again to emerge in an integrated source modulator combo known as electro-absorptive modulated (EAM) lasers. In spite of these amazing accomplishments, edge emitter innovations now seem to be few and far between. In fact, recent improvements have been accompanied by some level of manufacturing/process heroics that come with a hefty yield-cost impact — not exactly the solution carriers are looking for.

To date, VCSEL have earned a reputation as a superior technology for short reach applications such as fibre channel, Ethernet and intra-systems links (among switches, routers and hubs inside central offices). Within the first two years of commercial availability VCSEL displaced edge emitters in the local area networks. Consequently, since VCSEL were first adopted in short reach applications, many have prematurely concluded that VCSEL are limited to low power, short wavelength applications. Conversely, a few companies are poised to prove otherwise as recent announcements suggest significant strides in the areas of 1 310 nm source lasers, 1 550 nm tunable lasers and finally high power pump lasers.

VCSEL are grown, processed and tested while still in the wafer form. As such, there is significant economy of scale resulting from the ability to conduct parallel device processing, whereby equipment utilization and yields are maximized and set up times and labor content are minimized. In the case of a VCSEL, the mirrors and active region are sequentially stacked along the x-axis (perpendicularly to the pn-junction, see Fig. 3.6 on Page 58) during epitaxial growth. The VCSEL wafer then goes through etching and metallization steps to form the electrical contacts. At this point the wafer goes to test where individual laser devices are characterized on a pass-fail basis.

[56] See reference Footnote 55 on Page 98

[57] Hays, T.: A new breed of laser emerges on the optical frontier. Fiber Optic Technology — formerly Fiberoptic Products News.(2005). The following section is literally quoted from `http://fiberoptictechnology.net/Scripts`.

Finally, the wafer is diced and the lasers are binned for either higher-level assembly (typically $> 95\,\%$) or scrap (typically $5\,\%$).

In a simple Fabry-Perot edge emitter (DFB edge emitters require additional etch and re-growth steps) the growth process also occurs along the x-axis, but only to create the active region, as mirror coatings are later applied along the y-axis (in the junction plane). After epitaxial growth, the wafer goes through the metallization step and is subsequently cleaved along the y-axis, forming a series of wafer strips. The wafer strips are then stacked and mounted into a coating fixture. The x-axis edges of the wafer strips are then coated to form the device mirrors. Now the wafer strips are diced to form discrete laser chips, which are then mounted onto carriers. Finally, the laser devices go in to test where typically more than $50\,\%$ of DFB are scrapped.

It is also important to understand that VCSEL consume less material. In the case of a 3-inch wafer, a laser manufacturer can build about 15 000 VCSEL or approximately 4 000 edge emitters. Considering the $2:1$ yield advantage (DFB edge emitter) combined with a $4:1$ wafer throughput edge, the VCSEL cost advantage is obvious.

The main disadvantage comes from the fact that VCSEL are routinely fabricated only in the short-wavelength region $\lambda < 1\,\mu$m which prevents applications in long-haul transmission systems operating near the fibre loss minimum at $\lambda = 1.55\,\mu$m. However, recent progress lead to VCSEL at $1.55\,\mu$m with threshold currents of $0.5\,$mA. If biased at $5\,$mA they can be directly modulated with $2.5\,$Gbit/s for a robust upstream WDM transmission in a low-cost passive optical access network.[58]

[58] Wong, E.; Zhao, X.-x.; Chang-Hasnain, C. J.; Hofmann, W.; Amann, M. C.: Uncooled, optical injection-locked $1.55\,\mu$m VCSELs for upstream transmitters in WDM-PONs. Technical Digest Optical Fiber Communication Conference (OFC'06), Anaheim (CA), USA, 05.–10.03.2006. Postdeadline Paper PDP50

Chapter 4

Optical amplifiers

The transmission distance of a fibre-optic communications system is limited by fibre loss and dispersion. For long-haul lightwave systems, the loss limitation[1] has traditionally been overcome using optoelectronic repeaters in which the optical signal is first converted into an electric current and then regenerated using a transmitter. Such regenerators become quite complex and expensive for multichannel lightwave systems. An alternative approach makes use of optical amplifiers, which amplify the optical signal directly without requiring its conversion to the electric domain.

Optical amplifiers amplify incident light through stimulated emission, the same mechanism as that used by lasers, see Sect. 3.2.5 on Page 68. Indeed, an optical amplifier is nothing but a laser without feedback. Its main ingredient is the optical gain realized when the amplifier is pumped to achieve population inversion.[2]

4.1 Semiconductor amplifier

Starting from the gain relations (3.73)–(3.75) on Page 80 as discussed for the Fabry-Perot laser, we formulate the equations for the gain of an semiconductor optical amplifier (SOA) with residual mirror reflectivities $R_{1,2} \neq 0$. A transition $R_{1,2} \to 0$ leads to a true travelling-wave amplifier. If $R_{1,2} \ll 1$ holds as it is the case for real devices, we talk of a near-travelling-wave amplifier (TWA). However, now the concept of a spatial average of gain and loss must not be applied any more.

Frequently, the symbol G is used for the power gain of an amplifier.[3,4] However, we already associated the character G with the gain rate, see Eq. (3.43) on Page 70. Therefore the calligraphic symbol $\mathcal{G}$ stands for the amplifier power gain in the following text.

Assume an amplifying waveguide region with length L, having a propagation constant β and an effective (modal) refractive index n_e according to Eq. (2.37) on Page 22. For the single-pass power gain $\mathcal{G}_s$ and the phase shift φ along the amplifying region we know from Eqs. (3.73)–(3.77)

$$\mathcal{G}_s = \exp\left[(\Gamma g - \alpha_{Ve})\, L\right], \qquad \varphi = \beta L = k_0 n_e L\,. \tag{4.1}$$

Taking into account the multiple reflections at the mirrors, the amplification factor $\mathcal{G}$ is obtained using the standard theory of a Fabry-Perot interferometer,[5]

$$\mathcal{G} = \frac{\mathcal{G}_s(1-R_1)(1-R_2)}{\left(1-\mathcal{G}_s\sqrt{R_1R_2}\,\right)^2 + 4\mathcal{G}_s\sqrt{R_1R_2}\,\sin^2\varphi}\,, \tag{4.2}$$

$$\varphi = \beta L, \qquad \text{resonances Eq. (3.67): } \varphi_z = \omega_z n_e L/c = m_z\pi\,, \quad m_z = 1, 2, 3, \ldots$$

[1] See Footnote 20 on Page 6
[2] See Chapter 8 Page 361 in reference Footnote 9 on Page 4
[3] See Sect. 8.2. Page 368 ff. in reference Footnote 9 on Page 4
[4] See Sect. 5.5. Page 209 ff. in reference Footnote 3 on Page 47
[5] See references in Footnotes 6, 7, 1 on Pages 48 and 7

As is evident from Eq. (4.2), $\mathcal{G}$ peaks whenever the frequency $f = \omega/(2\pi)$ coincides with one of the cavity-resonance frequencies f_z and drops sharply in between them. Because spontaneous emission was disregarded, we find infinite gain for $\mathcal{G}_s\sqrt{R_1R_2} = 1$ at the resonance points $\varphi_z = m_z\pi$ spaced apart by the free spectral range $\Delta f_z = c/(2n_{eg}L)$, see Eq. (3.68) on Page 79 and Fig. 3.26 on Page 93. The usable gain, however, is limited by the existence of spontaneous emission in the region of the lasing threshold, Eq. (3.86) on Page 83. The power gain $\mathcal{G}$ depends on the field confinement factor Γ, Eq. (4.1) and following remarks on Page 101, and on the carrier concentration n_T, Eq. (3.80) on Page 81. Therefore, $\mathcal{G}$ changes with the polarization of the field (the gain for TE polarization is about $5 \ldots 10$ dB larger in bulk amplifiers than for TM polarization), with the injection current I via n_T, Eq. (3.93) on Page 85, and, because of gain saturation, it varies also with the power level of the input signal. Further, because of amplitude-phase coupling, Eq. (3.126) on Page 92, we find a signal-dependent phase shift of the optical output (a chirp of its frequency). The signal power dependency and/or the chirp may be exploited for frequency conversion.[6,7,8,9,10]

At resonance and anti-resonance we find the maximum and minimum gain factors

$$\mathcal{G}_{\max} = \frac{\mathcal{G}_s(1-R_1)(1-R_2)}{(1-\mathcal{G}_s\sqrt{R_1R_2})^2}, \qquad \mathcal{G}_{\min} = \frac{\mathcal{G}_s(1-R_1)(1-R_2)}{(1+\mathcal{G}_s\sqrt{R_1R_2})^2}\,. \tag{4.3}$$

From the ripple of the gain curve the single-pass gain $\mathcal{G}_s\sqrt{R_1R_2}$ may be derived,[11,12,13]

$$\mathcal{G}_s\sqrt{R_1R_2} = \frac{\sqrt{\mathcal{G}_{\max}/\mathcal{G}_{\min}}-1}{\sqrt{\mathcal{G}_{\max}/\mathcal{G}_{\min}}+1}\,. \tag{4.4}$$

For a 3 dB ripple we have $\mathcal{G}_s\sqrt{R_1R_2} = 0.17$, so for a single-pass gain $10\lg\mathcal{G}_s = 20$ dB the mean mirror reflection factor must be $\sqrt{R_1R_2} < 0.17 \times 10^{-2}$. Variations in the single-pass gain $\mathcal{G}_s$ have the more effect the larger $\sqrt{R_1R_2}$ is.

4.1.1 Fabry-Perot amplifier

The bandwidth of a Fabry-Perot amplifier (FPA) is determined by the sharpness of the cavity resonance at f_z. The spectral distance between the half-maximum points of the gain $\mathcal{G}$ inside one single Fabry-Perot mode m_z in Fig. 3.26 can be computed from Eqs. (4.2), (4.3),

$$B_{\mathcal{G}} = \frac{c}{\pi n_{eg}L}\arcsin\left(\frac{1-\mathcal{G}_s\sqrt{R_1R_2}}{\sqrt{4\mathcal{G}_s\sqrt{R_1R_2}}}\right) = \frac{c}{\pi n_{eg}L}\arcsin\sqrt{\frac{(1-R_1)(1-R_2)}{4\mathcal{G}_{\max}\sqrt{R_1R_2}}}\,. \tag{4.5}$$

Measuring $B_{\mathcal{G}}$ is an alternative method to determine $\mathcal{G}_s\sqrt{R_1R_2}$. Usually, the arcsin-function may be approximated by its argument. For a Fabry-Perot amplifier the maximum power gain $\mathcal{G}_{\max}$

[6] Nielsen, M. L.; Nord, M.; Petersen, M. N.; Dagens, B.; Labrousse, A.; Brenot, R.; Martin, B.; Squedin, S.; Renaud, M.: 40 Gbit/s standard-mode wavelength conversion in all-active MZI with very fast response. Electron. Lett. 39 (2003) 20th Feb., No. 4

[7] Nielsen, M. L.; Lavigne, B.; Dagens, B.: Polarity-preserving SOA-based wavelength conversion at 40 Gbit/s using bandpass filtering. Electron. Lett. 39 (2003) 4th Sep., No. 18

[8] Leuthold, J.; Ryf, R.; Maywar, D. N.; Cabot, S.; Jaques, J.; Patel, S. S.: Nonblocking all-optical cross connect based on regenerative all-optical wavelength converter in a transparent demonstration over 42 nodes and 16 800 km. IEEE J. Lightwave Technol. 21 (2003) 2863–2870

[9] Leuthold, J.; Marom, D. M.; Cabot, S.; Jaques, J. J.; Ryf, R.; Giles, C. R.: All-optical wavelength conversion using a pulse reformatting optical filter. IEEE J. Lightwave Technol. 22 (2004) 186–192

[10] Leuthold, J.; Möller, L.; Jaques, J.; Cabot, D.; Zhang, L.; Bernasconi, P.; Cappuzzo, M.; Gomez, L.; Laskowski, E.; Chen, E.; Wong-Foy, A.; Griffin, A.: 160 Gbit/s SOA all-optical wavelength converter and assessment of its regenerative properties. Electron. Lett. 40 (2004) 29th Apr., No. 9

[11] Hakki, B. W.; Paoli, T. L.: Gain spectra in GaAs double-heterostructure injection lasers. J. Appl. Phys. 46 (1975) 1299–1306

[12] Guo, Wei-Hua; Huang, Yong-Zhen; Han, Chun-Lin; Yu, Li-Juan: Measurement of gain spectrum for Fabry–Pérot semiconductor lasers by the Fourier transform method with a deconvolution process. IEEE J. Quantum Electron. 39 (2003) 716–721

[13] Fazludeen, R.; Samit Barai; Prasant Kumar Pattnaik; Srinivas, T.; Selvarajan, A.: A novel technique to measure the propagation loss of integrated optical waveguides. IEEE Photonics Technol. Lett. 17 (2005) 360–362

changes with the operating current, so the product of bandwidth $B_\mathcal{G}$ in mode m_z and maximum amplitude gain $\sqrt{\mathcal{G}_{\max}}$ remains constant,

$$B_\mathcal{G}\sqrt{\mathcal{G}_{\max}} = \frac{c}{2\pi n_{eg}L}\sqrt{\frac{(1-R_1)(1-R_2)}{\sqrt{R_1R_2}}} = \frac{\Delta f_z}{\pi}\sqrt{\frac{(1-R_1)(1-R_2)}{\sqrt{R_1R_2}}} = \text{const}\ . \tag{4.6}$$

As before, neighbouring modes are spectrally separated by Δf_z, Eq. (3.68). For semiconductor lasers without antireflection coated facets the power reflection factors are $R_1 = R_2 = 0.32$. If the crystal length is $L = 300\,\mu$m and the effective group index $n_{eg} = 3.5$, a bandwidth-gain product of $B_\mathcal{G}\sqrt{\mathcal{G}_{\max}} = 55\,$GHz results. Such a small bandwidth makes Fabry-Perot amplifiers unsuitable for most lightwave system applications.

4.1.2 Travelling-wave amplifier

Antireflection coatings are necessary for travelling-wave amplifiers (TWA). Minimum reflectivities of $\sqrt{R_1R_2} = 10^{-5}$ are achieved, but considerable technological effort is required. For this reason, alternative techniques help reducing the reflection feedback. In one method,[14] the active-region stripe is tilted from the facet normal. If the vertical bars | denote the facets and the dash — or the slash ⁄ represent the stripe, the conventional amplifier is characterized by |—|, while the angled-facet or tilted-strip structure appears like |⁄|. In practice, tilted stripes with antireflection-coated facets have effective reflectivities down to $R_{1,2} \approx 10^{-4}$. The 3 dB bandwidth of typical amplifiers is $\Delta\lambda_H = 70\,$nm near $\lambda = 1.5\,\mu$m. This corresponds to a frequency bandwidth of $\Delta f_H = 9.3\,$THz and compares well with the rough estimate $\Delta f_H = 12.1\,$THz in Eq. (3.66) on Page 78. A gain of 28 dB with a residual ripple $< 3\,$dB and a polarization dependence of the gain $< 1\,$dB is possible.

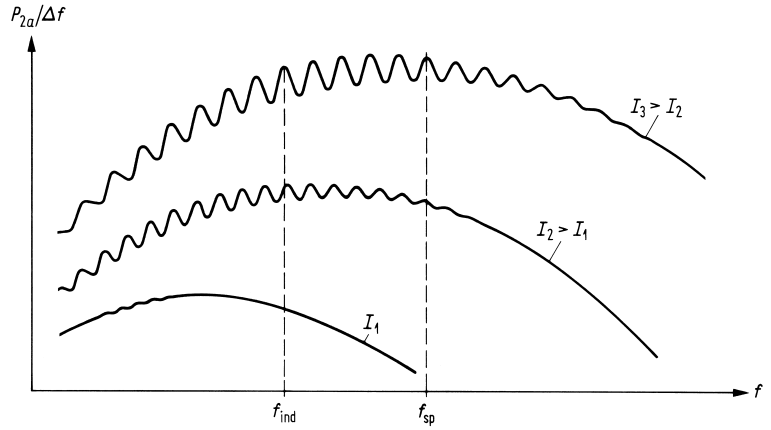


Fig. 4.1. Near-travelling-wave amplifier. Schematic of the spectral output power density $P_{2a}/\Delta F$ of amplified spontaneous emission as transmitted through mirror R_2 for varying injection currents $I_1 < I_2 < I_3$. The frequencies of maximum gain and maximum spontaneous emission are denoted as f_{ind} and f_{sp} for an operating current $I = I_3$.

Figure 4.1 displays the schematic spectral output power density $P_{2a}/\Delta f$ at the exit facet of a near-travelling-wave amplifier. Because of bandfilling at larger injection currents the quasi Fermi levels move deeper into the bands, and the frequencies of maximum spontaneous emission f_{sp} and of maximum induced amplification f_{ind} shift to higher frequencies. This may be also deduced from the diagrams Fig. 3.15 on Page 71, where the point of zero gain $x_0 = (W_{Fn} - W_{Fp} - W_G)/(kT_0)$ depends on the difference $W_{Fn} - W_{Fp}$ of the quasi Fermi levels. As seen in Fig. 3.15, the relation $f_{\text{ind}} < f_{\text{sp}}$ holds.

[14] Zah, C. E.; Osinski, J. S.; Caneau, C.; Menocal, S. G.; Reith, L. A.; Salzman, J.; Shokoohi, F. K.; Lee, T. P.: Electron. Lett. 23 (1987) 990

4.2 Doped fibre amplifier

An alternative to current-pumped semiconductor laser amplifiers are optically pumped $E\mathrm{r}^{3+}$-*d*oped glass *f*ibre *a*mplifiers (EDFA). The amplification process is described in Fig. 3.3 on Page 54, 54. The optical bandwidth B_O is in the order $B_{\mathrm{OA}} = 4\,\mathrm{THz}$ for an ordinary EDFA. They are commercially available at $\lambda_S = 1.55\,\mu\mathrm{m}$ ($f_S = 193\,\mathrm{THz}$) in the S band (*s*hort-wavelength band, $\lambda \leq 1.528\,\mu\mathrm{m}$), C band (*c*onventional or *c*entral band, $1.528\,\mu\mathrm{m} \leq \lambda \leq 1\,563\,\mu\mathrm{m}$, $\Delta\lambda = 35\,\mathrm{nm}$), and in the L band (*l*ong-wavelength band, $1.563\,\mu\mathrm{m} \leq \lambda \leq 1.606\,\mu\mathrm{m}$, $\Delta\lambda = 43\,\mathrm{nm}$). Combining a C and an S band amplifier[15] the following record results were achieved, Fig. 4.2:

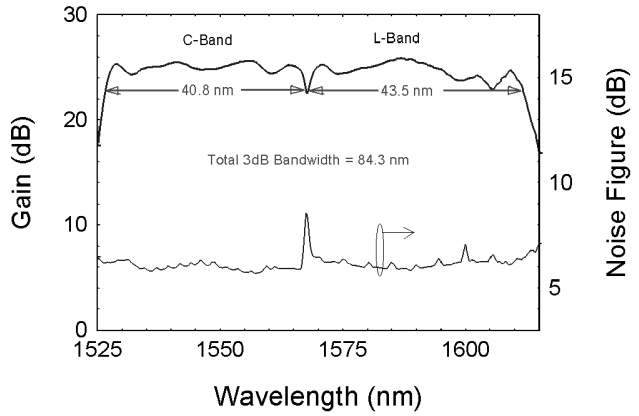


Fig. 4.2. Er-doped optical ultra-broadband amplifier[15](ECOC'1998)

- With 84.3 nm of bandwidth this ultra wide band amplifier can accommodate 100 WDM (*w*avelength *d*ivision *m*ultiplexing) channels with the proposed ITU (*I*nternational *T*elecommunication *U*nion) standard channel spacing of 100 GHz (0.8 nm @ $\lambda_S = 1.55\,\mu\mathrm{m}$), or 200 WDM channels with 50 GHz (0.4 nm @ $\lambda_S = 1.55\,\mu\mathrm{m}$). There is enough power to support 200 WDM channels. On the other hand, for a given number of WDM channels, this amplifier can be used to allow for a wider channel spacing which may be needed in optical networks with multiple cascaded filters.
- UWB EDFA with a two-section, split band structure. Maximum operating gain is 25 dB and the noise figure is about 6 dB. The output power is very high at 25 dBm, which is required for large numbers of WDM channels. The split bands allow independent optimization of each band for dispersion compensation and span loss variations. The present EDFA, based on field tested erbium-doped silica fiber technology, can be used in Tbit/s capacity DWDM (dense WDM) systems and networks.
- Bandwidth $\Delta\lambda = 84\,\mathrm{nm}$ ($\Delta f = 10\,\mathrm{THz}$ @ $\lambda_S = 1.563\,\mu\mathrm{m}$)
- Noise figure $F = 6\,\mathrm{dB}$, power gain $\mathcal{G}_s = (25 \pm 1.5)\,\mathrm{dB}$, output power 25 dBm $\widehat{=}$ 320 mW
- 100 channels @ 10 Gbit/s @ 400 km, i. e., a length-bandwidth product of 400 Tbit/s · km (!). The amplifier bandwidth B_{OA} centred at the optical frequency f_S is usually much larger than the signal bandwidth, $B \ll B_{\mathrm{OA}} \ll f_S$. With passive optical filters the optical bandwidth may be varied in the range $100\,\mathrm{GHz} \leq B_O \leq B_{\mathrm{OA}}$.

[15]Sun, Y.; Sulhoff, J. W.; Srivastava, A. K.; Abramov, A.; Strasser, T. A.; Wysocki, P. F.; Pedrazzani, J. R.; Judkins, J. B.; Espindola, R. P.; Wolf, C.; Zyskind, J. L.; Vengsarkar, A. M.; Zhou, J.: A gain-flattened ultra wide band EDFA for high capacity WDM optical communications system. Proc. 24th Europ. Conf. Opt. Commun. Madrid (ECOC 1998), 20.–24. Sept. 1998. Vol. 1 pp. 53–54 (Lucent Technologies — Bell Laboratories, Holmdel, NJ)

Chapter 5

Pin photodiode

The role of an optical receiver is to convert the optical signal back into electrical form and to recover data transmitted through the lightwave system. Its main component is a photodetector that converts light into electricity through the photoelectric effect. Because of its speed and sensitivity photoconductive detectors in the form of reverse-biased semiconductor pn-junctions are commonly used for lightwave systems. The following sections focus on the pin-photodiode. With *a*valanche *p*hoto*d*iodes (APD) the responsivity may be increased by impact ionization at the cost of additional avalanche noise. Usually, the combination of optical amplifiers and pin-photodiodes yields a better sensitivity and linearity than the use of APD.

5.1 Basic relations

A reverse-biased pn-junction consists of a region, known as the depletion or space-charge region, that is essentially devoid of free charge carriers and where a large built-in electric field (Eq. (3.34)) opposes flow of electrons from the n-side to the p-side and of holes from the p-side to the n-side. When such a pn-junction is illuminated with light, electron-hole pairs are created through absorption of photons with energies $hf_S > W_G$ (light frequency f_S) larger than the bandgap energy W_G. The probability that a photon generates an electron-hole pair is the quantum efficiency η. Because of the large built-in electric field, electrons and holes generated inside the depletion region accelerate in opposite directions and drift to the regions where they are majority carriers.[1,2]

A limiting factor for the bandwidth of pn-photodiodes is the presence of a diffusive component in the photocurrent. Electrons generated in the p-region have to diffuse to the depletion-region boundary before they can drift to the n-side; similarly, holes generated in the n-region must diffuse to the depletion-region boundary. Diffusion is an inherently slow process. Carriers take 1 ns or longer to diffuse over a distance of about 1 μm. In practice, the diffusion contribution depends on the bit rate and becomes negligible by decreasing the widths of the p- and n-regions and increasing the depletion-region width so that most of the incident optical power is absorbed inside it. This is the approach adopted for pin-photodiodes, discussed next.[3]

A simple way to increase the depletion-region width is to insert a layer of undoped (or lightly doped) semiconductor material between the pn-junction. Since the middle layer consists of (nearly) intrinsic material, such a structure is referred to as the pin-photodiode.[4] Figure 5.1 shows the device structure (not drawn to scale) together with the electric-field distribution inside it under reverse-bias operation. Because of its intrinsic nature, the middle i-layer offers a high resistance, and most of the voltage drop occurs across it. As a result, a large electric field exists in the i-layer. In essence, the depletion region extends throughout the i-region $-w_A \leq x \leq 0$, and the width

[1] See Sect. 4.2.1 Page 141 in reference Footnote 9 on Page 4
[2] See Chapter 7 Page 253 ff. in reference Footnote 3 on Page 47
[3] See Sect. 4.2.1 Page 143 in reference Footnote 9 on Page 4
[4] See Sect. 4.2.2 Page 144 in reference Footnote 9 on Page 4

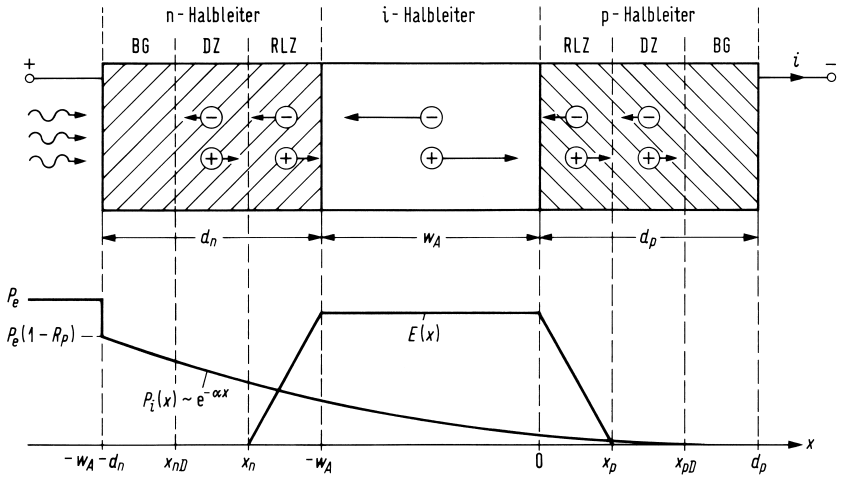


Fig. 5.1. Schematic of a pin-diode. BG contact region (= *B*ahn*g*ebiet), DZ diffusion zone, RLZ space-charge (or depletion) region (= *R*aum*l*adungs*z*one). P_e light power incident from region external of semiconductor, R_P power reflection factor of the semiconductor surface, $P_i(x)$ light power inside the semiconductor, α light power attenuation constant, d_n (d_p) length of n-doped (p-doped) semiconductor, w_A length of intrinsic absorption zone, $E(x)$ x-component of electric field. Halbleiter = semiconductor

w_A of this absorption zone can be controlled by changing the middle-layer thickness. The main difference from the pn-photodiode is that the drift component of the photocurrent dominates over the diffusion component simply because most of the incident power is absorbed inside the i-region (absorption region) of the pin-photodiode.

The external light power P_e is incident perpendicularly to the PD surface and partially reflected with a power reflection factor R_P, see Eq. (2.35) on Page 19 where $R_P = R_E = R_H$. The power inside the semiconductor is attenuated exponentially. The absorption length corresponding to the reciprocal attenuation constant α for direct semiconductors (e. g., GaAs) is in the order $1/\alpha = 1\,\mu$m, for indirect semiconductors (e. g., Si) it is $1/\alpha = 10 \ldots 20\,\mu$m, see also Fig. 5.4 on Page 110. A high-speed operation requires small carrier transit times in the depletion region, i. e., small w_A. However, because the depletion-layer capacitance increases with decreasing w_A, an optimum absorption layer width results, and an optimum absorbed power $P_e[1 - \exp(-\alpha w_A)]$ for $R_P = 0$, $d_n = d_p = 0$ has to be found. Thus, high speed (small w_A) means low quantum efficiency and low sensitivity. To avoid this dilemma, an optical waveguide structure can be used to which the optical signal is edge-coupled, see Fig. 5.7 on Page 115. When the power $P_i(x)$ inside the absorption region changes significantly with x, then the light should be radiated into the direction of the faster moving charge carriers because in this case a larger percentage of the carriers leaves the absorption region faster. In Fig. 5.1 the holes are assumed to drift at a higher velocity than the electrons, $v_p > v_n$. This rule is not necessarily obeyed if discontinuities in the bandedges of heterostructures inhibit electrons or holes from leaving the i-region.

5.1.1 Short-circuit photocurrent

We assume a semiconductor without external magnetic field and neglect the magnetic fields associated with flowing currents. The basic equations[5] for semiconductor-device operation can be classified in three groups:[6] Continuity equations, current-density or transport equations, and Maxwell's equations. For the *continuity equations* we need the electron and hole concentrations n_T and p, the electron and hole (convection) current densities $\vec{J}_n$ and $\vec{J}_p$, the generation (g_n and g_p) and the

[5] Sze, S. M.: Physics of semiconductor devices. New York: John Wiley & Sons 1985. Chapter 3 Page 70
[6] See Sect. 5.4.2 Page 196 in reference Footnote 23 on Page 30

recombination rates (r_n and r_p) of electrons and holes,

$$\begin{aligned} \partial p/\partial t + \text{div}\ \vec{J}_p/e &= g_p - r_p\,, \\ \partial n_T/\partial t - \text{div}\ \vec{J}_n/e &= g_n - r_n\,. \end{aligned} \tag{5.1}$$

The current-density or *transport equations* consist of the drift component caused by the field and the diffusion component, which in turn is determined by the carrier concentration gradient. Parameters are the drift velocities $\vec{v}_n$ and $\vec{v}_p$, the diffusion constants D_n and D_p, the mobilities μ_n and μ_p for electrons and holes, and the electric field $\vec{E}$,

$$\begin{aligned} \vec{J}_p &= ep\vec{v}_p \quad - eD_p\ \text{grad}\ p, \qquad \vec{v}_p = \mu_p\vec{E}, \\ \vec{J}_n &= -en_T\vec{v}_n + eD_n\ \text{grad}\ n_T, \qquad \vec{v}_n = -\mu_n\vec{E}, \end{aligned} \tag{5.2}$$

From Maxwell's equations the most important is the Poisson equation which determines the properties of the pn-junction depletion layer. Poisson's equation connects the electric field $\vec{E}$, the total electric charge density ρ, the dielectric displacement $\vec{D}$, and the permittivity ϵ,

$$\text{div}\,\vec{D} = \rho, \qquad \vec{D} = \epsilon\vec{E}. \tag{5.3}$$

The total current density is source-free,

$$\text{div}\left(\vec{J}_n + \vec{J}_p + \frac{\partial\vec{D}}{\partial t}\right) = 0. \tag{5.4}$$

Equations (5.1)–(5.4) are applied to an i-semiconductor Fig. 5.2 for the one-dimensional case. The electric field due to the externally connected voltage is so high that electrons and holes drift with their saturation velocities v_n and v_p. The recombination rates r_n and r_p are negligibly small

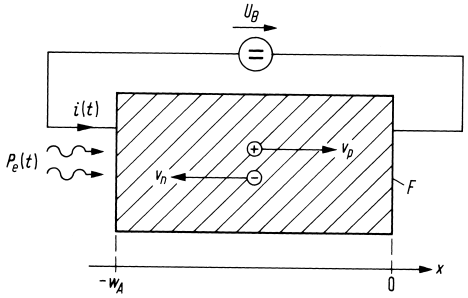


Fig. 5.2. i-layer of a pin-photodiode (one-dimensional case, cross-section area F). Saturation drift velocities $v_n > 0$ and $v_p > 0$ for electrons and holes, incident external optical power $P_e(t)$, total conduction current $i(t)$, open-circuit voltage U_B of a battery with an internal resistance of zero

because the carrier lifetime is much larger than the drift time in the absorption zone $-w_A \le x \le 0$. Diffusion currents may be neglected compared to drift currents. Photogeneration dominates the carrier generation processes, $g_p = g_n = g$. Introducing convection and conduction currents i instead of current densities J (cross-section area F) we find from Eq. (5.1)–(5.4)

$$\begin{aligned} \frac{1}{v_p}\frac{\partial i_p}{\partial t} + \frac{\partial i_p}{\partial x} &= eFg\,, \qquad i_p = Fepv_p\,, \\ \frac{1}{v_n}\frac{\partial i_n}{\partial t} - \frac{\partial i_n}{\partial x} &= eFg\,, \qquad i_n = Fen_Tv_n\,, \end{aligned} \tag{5.5}$$

and

$$\epsilon\frac{\partial E}{\partial x} = e(p - n_T), \qquad \frac{\partial}{\partial x}\left(i_n + i_p + F\epsilon\frac{\partial E}{\partial t}\right) = 0\,. \tag{5.6}$$

The total time-dependent conduction current is

$$i(t) = i_n(x,t) + i_p(x,t) + F\epsilon \frac{\partial E(x,t)}{\partial t}\,. \tag{5.7}$$

Because of

$$\int_{-w_A}^{0} E(x,t)\,\mathrm{d}x = U_B = \text{const} \quad \rightarrow \quad \int_{-w_A}^{0} \frac{\partial E(x,t)}{\partial t}\,\mathrm{d}x = \frac{\mathrm{d}\,U_B}{\mathrm{d}t} = 0\,, \tag{5.8}$$

we find the total conduction current in the external circuit (i. e., the external short-circuit current, $\mathrm{d}\,U_B = 0$) as an average of the sum of the carrier convection currents in the drift region $-w_A \leq x \leq 0$,

$$i(t) = \frac{1}{w_A} \int_{-w_A}^{0} [\, i_n(x,t) + i_p(x,t)\,]\,\mathrm{d}x\,. \tag{5.9}$$

The carrier transit times τ_n, τ_p are defined by

$$w_A = v_n\tau_n = v_p\tau_p\,. \tag{5.10}$$

The total number of electrons and holes in the semiconductor are

$$N_n(t) = F \int_{-w_A}^{0} n_T(x,t)\,\mathrm{d}x\,, \qquad N_p(t) = F \int_{-w_A}^{0} p(x,t)\,\mathrm{d}x\,. \tag{5.11}$$

From Eqs. (5.9)–(5.11) we calculate with the help of Eq. (5.5) ($i_p = Fepv_p$, $i_n = Fen_Tv_n$) the total conduction current,

$$i(t) = \frac{e}{\tau_n} N_n(t) + \frac{e}{\tau_p} N_p(t). \tag{5.12}$$

If the irradiated n-region is much shorter than the absorption length (i. e., $\alpha\, d_n \rightarrow 0$ in Fig. 5.1), the light power in the i-region $P_i(x,t)$ reads

$$P_i(x,t) = P_e(t)\,(1 - R_P)\,\mathrm{e}^{-\alpha(x+w_A)}\,. \tag{5.13}$$

Any light propagation times are neglected. P_e, P_i are classical optical powers averaged over a few optical cycles. The power fraction which is absorbed inside the i-zone represents the quantum efficiency η of the photodiode,

$$\eta = \frac{P_i(-w_A,t) - P_i(0,t)}{P_e(t)} = (1 - R_P)\left(1 - \mathrm{e}^{-\alpha w_A}\right). \tag{5.14}$$

The mean generation rate of electron-hole pairs equals the mean absorption rate of photons, i. e., the absorbed power per quantum energy hf_S. Because photodiodes cannot emit optical power, no second harmonic light frequency is generated as it is common for classical microwave detectors. The generation rate g (unit $\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$) is

$$g(x,t) = -\frac{1}{Fhf_S}\frac{\partial P_i(x,t)}{\partial x} = \frac{\alpha P_i(x,t)}{Fhf_S}\,. \tag{5.15}$$

With Eq. (5.13) (substitute $P_i(x,t)$) and Eq. (5.14) (substitute $(1 - R_P)$) we find

$$eFg(x,t) = \frac{\eta e}{hf_S}\,P_e(t)\,\frac{\alpha\,\mathrm{e}^{-\alpha(x+w_A)}}{1 - \mathrm{e}^{-\alpha w_A}}\,. \tag{5.16}$$

5.1.2 Equivalent electrical circuit

For $\mathrm{d}/\mathrm{d}t = 0$ with $P_e(t) \equiv P_e = \mathrm{const}_t$ the Eqs. (5.5), (5.16) for the static short-circuit current $i(t) \equiv i = \mathrm{const}_t$ may be easily solved. We integrate Eq. (5.5), (5.16) with the notation $f(x,t) \equiv f(x)$ for space and time dependent functions f in the stationary case. Further, we observe that in Figs. 5.1, 5.2 the minority current injection may be neglected, $i_p(-w_A) = 0$, $i_n(0) = 0$,

$$i = i_p(0) = i_n(-w_A) = \int_{-w_A}^{0} eFg(x)\,\mathrm{d}x = \frac{\eta e}{hf_S}\,P_e\,, \quad i = SP_e\,, \quad S = \frac{\eta e}{hf_S}\,, \quad \frac{S}{\mathrm{A/W}} = 0.806\,\eta\,\frac{\lambda_S}{\mu\mathrm{m}}\,. \tag{5.17}$$

The quantity S is called photodetector sensitivity (responsivity). The absorbed power is ηP_e and corresponds to a photon absorption rate of $\eta P_e/(hf_S)$. Each absorbed photon generates an electron-hole pair leading to the transport of one elementary charge e through the external circuit. The rate of generated charges i/e equals the photon absorption rate $\eta P_e/(hf_S)$.

For the time-dependent case we substitute Eq. (5.16) in Eq. (5.5). The current $i(t)$ as calculated from Eq. (5.9) or Eq. (5.12) represents the short-circuit current which feeds the equivalent circuit of the electrical embedding network of the photodiode Fig. 5.3. The Fourier transforms Eq. (2.2)

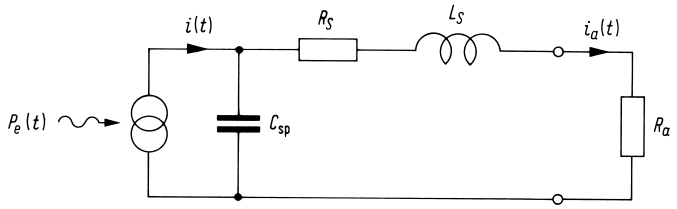


Fig. 5.3. Equivalent electrical circuit of a photodiode. P_e incident external optical power, C_{sp} depletion-layer capacity, R_S series resistance, L_S series inductance, R_a load resistance

of $i(t)$, $i_a(t)$ are denoted as $I(f)$, $I_a(f)$. The transfer function $H_S(f)$ from the photodiode to the load resistance R_a reads

$$H_S(f) = \frac{I_a(f)}{I(f)} = \frac{\omega_r^2}{(\mathrm{j}\,\omega)^2 + 2\gamma_r(\mathrm{j}\,\omega) + \omega_r^2}\,, \qquad \omega_r^2 = \frac{1}{L_S C_{\mathrm{sp}}}\,, \qquad 2\gamma_r = \frac{R_S + R_a}{L_S}\,. \tag{5.18}$$

The transfer function $H_S(f)$ has the same structure as the small-signal transfer function of the current-modulated laser diode, Eq. (3.105), (3.106) on Page (87) and Fig. 3.22. Typical values of the circuit elements are: $C_{\mathrm{sp}} = 0.04 \ldots 0.2\,\mathrm{pF}$, $R_S = 10 \ldots 50\,\Omega$, $L_S = 0.15 \ldots 0.5\,\mathrm{nH}$, $R_a = 50\,\Omega$. The reverse diode voltage depends on the material and on the width if the depletion-layer and is in the order of a few volts. The depletion-layer capacitance C_{sp} may be calculated with the formula for a parallel-plate capacitor, Eq. (5.40). Relative dielectric constants ϵ_r are specified in Sect. 5.2. The cross-sectional area F of a photodiode is usually circular having a diameter in the order $7 \ldots 200\,\mu\mathrm{m}$, typically in the range $10 \ldots 80\,\mu\mathrm{m}$.

5.2 Materials

Materials commonly used to make photodiodes may be elemental or compound semiconductors. From the elemental semiconductors Eq. (3.16) Ge ($\epsilon_r = 16$) is suitable in the long-wavelength region (direct semiconductor for $\lambda < 1.55\,\mu\mathrm{m}$, indirect for $\lambda < 1.85\,\mu\mathrm{m}$), while Si exhibits excellent

properties in the short-wavelength and visible range ($\epsilon_r = 11.7$, direct for $\lambda < 0.36\,\mu$m, indirect for $\lambda < 1.1\,\mu$m).

Compound semiconductors (see Tables 3.1, 3.2) are common in the long-wavelength domain. InP substrates (transparent for $\lambda > 0.92\,\mu$m, $\epsilon_r = 12.35$) with an lattice-matched $In_{0.53}Ga_{0.47}As$ absorption layer are used (in short InGaAs, good for $\lambda < 1.65\,\mu$m, $\epsilon_r = 13.6$).

The bandgap difference $|\Delta W_G| = 0.6\,$eV (InP: $W_G = 1.35\,$eV; InGaAs: $W_G = 0.75\,$eV) leads to discontinuities for the CB edge of $|\Delta W_L| = 0.2\,$eV and for the VB edge of $|\Delta W_V| = 0.4\,$eV, see Eq. (3.33) and Fig. 3.13 on Page 67. For an isotype nN-junction between weakly n-doped InGaAs and n-doped InP the built-in voltage of typically $U_D = 0.22\,$eV has to be added. Thus, the CB edges of InGaAs and InP on both sides of the contact have nearly the same energy levels. The remaining CB spike is narrow, and electrons may easily penetrate the barrier by tunneling. However, VB holes injected from InGaAs into InP see a VB potential barrier which increased to $|\Delta W_V| + U_D = 0.62\,$eV.

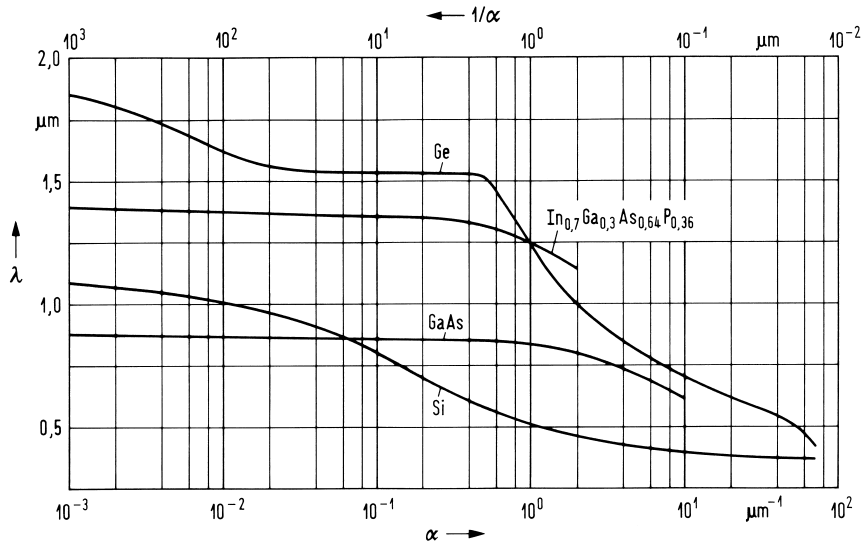


Fig. 5.4. Wavelength dependence of the absorption constant α (penetration depth $1/\alpha$) for several semiconductor materials

Figure 5.4 shows the wavelength dependence of the absorption constant α and of the penetration depth $1/\alpha$ for Ge, Si, GaAs and $(In_{0.7}Ga_{0.3})(As_{0.64}P_{0.36})$; these materials are commonly used to make photodiodes for lightwave systems. Some special values for InGaAs are: $\alpha = 0.68\,\mu\text{m}^{-1}$ at $\lambda = 1.55\,\mu$m, $\alpha = 1.16\,\mu\text{m}^{-1}$ at $\lambda = 1.36\,\mu$m, $\alpha = 2.15\,\mu\text{m}^{-1}$ at $\lambda = 1.06\,\mu$m.

5.3 Time and frequency response

The response time of a pin-photodiode is determined by the speed with which it responds to variations of the incident optical power. The absorption layer of a pin-photodiode is displayed in Fig. 5.5. $P_e(t)$ is the incident external light power. The quantum efficiency η was defined in Eq. (5.14). The internal light power dependence $P_i(x,t)$ in the i-region is given by Eq. (5.13).

If we set formally (irrespective of the physical units) $P_e(t) = \delta(t)$, then the short-circuit "current" $i(t)$ computed from Eq. (5.9), (5.12) is denoted as impulse response $h_P(t;\text{pin})$ (unit A / Ws; the subscript P relates the impulse response to an impulse of the optical power). We substitute the generation term eFg from Eq. (5.16) using $P_e(t) = \delta(t)$ into Eq. (5.5), and integrate over the small interval $-\Delta t \le t \le \Delta t$. The currents before the power impulse are zero, and

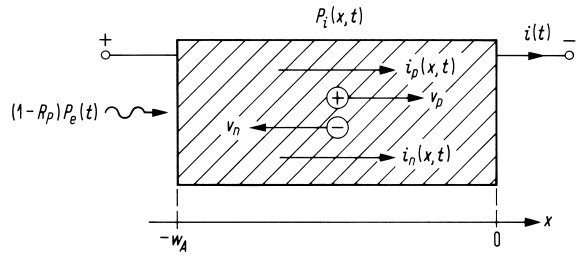


Fig. 5.5. Absorption layer of a pin-photodiode. P_e incicent external light power, R_P power reflection coefficient, $i(t)$ external short-circuit current, P_i internal optical power; i_p, i_n convection currents of electrons and holes; v_p, v_n saturation drift velocities, w_A length of absorption region

$\int_{-\Delta t}^{+\Delta t} \frac{\partial i_p}{\partial x}\,\mathrm{d}t = \frac{\partial}{\partial x}\int_{-\Delta t}^{+\Delta t} i_p\,\mathrm{d}t \to 0$ for $\Delta t \to 0$ because $i_p(x,t)$ has no singularity,

$$\frac{1}{v_p}\int_{-\Delta t}^{+\Delta t}\frac{\partial i_p}{\partial t}\,\mathrm{d}t + \int_{-\Delta t}^{+\Delta t}\frac{\partial i_p}{\partial x}\,\mathrm{d}t = \frac{\eta e}{h f_S}\,\frac{\alpha\,\mathrm{e}^{-\alpha(x+w_A)}}{1-\mathrm{e}^{-\alpha w_A}}\int_{-\Delta t}^{+\Delta t}\delta(t)\,\mathrm{d}t \quad \text{for } \Delta t \to 0\,, \tag{5.19}$$

$$\frac{1}{v_p}\Big(i_p(x,+0) - \underbrace{i_p(x,-0)}_{=0}\Big) + \frac{\partial}{\partial x}\underbrace{\int_{-\Delta t}^{+\Delta t} i_p(x,t)\,\mathrm{d}t}_{=0 \text{ for } \Delta t \to 0} = \frac{\eta e}{h f_S}\,\frac{\alpha\,\mathrm{e}^{-\alpha(x+w_A)}}{1-\mathrm{e}^{-\alpha w_A}}$$

For $\Delta t \to 0$ we find as an initial condition the convection "currents" at $t = +0$,

$$\frac{1}{v_p}\,i_p(x,+0) = \frac{1}{v_n}\,i_n(x,+0) = \frac{\eta e}{h f_S}\,\frac{\alpha\,\mathrm{e}^{-\alpha(x+w_A)}}{1-\mathrm{e}^{-\alpha w_A}}\,. \tag{5.20}$$

For the δ-excitation $g(x, t > 0) = 0$ is valid. The homogeneous differential equations (5.5) are solved by arbitrary functions $i_p(x,t) = i_p(x - v_p t)$, $i_n(x,t) = i_n(x + v_n t)$ which fulfill the initial conditions Eq. (5.20). These initial carrier distributions drift to the right and to the left with the saturation velocities v_p and v_n, respectively. With the Heaviside function $H(z)$ and the carrier transit times τ_n, τ_p defined by $w_A = v_p\tau_p = v_n\tau_n$ we calculate

$$\left.\begin{matrix} i_p(x,t) \\ i_n(x,t) \end{matrix}\right\} = \frac{\eta e}{h f_S}\,\frac{\alpha\,\mathrm{e}^{-\alpha w_A}}{1-\mathrm{e}^{-\alpha w_A}}\left\{\begin{matrix} v_p\,\mathrm{e}^{-\alpha(x-v_p t)} \times \\ v_n\,\mathrm{e}^{-\alpha(x+v_n t)} \times \end{matrix}\right. \\ \times\big[H(t) - H(t-\tau_p)\big]\big[H(x - v_p t + w_A) - H(x)\big], \\ \times\big[H(t) - H(t-\tau_n)\big]\big[H(x + w_A) - H(x + v_n t)\big]. \tag{5.21}$$

Substituting these "currents" into Eq. (5.9), $i(t) = \frac{1}{w_A}\int_{-w_A}^{0}\big[\,i_n(x,t) + i_p(x,t)\,\big]\,\mathrm{d}x$, we find for the impulse response

$$h_P(t;\mathrm{pin}) = \frac{\eta e}{h f_S}\,\frac{1}{1-\mathrm{e}^{-\alpha w_A}}\left\{\frac{1-\mathrm{e}^{-\alpha(w_A - v_p t)}}{\tau_p}\big[H(t) - H(t-\tau_p)\big] + \right. \\ \left. + \frac{\mathrm{e}^{-\alpha v_n t} - \mathrm{e}^{-\alpha w_A}}{\tau_n}\big[H(t) - H(t-\tau_n)\big]\right\}, \tag{5.22}$$

$$\int_{-\infty}^{+\infty} h_P(t;\mathrm{pin})\,\mathrm{d}t = \frac{\eta e}{h f_S}\,.$$

The total "charge" transported in an external circuit, originating from an optical "power" $P_e(t) = \delta(t)$, is $\eta e/(h f_S)$. The corresponding absorbed light "energy" is η. Thus, $\eta/(h f_S)$ represents the absorbed photon "number". Each absorbed photon generates an electron-hole pair leading to the transport of an elementary charge e through the external circuit.

The Fourier transform $H_P(f;\mathrm{pin})$ of the impulse response $h_P(t;\mathrm{pin})$ is the transfer function of the pin-photodiode, see Eq. (2.2),

$$H_P(f;\mathrm{pin}) = \frac{\eta e}{h f_S}\frac{1}{1-\mathrm{e}^{-\alpha w_A}}\left[\frac{1-\mathrm{e}^{-\mathrm{j}\,\omega\tau_p}}{\mathrm{j}\,\omega\tau_p} + \frac{\mathrm{e}^{-\alpha w_A}-\mathrm{e}^{-\mathrm{j}\,\omega\tau_p}}{\alpha w_A - \mathrm{j}\,\omega\tau_p} + \right.$$
$$\left. + \frac{1-\mathrm{e}^{-\alpha w_A}\,\mathrm{e}^{-\mathrm{j}\,\omega\tau_n}}{\alpha w_A + \mathrm{j}\,\omega\tau_n} - \mathrm{e}^{-\alpha w_A}\,\frac{1-\mathrm{e}^{-\mathrm{j}\,\omega\tau_n}}{\mathrm{j}\,\omega\tau_n}\right]. \tag{5.23}$$

Thus, the external short-circuit current $i(t;\mathrm{pin})$ and its frequency response $I(f;\mathrm{pin})$ for an arbitrary time dependent illumination $P_e(t)$ are

$$\begin{aligned} i(t;\mathrm{pin}) &= \int_{-\infty}^{+\infty} P_e(t')\,h_P(t-t';\mathrm{pin})\,\mathrm{d}t', \\ I(f;\mathrm{pin}) &= \breve{P}_e(f) H_P(f;\mathrm{pin}) \end{aligned} \tag{5.24}$$

$\breve{P}_e(f)$ is the Fourier transform of $P_e(t)$. According to Fig. 5.3 and Eq. (5.18) the transfer function $H_S(f)$ from the photodiode to the load resistance is

$$I_a(f) = \breve{P}_e(f) H_P(f;\mathrm{pin}) H_S(f). \tag{5.25}$$

Strong absorption In the limiting case of strong absorption $\alpha w_A \to \infty$ all the light power is absorbed inside an infinitely thin layer at $x = -w_A$. The quantum efficiency η (Eq. (5.14)) simplifies to

$$\eta = 1 - R_P, \qquad (\alpha w_A \to \infty) \tag{5.26}$$

With an antireflection coating $\eta \to 1$ is achievable. The initial convection currents Eq. (5.20) valid for $-w_A \le x \le 0$ are

$$\frac{1}{v_p}\,i_p(x,+0) = \frac{1}{v_n}\,i_n(x,+0) = \frac{\eta e}{h f_S}\frac{\alpha\,\mathrm{e}^{-\alpha(x+w_A)}}{1-\mathrm{e}^{-\alpha w_A}}\left[H(x+w_A) - H(x)\right]. \tag{5.27}$$

With $\lim_{\alpha\to\infty}\left[\alpha\,\mathrm{e}^{-\alpha(x-v_p t+w_A)}\,H(x-v_p t+w_A)\right] = \delta(x-v_p t+w_A)$ we calculate[7] the convection "current" response $i_p(x,t)$ from Eq. (5.21),

$$\begin{aligned} i_p(x,t) &= \frac{\eta e}{h f_S}\frac{\alpha\,\mathrm{e}^{-\alpha w_A}}{1-\mathrm{e}^{-\alpha w_A}}\,v_p\,\mathrm{e}^{-\alpha(x-v_p t)}\Big[H(t)-H(t-\tau_p)\Big]\Big[H(x-v_p t+w_A)-H(x)\Big] \\ &= \frac{\eta e}{h f_S}\frac{v_p}{1-\mathrm{e}^{-\alpha w_A}}\Big[H(t)-H(t-\tau_p)\Big]\,\mathrm{e}^{-\alpha(x-v_p t+w_A)}\Big[H(x-v_p t+w_A)-H(x)\Big], \end{aligned}$$

[7] The meaning of $\alpha\,\mathrm{e}^{-\alpha(x)}\,H(x)$ may be seen through an integration by parts, $\int uv'\,\mathrm{d}x = uv - \int u'v\,\mathrm{d}x$,

$$\begin{aligned} \lim_{\alpha\to\infty}\alpha\int_{-\infty}^{+\infty}\mathrm{e}^{-\alpha x}\,H(x)\,\Phi(x)\,\mathrm{d}x &= \lim_{\alpha\to\infty}\alpha\int_{0}^{+\infty}\Phi(x)\,\mathrm{e}^{-\alpha x}\,\mathrm{d}x \\ &= \lim_{\alpha\to\infty}\alpha\left[\Phi(x)\frac{-1}{\alpha}\,\mathrm{e}^{-\alpha x}\Big|_0^{+\infty} - \int_0^{+\infty}\Phi'(x)\frac{-1}{\alpha}\,\mathrm{e}^{-\alpha x}\,\mathrm{d}x\right] \\ &= \lim_{\alpha\to\infty}\alpha\left[\Phi(0)\frac{1}{\alpha} - \Phi'(0)\left(\frac{-1}{\alpha}\right)^2 + \int_0^{+\infty}\Phi''(x)\left(\frac{-1}{\alpha}\right)^2\mathrm{e}^{-\alpha x}\,\mathrm{d}x\right] \\ &= \Phi(0) = \int_{-\infty}^{+\infty}\delta(x)\,\Phi(x)\,\mathrm{d}x \quad \rightsquigarrow \quad \delta(x) = \lim_{\alpha\to\infty}\left[\alpha\,\mathrm{e}^{-\alpha x}\,H(x)\right]. \end{aligned} \tag{5.28}$$

Thus, the initial conditions for the convection currents (i. e., the carrier concentrations) in the region $-w_A \le x \le 0$ are in proportion to $\delta(x+w_A)$.

which finally leads to

$$
\begin{aligned}
i_p(x,t) &= \frac{\eta e}{h f_S}\,\frac{v_p}{1-\mathrm{e}^{-\alpha w_A}}\Big[\,H(t)-H(t-\tau_p)\,\Big] \qquad (5.29)\\
&\quad\times\Big[\underbrace{\alpha\,\mathrm{e}^{-\alpha(x-v_pt+w_A)}\;H(x-v_pt+w_A)}_{=\delta(x-v_pt+w_A)\ \text{for}\ \alpha\to\infty}-\underbrace{\alpha\,\mathrm{e}^{-\alpha x}\;H(x)\,\mathrm{e}^{\alpha(v_pt-w_A)}}_{=0\ \text{for}\ \alpha\to\infty}\Big]\\
&=\{\alpha\to\infty\}=\frac{\eta e}{h f_S}\,v_p\Big[\,H(t)-H(t-\tau_p)\,\Big]\,\delta(x-v_pt+w_A).
\end{aligned}
$$

An analogous result may be found for $i_n(x,t)$. With Eq. (5.9) the total external current is $i(t)=\frac{1}{w_A}\int_{-w_A}^{0}\big[\,i_n(x,t)+i_p(x,t)\,\big]\,\mathrm{d}x$. For the hole current, $\int_{-w_A}^{0}\delta(x-v_pt+w_A)\,\mathrm{d}x=\int_{-v_pt}^{-v_pt+w_A}\delta(\xi)\times\mathrm{d}\xi=\{t=0\}=\int_{-0}^{+w_A}\delta(\xi)\,\mathrm{d}\xi=1$ holds. The electron current contribution disappears for $t>0$. From Eqs. (5.22), (5.23) we calculate the external current impulse and frequency responses

$$
\left.
\begin{aligned}
h_P(t;\mathrm{pin}) &= \begin{cases}\dfrac{\eta e}{h f_S}\left(\dfrac{1}{\tau_p}+\dfrac{1}{\tau_n}\right) & (t=0),\\[2ex] \dfrac{\eta e}{h f_S\tau_p}\,[\,H(t)-H(t-\tau_p)\,] & (t>0),\end{cases}\\
H_P(f;\mathrm{pin}) &= \frac{\eta e}{h f_S}\;\mathrm{e}^{-\mathrm{j}\,\omega\tau_p/2}\;\frac{\sin(\omega\tau_p/2)}{\omega\tau_p/2}\,,
\end{aligned}
\right\}\quad(\alpha w_A\to\infty). \qquad (5.30)
$$

Weak absorption For the limiting case of weak absorption $\alpha w_A\to 0$ the responses are

$$
\left.
\begin{aligned}
h_P(t;\mathrm{pin}) &= \frac{\eta e}{h f_S}\left\{\frac{1-t/\tau_p}{\tau_p}\,[\,H(t)-H(t-\tau_p)\,]\;+\right.\\
&\qquad\left.+\;\frac{1-t/\tau_n}{\tau_n}\,[\,H(t)-H(t-\tau_n)\,]\right\},\\
H_P(f;\mathrm{pin}) &= \frac{\eta e}{\mathrm{j}\,\omega\tau_p h f_S}\left[1-\mathrm{e}^{-\mathrm{j}\,\omega\tau_p/2}\;\frac{\sin(\omega\tau_p/2)}{\omega\tau_p/2}\right]+\\
&\qquad+\frac{\eta e}{\mathrm{j}\,\omega\tau_n h f_S}\left[1-\mathrm{e}^{-\mathrm{j}\,\omega\tau_n/2}\;\frac{\sin(\omega\tau_n/2)}{\omega\tau_n/2}\right].
\end{aligned}
\right\}\quad(\alpha w_A\to 0). \qquad (5.31)
$$

Figure 5.6 shows the external current impulse responses of a pin-photodiode for the limiting cases of strong ($\alpha w_A\to\infty$) and weak absorption ($\alpha w_A\to 0$) assuming $v_p=v_n=w_A/\tau$.

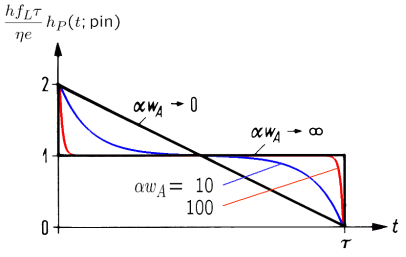


Fig. 5.6. Transit-time limited impulse responses of a pin-photodiode for $v_p=v_n$ ($\tau_p=\tau_n=\tau$) and the cases of strong absorption ($\alpha w_A\to\infty$, in practice $\alpha w_A\geq 1000$) and weak absorption ($\alpha w_A\to 0$, in practice $\alpha w_A\leq 1$). Intermediate-absorption graphs for $\alpha w_A=100$ and $\alpha w_A=10$

5.4 Cutoff frequency, quantum efficiency and responsivity

According to Eq. (5.18) on Page 109 the transfer function $H_S(f)$ from the photodiode to the load resistance R_a has the same structure as the small-signal transfer function of the current-modulated laser diode, Eq. (3.105), (3.106) on Page (87) and Fig. 3.22. We are only interested

in the short-circuit current $I(f;\mathrm{pin})$ originating from the alternating part of the sinusoidal light power modulation,

$$P_e(t) = P_0 + P_1\cos(\omega t) \quad (P_0 \geq P_1 = \mathrm{const}_t) \tag{5.32}$$

Again, the limiting cases of strong (Eq. (5.30)) and weak absorption (Eq. (5.31), $\tau_n = \tau_p = \tau$) are regarded,

$$\begin{aligned} \frac{I(f;\mathrm{pin})}{I(0;\mathrm{pin})} &= \mathrm{e}^{-\mathrm{j}\,\omega\tau_p/2}\,\frac{\sin(\omega\tau_p/2)}{\omega\tau_p/2}\,, && (\alpha w_A \to \infty), \\ \frac{I(f;\mathrm{pin})}{I(0;\mathrm{pin})} &= \frac{1}{\mathrm{j}\,\omega\tau/2}\left[1 - \mathrm{e}^{-\mathrm{j}\,\omega\tau/2}\,\frac{\sin(\omega\tau/2)}{\omega\tau/2}\right], && (\alpha w_A \to 0). \end{aligned} \tag{5.33}$$

Transit-time cutoff frequency The transit-time limited 3-dB cutoff frequency is computed from

$$\left|\frac{I(f_{3\,\mathrm{dB}};\mathrm{pin})}{I(0;\mathrm{pin})}\right| = \frac{1}{\sqrt{2}} \tag{5.34}$$

resulting in

$$f_{3\,\mathrm{dB}} = \begin{cases} 0.44/\tau_p & (\alpha w_A \to \infty), \\ 0.55/\tau & (\alpha w_A \to 0,\ \tau_n = \tau_p = \tau). \end{cases} \tag{5.35}$$

Quantum efficiency For weak absorption $\alpha w_A \to 0$ we see from Eq. (5.14) that the quantum efficiency is $\eta = (1-R_P)\alpha w_A$ (compare $\eta = 1 - R_P$ for $\alpha w_A \to \infty$, Eq. (5.26)), thus the product of quantum efficiency and 3-dB cutoff frequency is

$$\begin{aligned} \eta &= (1-R_P)\,(1-\mathrm{e}^{-\alpha w_A}) && (\text{Eq. (5.14)}), \\ \eta f_{3\,\mathrm{dB}} &= 0.55\,(1-R_P)\,\alpha v\,, && (\alpha w_A \to 0). \end{aligned} \tag{5.36}$$

Assuming $R_P = 0$ the efficiency-bandwidth products depends only on the material parameters α (absorption constant) and v (saturation velocity of carriers). With the InGaAs data of Sect. 5.2 at $\lambda = 1.55\,\mu$m ($\alpha = 0.68\,\mu\mathrm{m}^{-1}$, $v = (v_n + v_p)/2 = 56.5\,\mu\mathrm{m}\,/\,\mathrm{ns}$) an efficiency-bandwidth product $\eta f_{3\,\mathrm{dB}} = 21\,\mathrm{GHz}$ results, for $\lambda = 1.36\,\mu$m ($\alpha = 1.16\,\mu\mathrm{m}^{-1}$, $v = 56.5\,\mu\mathrm{m}\,/\,\mathrm{ns}$) it is $\eta f_{3\,\mathrm{dB}} = 36\,\mathrm{GHz}$.

Edge-coupling Better $\eta f_{3\,\mathrm{dB}}$-values may be achieved if the absorption length for photons along a waveguide and the transport distance of charge carriers is decoupled by an edge-coupled photodiode, Fig. 5.7. The active zone is an InGaAs absorbing layer embedded between n-InP and p-InP heterolayers. The structure is operated with reverse bias. The light is coupled into the vertical InP/InGaAs/InP slab waveguide along the z-direction. An InGaAs layer with a height of $w_A = 0.2\,\mu$m results in a field confinement factor of $\Gamma = 0.4$. If InGaAs has an absorption coefficient α, the waveguide shows an effective absorption constant $\Gamma\alpha$ for the fundamental mode propagating into the z-direction. The pin-photodiode has a z-extension of L. The input coupling efficiency from an external source into the waveguide is η_{coupl}. The quantum efficiency of the photodiode is

$$\eta = \eta_{\mathrm{coupl}}\left(1 - \mathrm{e}^{-\alpha\Gamma L}\right). \tag{5.37}$$

With the above values for α, Γ and waveguide lengths $L > 10\,\mu$m the quantum efficiency is $\eta \approx \eta_{\mathrm{coupl}}$. Coupling efficiencies of $\eta_{\mathrm{coupl}} = 0.8$ are feasible. The cutoff frequency for small absorption layers w_A is given by the second line of Eq. (5.35), $f_{3\,\mathrm{dB}} = 0.55/\tau$. Independently of $f_{3\,\mathrm{dB}}$ and of the operating wavelength the quantum efficiency is near $\eta \approx 0.8$. For $w_A = 0.2\,\mu$m the transit-time limited cutoff frequency would be $f_{3\,\mathrm{dB}} = 124\,\mathrm{GHz}$. The bandwidth of such waveguide photodiodes is limited by the $R_a C_{\mathrm{sp}}$ time constant in Fig. 5.3 which can be decreases by controlling the waveguide cross-section area.

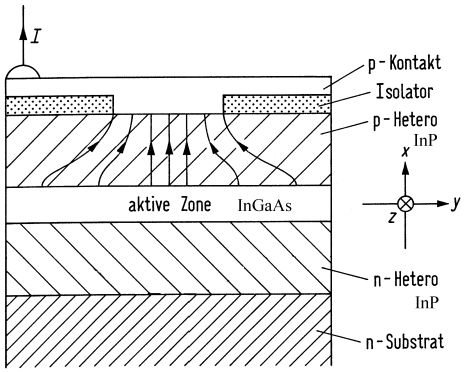


Fig. 5.7. Pin-diode with edge-coupling of light, w_A is the height of the active zone. x-axis: direction of current flow; z-axis: direction of light propagation (Aktive Zone = active zone, p-Kontakt = p-contact, Isolator = insulator, n-substrat = n-substrate).

Diffusion cutoff frequency When absorption inside the diffusion zones must not be neglected (see Fig. 5.1 for a schematic of the pin-photodiode) the cutoff frequency is limited by relatively slow diffusion processes. Carriers having a diffusion constant D move by a mean distance $\Delta x = \sqrt{D\tau}$ inside a time interval τ (random walk). For the case of Fig. 5.1 $\Delta x = x_n - x_{nD} = L_n = \sqrt{D_n \tau_n}$ would be given by the diffusion length L_n, and $\tau = \tau_n$ would equal the minority lifetime τ_n. If the n-doped layer is short (the contact region has a length of zero, the diffusion zone has a length $d_{\rm dif} < L_n$) we have $\Delta x = d_{\rm dif}$. The diffusion-limited cutoff frequency is defined by

$$f_{\rm dif} = \frac{1}{\tau} = \frac{D}{(\Delta x)^2} = \frac{\mu k T_0}{e(\Delta x)^2}\,. \tag{5.38}$$

Assuming a short semiconductor with $\Delta x = 0.2\,\mu$m and $\mu_n = 8\,500\,{\rm cm}^2/({\rm V\,s})$ ($1\,500\,{\rm cm}^2/({\rm V\,s})$) for GaAs (Si) diffusion-limited cutoff frequencies are computed to be $f_{\rm dif} = 540\,$GHz (94 GHz). For this example the actual cutoff frequency will not be diffusion-limited. However, for indirect semiconductors the minority carrier lifetime increases up to $\tau = 1\,$ms. Therefore the cutoff frequency may be as small as $f_{\rm dif} = 1\,$kHz.

Responsivity According to Eq. (5.17) the photocurrent i is directly proportional to the externally available optical power P_e. The proportionality factor S is called the sensitivity or responsivity of the photodiode ($hf_S = (\lambda_S/\,\mu{\rm m})/1.24$, $f_S = c/\lambda_S$),

$$i = SP_e\,, \qquad S = \frac{\eta e}{hf_S}, \qquad \frac{S}{{\rm A\,/\,W}} = \eta\,\frac{\lambda_S/\,\mu{\rm m}}{1.24} = 0.806\,\eta\,\frac{\lambda_S}{\mu{\rm m}}\,. \tag{5.39}$$

The sensitivity of a photodiode increases with the wavelength λ_S simply because more photons are present for the same optical power. Such a linear dependence on λ_S is not expected to continue forever, since eventually the photon energy hf_S becomes smaller than the bandgap energy W_G. The quantum efficiency η then drops to zero. The dependence of η on λ_S enters through the absorption coefficient α, Eq. (5.14) and Fig. 5.4.[8]

5.5 Device structures

The depletion-layer width w and the depletion-layer capacitance $C_{\rm sp}$ for an abrupt pn-junction may be computed if the impurity concentrations n_A, n_D, the intrinsic density n_i, the thermal voltage $U_T = kT/e$, the built-in voltage U_D (*German* Diffusionsspannung), and the reverse voltage

[8] See Sect. 4.1 Page 139 in reference Footnote 9 on Page 4

U across the depletion region are known. $U > 0$ means that the positive battery contact was connected to the n-doped semiconductor,

$$w = \sqrt{\frac{2\epsilon_0\epsilon_r(U_D + U)}{e}\left(\frac{1}{n_A} + \frac{1}{n_D}\right)}, \qquad C_{\mathrm{sp}} = \frac{\epsilon_0\epsilon_r F}{w}, \qquad U_D = U_T \ln\frac{n_A n_D}{n_i^2}. \tag{5.40}$$

In the short-wavelength region most photodiodes are made from Si. The light penetration depth is $1/\alpha = 15\,\mu$m at $\lambda = 0.85\,\mu$m. For the pin-photodiode of Fig. 5.8(a) the light penetrates an anti-

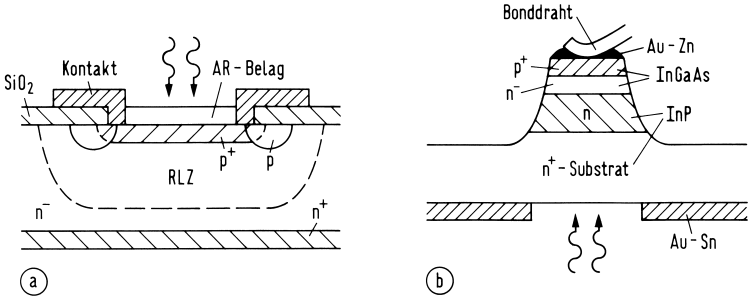


Fig. 5.8. Pin-photodiodes. (a) Planar Si photodiode (b) InGaAs/InP photodiode with mesa structure and illumination through the InP substrate (AR-Belag = anti-reflection coating, RLZ Raumladungszone = depletion region, Kontakt = contact. Bonddraht = bond wire).

reflection coating (SiO_2, Si_3N_4) and a thin p-doped layer ($< 1\,\mu$m) without significant reflection or absorption. The absorption region is made of n^--doped material (e. g., $n_D = 1.3\times10^{14}\,\mathrm{cm}^{-3}$). The propagation direction of the light as discussed on Page 106 coincides with the drift of the faster carriers (electrons in Si). A p-doped guard ring prevents a breakdown at high reverse voltages. The width of the depletion layer for $U = 10\,\mathrm{V}$, $n_D = 1.3\times10^{14}\,\mathrm{cm}^{-3} \ll n_A$, $\epsilon_0 = 8.85\times10^{-12}\,\mathrm{F\,m}^{-1}$, $\epsilon_r = 11.7$ is about $w = 10\,\mu$m. With a mean saturation drift velocity $v = (v_n + v_p)/2 = 64\,\mu\mathrm{m}\,/\,\mathrm{ns}$ a transit-time cutoff frequency near $f_{3\,\mathrm{dB}} = 3\,\mathrm{GHz}$ results from Eq. (5.35). With an active photodiode area of $F = (200\,\mu\mathrm{m})^2$ and a resistance of $R_S + R_a = 60\,\Omega$ (see Eq. (5.18)) the RC cutoff frequency is 6.4 GHz. Thus, the photodiode is transit-time limited. For $R_P = 0$ a quantum efficiency of $\eta \approx 0.5$ is to be expected, Eq. (5.14).

The performance of pin-photodiodes can be improved considerably by using a double-heterostructure design,[9] Fig. 5.8(b) and Fig. 5.9. Similar to the case of semiconductor lasers, the middle i-type layer is sandwiched between the p-type and n-type layers of a different semiconductor whose bandgap is chosen such that light is absorbed only in the middle i-layer. Since the bandgap energy of InP is $W_G = 1.35\,\mathrm{eV}$, see Table 3.2 on Page 56, the material is transparent for light whose wavelength exceeds $\lambda = 0.918\,\mu$m. By contrast, the bandgap of lattice-matched $(In_{0.53}Ga_{0.47})As$ is $W_G = 0.75\,\mathrm{eV}$ corresponding to a wavelength of $\lambda = 1.653\,\mu$m. The middle InGaAs layer thus absorbs strongly in the wavelength region $1.3\ldots1.6\,\mu$m. The diffusive component of the detector current is eliminated completely in such a heterostructure photodiode simply because photons are absorbed only inside the depletion region.

For the long-wavelength region, 3-dB cutoff frequencies near 100 GHz were measured. Figure 5.8(b) shows a pin-photodiode with mesa structure. Epitaxial layers of n-InP (buffer layer about $3\,\mu$m, $n_D = 5\times10^{16}\,\mathrm{cm}^{-3}$) and InGaAs ($1.2\,\mu$m, nominally undoped, $n_D = 3\times10^{14}\,\mathrm{cm}^{-3}$) are grown on a n^+-InP-substrate. By diffusion of Zn into the InGaAs layer a p^+n-junction is formed at a distance of $0.5\,\mu$m away from the surface. An etching process forms a mesa to reduce the capacitance. The illumination is through the substrate which is transparent for $\lambda > 0.92\,\mu$m. By reflection of the light at the p-contact the quantum efficiency is increased to $\eta \approx 0.5$.

[9] See Sect. 7.2.2 Page 260 in reference Footnote 3 on Page 47

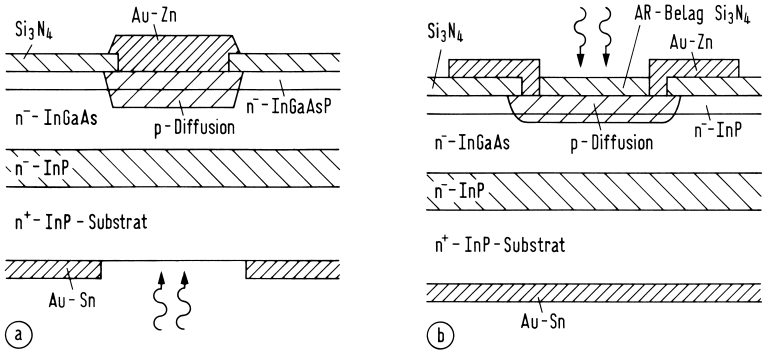


Fig. 5.9. Planar InGaAs/InP pin-photodiodes. (a) Illumination through the substrate (b) illumination from top, AR-Belag = anti-reflection coating

Figure 5.9 shows planar pin-photodiodes. In Fig. 5.9(a) the n-type ($n_D \approx 10^{15}\,\mathrm{cm}^{-3}$) InP, the InGaAs, and the (In,Ga)(As,P) ($W_G = 0.95\,\mathrm{eV} \mathrel{\widehat{=}} 1.3\,\mu\mathrm{m}$) layers are nominally undoped. The Zn p-diffusion reaches for about $1\,\mu$m into the InGaAs layer. The height of the InGaAs absorption layer is in the range $0.4 \ldots 5\,\mu$m, depending on the quantum efficiency design and on the tolerated depletion-layer capacitance. Measured efficiencies are $\eta = 0.3 \ldots 0.6$. Transit time 3-dB frequencies of 20 GHz and RC cutoff frequencies of 80 GHz were achieved.

With similar but top-illuminated planar pin-photodiodes (Fig. 5.9(b)) quantum efficiencies of $\eta = 0.7$ at $\lambda = 1.55\,\mu$m and $f_{3\,\mathrm{dB}} = 25\,\mathrm{GHz}$ are common. The cap layer ($0.5\,\mu$m) consists of n-InP ($n_D = 10^{16}\,\mathrm{cm}^{-3}$) which is moved to a very shallow depth ($0.1\,\mu$m) into the InGaAs layer for increasing the depletion layer over whole of the 2-μm InGaAs layer ($n_D < 5 \times 10^{15}\,\mathrm{cm}^{-3}$).

Instead of the n^+-InP substrate with a buffer layer of n^--InP the diode may be grown on a semi-insulation InP-substrate. In this case the buffer layer is replaced by a n^+-InP contact layer ($n_D = 10^{19}\,\mathrm{cm}^{-3}$) with a lateral contact.

Chapter 6

Noise

Optical receivers convert optical power P_e into electrical current i through a photodiode. The relation $i = SP_e$ in Eqs. (5.17), (5.39) assumes that the current resulting from such a conversion is noise free. However, this is not the case even for a perfect receiver. Two fundamental noise mechanisms, quantum (or shot) noise and thermal noise (Johnson or Nyquist noise) lead to current fluctuations even when the optical signal has a constant power P_e in the classical sense. The relation $i = SP_e$ still holds if we interpret i as the average current. However, unavoidable current fluctuations affect the receiver performance. The objective of this chapter is to briefly review some important noise mechanisms.[1]

6.1 Noise mechanisms

The classical light power $P_e(t)$ results from an average over a few optical cycles. Fluctuations in $P_e(t)$ are transferred to the photocurrent $i(t)$. The ideal classical signal exhibits a constant amplitude and phase, and no photocurrent fluctuations would be expected. However, quantum mechanics tells that this "ideal" signal (an ideal laser signal) consists of a sequence of independent photons which are Poisson distributed in time. Each photon generates an electron-hole pair with a quantum efficiency η Eqs. (5.14), (5.36). Thus, the short-circuit photocurrent consists of a stream of statistically independent elementary charges which are also Poisson distributed in time. This type of noise is known as shot[2] noise.

Therefore, the photodiode shot noise can be also denoted as quantum noise. The origin of this noise can be attributed neither to the source nor to the detector alone, because the noise shows only when photodiode and light source are interacting during the detection process.

Quantum noise, spontaneous emission noise and shot noise in electronic circuits are unavoidable. In addition to these noise mechanisms there are other noise sources which could be avoided or reduced (e. g., thermal noise of a resistor can be reduced by cooling the device).

6.2 Photocurrent noise

The photocurrent $i(t)$ (or i for short) fluctuates around its expectation $\overline{i(t)}$ (or $\bar{i}$ for short). The fluctuation is denoted as $\delta i(t)$ (or δi for short),

$$\delta i(t) = i(t) - \overline{i(t)}\,, \qquad \delta i = i - \bar{i}\,. \tag{6.1}$$

[1] See Sect. 4.4 Page 163 in reference Footnote 9 on Page 4

[2] shot (*pl.* same or shots): a small lead pellet used in quantity in a single charge or cartridge in a shotgun (The Concise Oxford Dictionary. Oxford: Oxford University Press 1990) — Onomatopoetically for the current noise to be heard in a loudspeaker resembling the falling of shot(s) onto a sheet metal.

The autocorrelation function $\vartheta_i(\tau)$ of i is related to the two-sided power spectrum $\Theta_i(f)$ via the Fourier transform,

$$\vartheta_i(\tau) = \overline{i(t+\tau)\, i(t)}\,, \qquad \Theta_i(f) = \int_{-\infty}^{+\infty} \vartheta_i(\tau)\, \mathrm{e}^{-\mathrm{j}\,2\pi f\tau}\, \mathrm{d}\tau\,. \tag{6.2}$$

Because $\vartheta_i(\tau)$ is real, the power spectrum has the property $\Theta_i(f) = \Theta_i^*(-f)$. From the definition, the symmetry relation $\vartheta_i(\tau) = \vartheta_i(-\tau)$ may be seen. As a consequence, the power spectrum is real, $\Theta_i(f) = \Theta_i^*(f)$, and $\vartheta_i(\tau) = \int_0^\infty 2\Theta_i(f)\,\mathrm{d}f$ defines a one-sided real spectrum $2\Theta_i(f)$. The noise variance is obtained by $\sigma_i^2 = \vartheta_i(\tau = 0)$,

$$\sigma_i^2 = \overline{(i-\bar{i}\,)^2} = \overline{\delta i^2} = \int_{-\infty}^{+\infty} \Theta_i(f)\,\mathrm{d}f = \int_0^\infty 2\Theta_i(f)\,\mathrm{d}f\,. \tag{6.3}$$

The spectral density of shot noise is constant and given by $\Theta_i(f) = e\bar{i}$. Usually, the differential fluctuations inside a differential bandwidth $\mathrm{d}f$ centred at the frequency f are of interest,

$$\mathrm{d}\Big(\overline{\delta i^2}\Big) = 2\Theta_i(f)\,\mathrm{d}f = 2e\bar{i}\,\mathrm{d}f, \qquad \overline{|i_{RD}|^2} = 2e\bar{i}B. \tag{6.4}$$

The complex phasor i_{RD} (effective or root mean square (RMS) value $|i_{RD}|$; subscript R for noise, *German* „Rauschen“) is defined to have the same power $\overline{|i_{RD}|^2}$ per frequency interval[3] $\mathrm{d}f$ as the actual noise process Eq. (6.4), and because the spectral power density is frequency-independent in the region of interest, the differential $\mathrm{d}f$ can be replaced by the receiver bandwidth B. Equation (6.4) expresses a property of the underlying Poisson statistics for the photons: The probability $w_N(N_P)$ for measuring N_P photons, if the expectation is $\overline{N_P} = N_e$, and the associated second central moment of the process are

$$w_N(N_P) = \frac{\overline{N_P}^{N_P}}{N_P!}\, \mathrm{e}^{-\overline{N_P}}, \quad \overline{\delta N_P^2} = \overline{(N_P - N_e)^2} = \overline{N_P}\,, \qquad \overline{N_P} = N_e\,. \tag{6.6}$$

The expected current $\bar{i} = S\overline{P_e}$ is computed from Eq. (5.39). Classical additional fluctuations from the laser source, which has a total output power P_a (Eq. (3.94) on Page 3.94), are described by the relative intensity noise (RIN),

$$\mathrm{RIN} = \int_0^\infty \mathrm{RIN}(f)\,\mathrm{d}f = \frac{\overline{\delta P_a^2}}{\overline{P_a}^2}\,, \qquad \mathrm{d}\left(\overline{\delta P_a^2}\right) = \overline{P_a}^2\,\mathrm{RIN}(f)\,\mathrm{d}f\,. \tag{6.7}$$

The total differential photocurrent noise fluctuation including the (uncorrelated) received RIN (mean power $\overline{P_a} \sim P_e \sim \bar{i}$) or the integrated noise current i_{RD} in a bandwidth B assuming $\mathrm{RIN}(f) = \mathrm{RIN}(f_e)$ are given by

$$\mathrm{d}\Big(\overline{\delta i^2}\Big) = \underbrace{2e\bar{i}\,\mathrm{d}f}_{\substack{\text{shot resp.}\\ \text{quantum}\\ \text{noise}}} + \underbrace{\bar{i}^2\,\mathrm{RIN}(f)\,\mathrm{d}f}_{\text{classical noise}}, \qquad \overline{|i_{RD}|^2} = 2e\bar{i}B + \bar{i}^2\,\mathrm{RIN}(f_e)B. \tag{6.8}$$

The spectral shot noise power density for $\bar{i} = 1\,\mathrm{mA}$ measured at a resistor of $R = 50\,\Omega$ amounts to $2e\bar{i}R = -168\,\mathrm{dBm\,/\,Hz}$.

[3] In Eq. (6.4) the spectral power density $2e\bar{i}$ is frequency-independent. An integration over the frequency range $0 \le f \le B$ results in the total current variance

$$\overline{\delta i^2} = \int_0^B \mathrm{d}\Big(\overline{\delta i^2}\Big) = \int_0^B 2e\bar{i}\,\mathrm{d}f = 2e\bar{i}B = \overline{|i_{RD}|_B^2}\,, \qquad \overline{|i_{RD}|_B^2} = 2e\bar{i}B = \overline{|i_{RD}|^2}\Big|_{\mathrm{d}f\to B}\,. \tag{6.5}$$

However, to avoid an over-complicated notation, we drop the subscript B in the total variance $\overline{|i_{RD}|_B^2}$ and replace $\mathrm{d}f \to B$ wherever it is appropriate, i. e., when the spectral noise current power density $\mathrm{d}\big(\overline{\delta i^2}\big)\big/\,\mathrm{d}f$ does not depend on frequency.

An avalanche photodiode (APD) amplifies the primary photocurrent $i_{\rm pr} = SP_e$ by a current multiplication factor M, the statistical average of which is denoted by M_0. The avalanche multiplication process contributes additional noise, which is described by an excess noise factor F_M,

$$F_M = \frac{\overline{M^2}}{\overline{M}^2} = \frac{\overline{M^2}}{M_0^2} = 1 + \frac{\overline{\delta M^2}}{M_0^2}. \tag{6.9}$$

It is common to approximate F_M by the function

$$F_M = M_0^x, \quad x > 0\,. \tag{6.10}$$

For the APD current $i = Mi_{\rm pr}$ and for the noise current $\mathrm{d}\big(\overline{\delta i^2}\big)$ in a differential bandwidth $\mathrm{d}f$ or for the integrated noise current i_{RD} in a bandwidth B assuming $\mathrm{RIN}(f) = \mathrm{RIN}(f_e)$ we find the relations

$$\begin{aligned} \bar{i} &= M_0\bar{i}_{\rm pr} = M_0\frac{\eta e}{hf_S}P_e, \\ \mathrm{d}\left(\overline{\delta i^2}\right) &= 2e\bar{i}_{\rm pr}M_0^2F_M\,\mathrm{d}f + (M_0\bar{i}_{\rm pr})^2\,\mathrm{RIN}(f)\,\mathrm{d}f, \quad \overline{|i_{RD}|^2} = 2e\bar{i}_{\rm pr}M_0^2F_MB + (M_0\bar{i}_{\rm pr})^2\,\mathrm{RIN}(f_e)B. \end{aligned} \tag{6.11}$$

For $M_0 = 1$, $F_M = 1$, Eq. (6.11) reduces to the case of the pin photodiode, Eq. (6.8) on Page 120.

6.3 Shot noise in semiconductor junctions

A similar relation as for the noise in photodiodes holds for ordinary semiconductor pn-junctions in diodes and transistors, too. The associated current fluctuation is also called shot noise[4], and no illumination with light is needed for this effect. Electrons and holes traverse the junction independently and at random times, depending on the thermal energy a carrier happens to have, and this leads to a Poisson distribution of the arrival times. A pn-junction which carries an

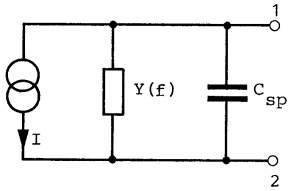


Fig. 6.1. Small-signal equivalent circuit of a pn-junction with shot noise RMS current I in a bandwidth B. Current fluctuation $\overline{|I|^2} = 2e\bar{i}B$, diffusion admittance $Y(f)$, junction capacitance $C_{\rm sp}$

average forward or reverse current $\bar{i}$ at a junction voltage U has a saturation current I_S, and a small-signal conductance G_0 for low frequencies $f \to 0$. With the thermal voltage U_T (*German* Temperaturspannung) we then find the well-known relation

$$i = I_S\left(\mathrm{e}^{U/U_T} - 1\right), \qquad G_0 = \frac{\partial i}{\partial U} = \frac{I_S}{U_T}\,\mathrm{e}^{U/U_T}, \qquad U_T = \frac{kT_0}{e}\,. \tag{6.12}$$

The diffusion admittance of the junction $Y(f) \sim \sqrt{1 + \mathrm{j}\,\omega\tau} \approx 1 + \frac{1}{2}\,(\mathrm{j}\,\omega\tau) - \frac{1}{8}\,(\mathrm{j}\,\omega\tau)^2$ (not all carriers cross the junction in a period $1/f$) is given by ($\omega = 2\pi f$, carrier recombination lifetime τ)

$$Y(f) = G_0\sqrt{1 + \mathrm{j}\,\omega\tau} = G(f) + \mathrm{j}\,B_Y(f), \qquad G(f) \approx G_0\left(1 + \frac{1}{8}\omega^2\tau^2\right). \tag{6.13}$$

[4]See Footnote 2 on Page 119

With Eq. (6.12) we write

$$\mathrm{d}\left(\overline{\delta i^2}\right) = 2e\bar{i}\,\mathrm{d}f + 4eI_S\,\mathrm{d}f + 4kT_0[G(f) - G_0]\,\mathrm{d}f = \begin{cases} 2e\bar{i}\,\mathrm{d}f & \text{for } f \to 0 \\ 4kT_0G(f)\,\mathrm{d}f & \text{for } \bar{i} = 0 \\ 4kT_0G_0\,\mathrm{d}f & \text{for } \bar{i} = 0 \text{ and } f \to 0 \end{cases} . \quad (6.14)$$

The equivalent circuit of a pn-junction with shot noise RMS current I and fluctuation $\overline{|I|^2} = 2e\bar{i}B$ in a bandwidth B, diffusion admittance $Y(f)$ and junction capacitance C_{sp} is depicted in Fig. 6.1.

6.4 Thermal noise

In any conductor at a finite temperature T, electrons move randomly. Random thermal motion of electrons in a resistor manifests as a fluctuating current even in the absence of an applied voltage. The load resistor R_a of a photodiode (Fig. 5.3) located in the front end of an optical receiver adds such fluctuations to the noise current generated by the photodiode. This additional noise component is referred to as thermal noise. It is also called Johnson noise or Nyquist noise after the two scientists who first studied it experimentally and theoretically.[5] The equivalent short-circuit

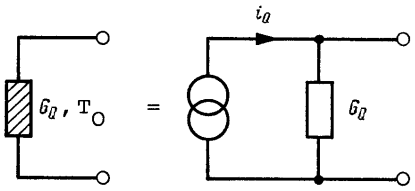


Fig. 6.2. Equivalent circuit of a conductance G_Q with thermal noise

noise current of the conductance G_Q and the available power at temperature T_0 in a bandwidth B are

$$\overline{|i_Q|^2} = 4kT_0G_QB, \qquad P_v = kT_0B, \qquad \text{room temperature } T_0 = 293\,\mathrm{K}\,. \quad (6.15)$$

To explain Eq. (6.15) on needs to know the number N_P of photons[6] with energy hf per mode in thermal equilibrium. In a bandwidth B these photons provide a power of N_PhfB. Electrical circuits are usually singlemoded, and the condition $hf \ll kT_0$ is easily fulfilled for $f < 1\,\mathrm{THz}$ ($hf < 4\,\mathrm{meV}$) at a room temperature of $T_0 = 293\,\mathrm{K}$ ($kT_0 = 25\,\mathrm{meV}$). Figure 6.2 displays the equivalent circuit of a resistance with thermal noise at temperature T_0. At a frequency $hf \ll kT_0$ the available noise power absorbed by an impedance-matched load in a bandwidth B is given by $P_v = kT_0B = \overline{|i_Q/2|^2}/G_Q$, the one-sided spectral density is $2\Theta_T(f) = kT_0$.

6.5 Electronic amplifier noise

If G_Q of Fig. 6.2 is the real part of the source admittance $Y_Q = G_Q + \mathrm{j}\,B_Q$, the additional noise of an electronic amplifier connected to this source may be described by an equivalent increase

[5] See Sect. 4.4.1 Page 164 in reference Footnote 9 on Page 4

[6] If a number $N = N_1 + N_2$ of microsystems (N_1 in state W_1, N_2 in state W_2) is in thermal equilibrium with an electromagnetic radiation mode, the average number of emissions must equal the average number of absorptions, $N_2w^{(\mathrm{eM})} = N_1w^{(\mathrm{aM})}$, i. e., with Eq. (3.36) on Page 68, $N_2(N_P + 1) = N_1N_P$. Because in thermal equilibrium we have $N_1/N_2 = \exp[(W_2 - W_1)/(kT_0)]$ from Eq. (3.11) on Page 53, we find Planck's formula for the expected number of photons in a mode with frequency f in thermal equilibrium (Bose-Einstein distribution),

$$N_P = \frac{1}{\exp\left(\frac{hf}{kT_0}\right) - 1} \approx \{hf \ll kT_0\} \approx \frac{kT_0}{hf}\,, \qquad hf = W_2 - W_1. \quad (6.16)$$

of the conductance temperature from the reference temperature T_0 to an elevated temperature FT_0, while the amplifier is regarded as noiseless. The noise figure $F = 1 + T_R/T_0$ specifies the equivalent fictitious temperature increase T_R which must be attributed to the source conductance G_Q. According to Fig. 7.1 on Page 125 in the next chapter, where the signal-to-noise ratios SNR are explicitly given, the equivalent short-circuit noise current for a noisy amplifier with noise figure F is

$$\overline{|i_R|^2} = 4kFT_0G_QB, \qquad F = \frac{\mathrm{SNR}_{\mathrm{input}}^{(\mathrm{available})}}{\mathrm{SNR}_{\mathrm{output}}} = \frac{\overline{|i_R|^2}}{\overline{|i_Q|^2}} = 1 + \frac{T_R}{T_0}\,. \tag{6.17}$$

6.6 Optical amplifier noise

Quantum noise (amplified spontaneous emission) is a limiting factor for optical amplifiers used in direct detection systems. A signal with optical power $P_S = \frac{1}{2}A^2$ is perturbed by narrowband noise with an optical bandwidth $B_O = 2B$ which equals two times the electrical (signal) bandwidth B.

Figure 6.3 shows schematically the one-sided optical power spectrum $w_s(f)$ in front of the optical amplifier, and the one-sided electrical power spectra $w_i(f)$ of the photocurrent i_S after optical amplification with a single-pass gain $\mathcal{G}_s$. The detecting photodiode has a sensitivity of $S = \eta e/(hf_S)$ (unit mA / mW; η quantum efficiency, e elementary charge, f_S central laser frequency). Quantum noise may be described by an equivalent classical Gaussian input noise source. Its available input noise power $P_{R\,\mathrm{eq}}$ has a spectral density w_O inside the optical bandwidth

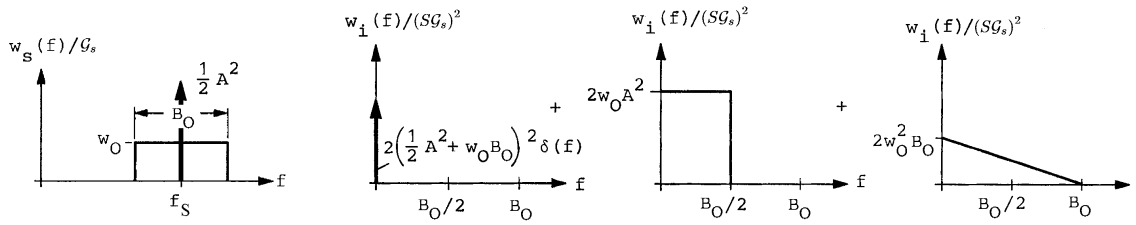


Fig. 6.3. One-sided power spectrum $w_s(f)$ of a sinusoidal optical signal with amplitude $A = \sqrt{2P_S}$ and frequency $f = f_S$ superimposed with narrowband noise from an ideal optical amplifier of bandwidth $B_O \ll f_S$ and optical power amplification $\mathcal{G}_s \gg 1$. Resulting signal and narrowband noise received with power detector having sensitivity S. Displayed are the one-sided partial power spectra $w_i(f)$ of the low-frequency components in the photocurrent $i = S\mathcal{G}_sP$ for $\mathcal{G}_s \gg 1$.

$B_O = 2B$ which is adjusted by appropriate optical filtering to correspond to double the electrical signal bandwidth B. Depending on the receiver type (homodyne, heterodyne, or, here, direct receiver) there are $\kappa = \frac{1}{2}$, 1, or 2 equivalent noise photons of energy hf_S in *one* transverse mode (observation time $\tau = 1/B$) and in *one* polarization,

$$P_{R\,\mathrm{eq}} = \frac{\kappa hf_S}{\tau} = \kappa hf_SB, \qquad w_O = hf_S, \qquad \kappa = \tfrac{1}{2}, 1, 2\,. \tag{6.18}$$

Because $P_{R\,\mathrm{eq}}$ represents minimum-uncertainty quantum fluctuations this power cannot be extracted from the system, so that after an optical amplifier having a single-pass gain $\mathcal{G}_s$ only the (unpolarized) optically amplified spontaneous noise power $P_R = 2\,(\mathcal{G}_s - 1)\,w_OB$ may be directly detected.

Noise figure The optimum electrical signal-to-noise ratio (SNR) γ_{dir} for direct detection is obtained if only the photodiode shot noise current i_{RD} (= quantum noise) of the direct current $i_S = SP_e$ is involved, the RMS noise of which is $\overline{|i_{RD}|^2} = 2ei_SB$; γ_1 denotes the ideal SNR for a detector quantum efficiency of $\eta = 1$,

$$\gamma_{\mathrm{dir}} = \frac{i_S^2}{\overline{|i_{RD}|^2}} = \frac{(SP_e)^2}{2eSP_eB} = \frac{\eta P_e}{2hf_SB}\,, \qquad \gamma_{\mathrm{dir}}(\eta = 1) = \gamma_1\,. \tag{6.19}$$

At the output of the optical amplifier the signal current is $i'_S = S\mathcal{G}_s P_e$, and if $B = B_O/2$ the dominant noise results from the mixing process between both the lower and upper noise sidebands and the coherent carrier.[7] The resulting RMS detector noise current is $\overline{|i'_{RD,2}|^2} = 4S\mathcal{G}_s P_e\, S\,(\mathcal{G}_s - 1)\, w'_O B$ where the (unpolarized) optically *a*mplified *s*pontaneous *e*mission (ASE) noise power is denoted by $P_{\mathrm{ASE}} = 2\,(\mathcal{G}_s - 1)\, 2w'_O B$; its spectral density w'_O as referred to the amplifier input is larger than $w_O \le w'_O$ because of non-complete inversion $n_{\mathrm{sp}} \neq 1$ of any real world optical amplifier, Eq. (3.45) on Page 71. Independently of the quantum efficiency η, this leads to an SNR of

$$\gamma_{\mathrm{OA\,dir}} = \frac{\overbrace{\left(S\mathcal{G}_s P_e\right)^2}^{i'^2_S}}{\underbrace{4\,(S\mathcal{G}_s P_e)\, S\,(\mathcal{G}_s - 1)\, w'_O B}_{\overline{|i'_{RD,2}|^2}}} = \frac{1}{2}\,\frac{P_e}{2w'_O B}\,\frac{1}{1 - 1/\mathcal{G}_s} = \left\{ \begin{array}{c} \text{ideal amplifier:} \\ \mathcal{G}_s \gg 1\,,\ w'_O = w_O \end{array} \right\} = \frac{\gamma_1}{2}\,. \quad (6.20)$$

According to the usual definition of the electrical noise figure F we find

$$\begin{aligned} F = \frac{\mathrm{SNR}^{(\mathrm{available})}_{\mathrm{input}}}{\mathrm{SNR}_{\mathrm{output}}} = \frac{\gamma_{\mathrm{dir}}(\eta = 1)}{\gamma_{\mathrm{OA\,dir}}} &= \frac{P_e}{2hf_S B}\,\frac{2\,(\mathcal{G}_s - 1)\, 2w'_O B}{\mathcal{G}_s P_e} = 2\,\frac{w'_O}{hf_S}\left(1 - \frac{1}{\mathcal{G}_s}\right) \\ &= \frac{P_{\mathrm{ASE}}/\,\mathcal{G}_s}{2hf_S B}\,, \qquad P_{\mathrm{ASE}} = 2\,(\mathcal{G}_s - 1)\, 2w'_O B\,. \end{aligned} \quad (6.21)$$

It has to be observed that the definition Eq. (6.21) only applies if a (nearly) sinusoidal optical carrier with frequency f_S is present during the noise measurement.[8,9] If the noise is measured in one polarization only, this value $P_{\mathrm{ASE},\,x}$ has to be doubled for determining $P_{\mathrm{ASE}} = 2P_{\mathrm{ASE},\,x}$. Naturally, the noise figure remains unaffected.

[7] W. Freude: Optische Empfänger und Fehlerwahrscheinlichkeit (OEF, Optical receivers and error probability). See Sect. 4.2 „Direkt- und Überlagerungsempfang im Vergleich (Direct and heterodyne reception compared)“. Compuscript for download at `http://www.ipq.kit.edu/studies/OEF/lecture_en.htm`

[8] Dausend, S.: Telcordia qualification: Gain and noise figure measurement techniques. Application note by Keopsys, 13.02.2006

[9] Deutsche Norm DIN EN 61290-3-1,2: Lichtwellenleiter-Verstärker – Prüfverfahren – Teil 3-1,2: Rauschzahlparameter. Prüfverfahren mit optischem und elektrischem Spektralanalysator. Mai 2004, Aug 2003

Chapter 7

Receivers and detection errors

The design of an optical receiver depends largely on the modulation format used by the transmitter. In particular, it depends on whether the signal is transmitted in an analog or digital format. For analog signals the signal-to-noise ratio quantifies the sensitivity, for digital systems it is the bit error ratio. Since most lightwave systems employ the digital format, we focus in this chapter on digital optical receivers.[1]

The front end of a receiver consists of a photodiode followed by an electronic amplifier which amplifies the weak electrical signal for further processing. The amplifier noise properties are vital for the overall reception quality.

7.1 Pin photodiode receiver limits

A schematic of an optical receiver with pin-photodiode is seen in Fig. 7.1. The noise sources i_{RD} (Eq. (6.4)) and i_R (Eq. (6.17)) come from different physical processes and are uncorrelated; therefore, the noise powers may be added. The average signal current $i_S \equiv \bar{i} = S\overline{P_e}$ in Eq. (6.1)–

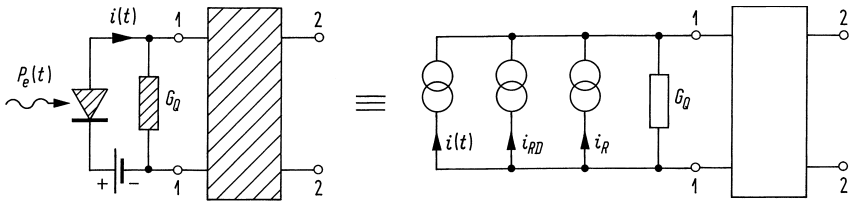


Fig. 7.1. Schematic of an optical receiver with pin-photodiode, source conductance G_Q and amplifier (noisy components are hatched). i_{RD} specifies the shot noise of the pin-photodiode, i_R represents the noise of the source conductance and of the amplifier

Eq. (6.8) is caused by the mean optical external signal power $\overline{P_e}$. The signal-to-noise ratio (SNR) γ for direct reception is defined as the ratio of the electrical signal power $P_S \sim i_S^2$ and the electrical noise power $P_R \sim \left(\overline{|i_{RD}|^2} + \overline{|i_R|^2}\right)$ in an electrical signal bandwidth B,

$$\gamma = \frac{P_S}{P_R} = \frac{i_S^2}{\overline{|i_{RD}|^2} + \overline{|i_R|^2}} = \frac{\eta\overline{P_e}}{2hf_SB} \, \frac{1}{1 + 4kFT_0G_Q/(2ei_S)} \, . \tag{7.1}$$

In most cases of practical interest, the amplifier noise $\overline{|i_R|^2} \gg \overline{|i_{RD}|^2}$ dominates receiver performance (*thermal noise limit*). However, if quantum noise becomes larger than amplifier noise, $\overline{|i_{RD}|^2} \gg \overline{|i_R|^2}$ or $4kFT_0G_Q/(2ei_S) \ll 1$, the maximum SNR may be achieved,

$$\gamma_{\max} = \frac{\eta\overline{P_e}}{2hf_SB} \qquad \text{(shot-noise limited, } \frac{4kFT_0G_Q}{2ei_S} \ll 1\text{).} \tag{7.2}$$

[1] See Sect. 4.3 Page 155 in reference Footnote 9 on Page 4

This is called the *shot noise limit*. In this case, the SNR increases linearly with $\overline{P_e}$ and depends only on the quantum efficiency η, the signal bandwidth B, and the photon energy hf_S. It is common to choose the bit period $T_t = 1/f_t$ according to the Nyquist condition $T_t = 1/(2B)$ (sampling theorem). The maximum shot-noise limited SNR is given by the mean number of absorbed photons ηN_e per bit period T_t (the absorbed energy in a bit period T_t is $\eta\overline{P_e}T_t = \eta\overline{P_e}/(2B)$),

$$\gamma_{\max} = \eta N_e \,. \tag{7.3}$$

In the shot noise limit, a SNR of 20 dB can be realized for $N_e \approx 100$ photons (assuming $\eta \approx 1$). By contrast, several 1 000 photons are required to obtain $\gamma \mathrel{\widehat{=}} 20\,\mathrm{dB}$ when thermal noise dominates the receiver. As a reference, for a 1.55-μm receiver operating at $f_t = 10\,\mathrm{Gbit/s}$, $N_e = 100$ when $\overline{P_e} \approx 130\,\mathrm{nW}$.

With a simple pin-photodiode receiver the shot noise limit cannot be reached because of the electronic amplifier noise. However, if an optical preamplifier boosts the signal power $\overline{P_e}$, the mean signal current i_S can become so large that condition (7.2) may be fulfilled. As a disadvantage, the optical amplifier adds some noise (minimum noise figure for direct reception is $F = 2$). Therefore, the shot-noise limited SNR is smaller than in Eq. (7.2), but it *can* be reached technically, and it becomes independent of the detector quantum efficiency η,

$$\gamma_{\mathrm{OA\ max}} = \frac{1}{2}\,\frac{\overline{P_e}}{2hf_SB} = \frac{\gamma_{\max}}{2\eta} = \frac{1}{2}N_e \,. \tag{7.4}$$

7.2 Detection errors

The data-recovery section of digital optical receivers consist of a decision circuit and a clock-recovery circuit. The clock-recovery unit isolates a spectral component $f_t = 1/T_t$ according to the clock frequency and helps to synchronize the decision process.[2] For the case of a binary modulation

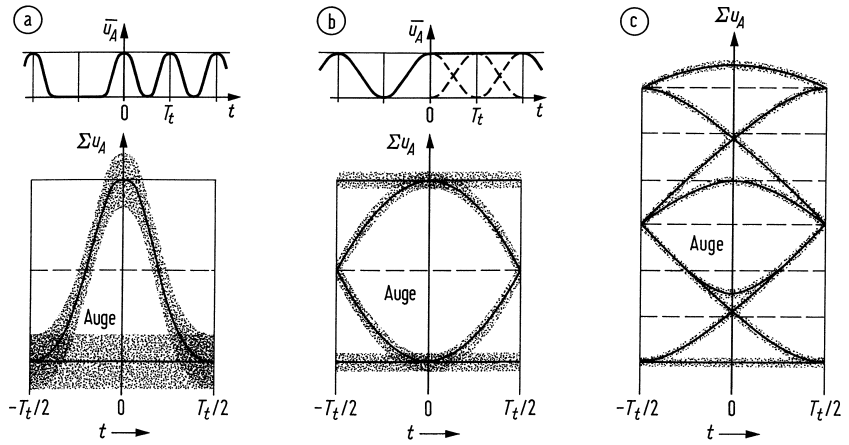


Fig. 7.2. Eye diagrams for RZ (return to zero) pulses, sampling time $t = 0$; solid lines: without noise. (a) large noise, no impulse overlap (b) optimum case: small noise, impulse overlap, but no intersymbol interference at sampling time (c) low noise, strong impulse overlap, strong intersymbol interference at sampling time. Auge = eye

the clock frequency equals the bit rate (unit bit/s). Inside each time slot T_t at sampling times t_A determined by the clock-recovery circuit the decision circuit compares the received signal $u(t_A)$ to a threshold level u_S, and decides whether the signal corresponds to a logical one ($u(t_A) > u_S$) or a logical zero ($u(t_A) < u_S$). The best sampling time corresponds to the situation in which the signal level difference between 1-bit and 0-bit is maximum. It can be determined from the eye diagram

[2] See Sect. 4.3.3 Page 159 in reference Footnote 9 on Page 4

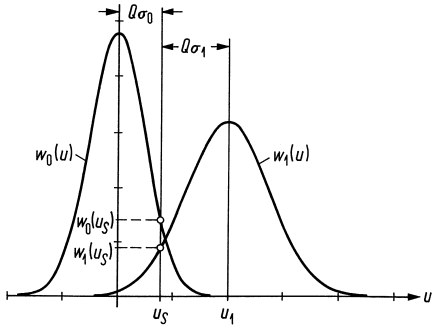


Fig. 7.3. Probability densities $w_0(u)$, $w_1(u)$ of the sampled input voltage of the decision circuit for the receive symbols zero, one. σ_0, σ_1 standard deviations, u_1 expectation of voltage for a received one, u_S specific choice of decision threshold fixed by the bit error parameter Q

formed by superposing electrical random bit sequences on top of each other, Fig. 7.2. The decision circuit receives a voltage

$$\begin{aligned} u(t) &= u_{R0}(t) && \text{(0-bit received)},\\ u(t) &= u_{R1}(t) + u_1 h_A(t) && \text{(1-bit received)}. \end{aligned} \tag{7.5}$$

The quantities $u_{R0}(t)$, $u_{R1}(t)$ are random voltages with a Gaussian distribution and an expectation zero. An equalizing filter shapes the received impulse such that the resultant $h_A(t)$ has $h_A(t_A) = 1$ for the sampling time $t_A = 0$, and that any interference with impulses $h_A(0 \pm nT_t)$ at neighbouring sampling points $n = 1, 2, \ldots$ disappears, $h_A(\pm nT_t) = 0$. The expectation of the voltage at the sampling time is denoted as u_1.

The probability density functions for the decision circuit voltages at the sampling times are $w_0(u)$ and $w_1(u)$ for a received 0 and 1, respectively,

$$\begin{aligned} w_0(u) &= \tfrac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\tfrac{u^2}{2\sigma_0^2}\right), \qquad \overline{u} = 0, \quad \overline{(u-\overline{u})^2} = \overline{u_{R0}^2} = \sigma_0^2,\\ w_1(u) &= \tfrac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\tfrac{(u-u_1)^2}{2\sigma_1^2}\right), \ \overline{u} = u_1, \ \overline{(u-\overline{u})^2} = \overline{u_{R1}^2} = \sigma_1^2. \end{aligned} \tag{7.6}$$

Because bit 1 is transmitted at a higher optical power level than bit 0, its shot noise variance is larger, $\sigma_1^2 > \sigma_0^2$, see Eq. (6.4). Figure 7.3 displays the probability densities $w_0(u)$, $w_1(u)$. In most practical cases we have $1 < \sigma_1^2/\sigma_0^2 \leq 2$. For an arbitrary decision threshold u_S all samples $u < u_S$ are detected as 0, and all samples $u > u_S$ as 1. The probability of erroneously *d*eciding 1 when 0 is *r*eceived is $p(1\mathrm{d}|0\mathrm{r})$ (and $p(0\mathrm{d}|1\mathrm{r})$ for the opposite case). The probabilities of receiving logical 0 and logical 1 are $p(0\mathrm{r})$ and $p(1\mathrm{r})$, respectively. The optimum decision threshold u_S minimizes the bit error ratio (BER, bit error probability) and requires $\partial\,\mathrm{BER}\,/\partial u_S = 0$,

$$\begin{aligned} \mathrm{BER} &= p(1\mathrm{r})p(0\mathrm{d}|1\mathrm{r}) + p(0\mathrm{r})p(1\mathrm{d}|0\mathrm{r}),\\ & \qquad p(0\mathrm{r}) + p(1\mathrm{r}) = 1. \end{aligned} \tag{7.7}$$

The formulae for the error function $\mathrm{erf}(z)$ and the complementary error function $\mathrm{erfc}(z)$ are:

$$\begin{aligned} &\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2)\,\mathrm{d}t, \qquad \mathrm{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty \exp(-t^2)\,\mathrm{d}t,\\ &\mathrm{erf}(\infty) = 1, \qquad \mathrm{erfc}(\pm z) = 1 \mp \mathrm{erf}(z). \end{aligned} \tag{7.8}$$

An approximation for $z > 2$ with an error $< 3\,\%$ is specified[3] in the first part of Eq. (7.9), while a

[3] Abramowitz, M.; Stegun, I. A. (Eds.): Handbook of mathematical functions, 9. Ed. New York: Dover Publications 1970. Chapter 7 Eq. (7.1.23) Page 298

display of $\lg\{-\lg[\mathrm{erfc}(z)]\}$ vs. $\lg z$ results essentially in a straight line:

$$\mathrm{erfc}(z) = \frac{\exp(-z^2)}{\sqrt{\pi z^2}}\left[1 - \frac{1}{2z^2} + \ldots\right], \quad \lg\{-\lg[\mathrm{erfc}(z)]\} \approx \{z \gg 1\} \approx 2\lg z + 0.434\,. \tag{7.9}$$

For Gaussian density functions

$$w(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-A)^2}{2\sigma^2}\right] \tag{7.10}$$

we apply Eqs. (7.8), (7.9) and find

$$\int_{z_1}^{z_2} w(z)\,\mathrm{d}z = \frac{1}{2}\,\mathrm{erf}\left(\frac{z-A}{\sigma\sqrt{2}}\right)\Bigg|_{z_1}^{z_2} = \frac{1}{2}\,\mathrm{erfc}\left(\frac{z-A}{\sigma\sqrt{2}}\right)\Bigg|_{z_2}^{z_1}. \tag{7.11}$$

Referring to Fig. 7.3 and Eq. (7.6) we have

$$\begin{aligned} p(0\mathrm{d}|1\mathrm{r}) &= \int_{-\infty}^{u_S} w_1(u)\,\mathrm{d}u = \tfrac{1}{2}\,\mathrm{erfc}\Big(\tfrac{u_1-u_S}{\sigma_1\sqrt{2}}\Big), \\ p(1\mathrm{d}|0\mathrm{r}) &= \int_{u_S}^{\infty} w_0(u)\,\mathrm{d}u = \tfrac{1}{2}\,\mathrm{erfc}\Big(\tfrac{u_S}{\sigma_0\sqrt{2}}\Big). \end{aligned} \tag{7.12}$$

The results Eq. (7.12) have to be substituted into Eq. (7.7).

Optimum decision With equal probabilities of logical 0 and 1, $p(1\mathrm{r}) = 1 - p(0\mathrm{r}) = \frac{1}{2}$, the optimum decision threshold for monomodal probability density functions $w_{0,1}(u)$ as in Fig. 7.3 (only one maximum) is to be found at the intersection $w_1(u_S) = w_0(u_S)$, independent of the detailed shapes of the probability density function. Because of the typical values

$$\begin{aligned} &1 < \sigma_1/\sigma_0 \le \sqrt{2}\,, \qquad \sigma_1 \approx \sigma_0 \\ &p(1\mathrm{r}) = \frac{\sigma_1}{\sigma_0}\,p(0\mathrm{r}) \gtrapprox p(0\mathrm{r}), \quad p(1\mathrm{r}) \approx p(0\mathrm{r}) \approx 1/2 \end{aligned} \tag{7.13}$$

the relation $p(1\mathrm{r})/\sigma_1 = p(0\mathrm{r})/\sigma_0$ can be replaced approximately by $p(1\mathrm{r}) \approx p(0\mathrm{r}) \approx 1/2$. For $p(1\mathrm{r})/\sigma_1 = p(0\mathrm{r})/\sigma_0$ we find the following simple rule for an optimum decision threshold (see Fig. 7.3),

$$\begin{aligned} &Q = \frac{u_1}{\sigma_0+\sigma_1} = \frac{u_S}{\sigma_0} = \frac{u_1-u_S}{\sigma_1}\,, \\ &p(0\mathrm{d}|1\mathrm{r}) = p(1\mathrm{d}|0\mathrm{r}) = \tfrac{1}{2}\,\mathrm{erfc}(Q/\sqrt{2}\,), \\ &\sigma_0 w_0(u_S) = \sigma_1 w_1(u_S) = \exp(-Q^2/2)/\sqrt{2\pi}\,. \end{aligned} \tag{7.14}$$

Substituting into Eq. (7.7), the minimum bit error probability reads

$$\mathrm{BER} = \frac{1}{2}\,\mathrm{erfc}\left(\frac{Q}{\sqrt{2}}\right). \tag{7.15}$$

The bit-error parameter Q assumes values of $Q = 3.7$, 6, 6.7, 7.3, 7.9 for BER = 10^{-4}, 10^{-9}, 10^{-11}, 10^{-13}, 10^{-15}. For a given Q and known standard deviations σ_0, σ_1 of the noise, the signal voltage u_1 and the optimum threshold may be computed from Eq. (7.14).

For $p(1\mathrm{r})/\sigma_1 \neq p(0\mathrm{r})/\sigma_0$ the choice of u_S according to Eq. (7.14) leads again to a bit error probability Eq. (7.15) (i. e., a BER value independent of $p(1\mathrm{r})$ and $p(0\mathrm{r})$), but this threshold is no more optimum, and the BER is no longer minimum. This drawback is counterbalanced by the advantage that the estimated BER does not depend on the actual bit probabilities $p(1\mathrm{r})$ and $p(0\mathrm{r})$.

When computing the BER, the shapes of the probability density functions $w_0(u)$, $w_1(u)$ are required. Thus, there is no unique dependency of the SNR defined by moments $\overline{u}$, $\overline{u^2}$ up to the second order, compare Eq. (7.1), and the BER. However, the Gaussian distribution is fully determined by moments up to the second order, and a unique connection between the SNR and the BER may be established, if the noise signals at the decision circuit input are truly Gaussian. From a simple measurement of the mean $\overline{u}$ and the effective fluctuation $\sqrt{\overline{u^2}}$ the SNR may be determined. This is also very important for numerical simulations, where it is practically impossible to simulate 100 erroneous bits out of 100×10^9 bits for BER $= 10^{-9}$. The real-time bit rate achieved by a numerical simulation is about 64 kbit/s for a computer with 3×10^9 floating-point operations per second. This leads to a computing time of 18 days *for only one value* of the optical power P_e.

Signal-to-noise power ratio We assume a 50 %-RZ signal according to Eq. (7.5). Intersymbol interference of the impulses $h_A(t)$ be neglected, i. e., the impulses are assumed not to interfere with a signal in neighbouring clock periods. Zeros and ones are equally distributed, f_t is the bit rate, $T_t = 1/f_t$ the clock period and $B = f_t/2$ the minimum electrical (Nyquist) bandwidth which is needed to transmit the impulses. With Eqs. (7.5), (7.6) the mean electrical power P at the decision circuit is

$$\begin{aligned} P &= \tfrac{1}{2}\left\{ \tfrac{1}{T_t}\int_{-T_t/2}^{+T_t/2} \overline{[u_1 h_A(t) + \underbrace{u_{R1}(t)}_{\overline{u_{R1}}=0}\,]^2}\,\mathrm{d}t + \tfrac{1}{T_t}\int_{-T_t/2}^{+T_t/2} \overline{u_{R0}^2(t)}\,\mathrm{d}t \right\} \\ &= \tfrac{u_1^2}{2} I(h_A) + \tfrac{1}{2}(\sigma_0^2 + \sigma_1^2), \\ I(h_A) &= \tfrac{1}{T_t}\int_{-T_t/2}^{+T_t/2} h_A^2(t)\,\mathrm{d}t\,. \end{aligned} \tag{7.16}$$

If the SRV is computed from Eq. (7.16), and if we substitute u_1 from Eq. (7.14), the ratio of *electrical* signal power P_S and the *electrical* noise power P_R is

$$\gamma = \frac{P_S}{P_R} = \frac{u_1^2 I(h_A)/2}{(\sigma_0^2 + \sigma_1^2)/2} = Q^2 \frac{(\sigma_0 + \sigma_1)^2 I(h_A)}{\sigma_0^2 + \sigma_1^2}\,. \tag{7.17}$$

Because we have typically $1 < \sigma_1^2/\sigma_0^2 \le 2$ (Eq. (7.13)), and the integral assumes values around $I(h_A) = 1/2$ for usual impulse shapes (e. g., for a raised cosine $h_A(t) = \frac{1}{2}\big[1 + \cos(2\pi t/T_t)\big] = \cos^2(\pi t/T_t)$ we find $I(h_A) = 3/8 \approx 1/2$), we have approximately

$$\gamma = \frac{P_S}{P_R} = (0.97 \ldots 1) \times Q^2, \qquad \rightarrow \qquad \gamma = \frac{P_S}{P_R} = Q^2. \tag{7.18}$$

In summary: If the noise at the decision circuit follows a Gaussian distribution, if the impulse shapes are such that $I(h_A) \approx 1/2$ and if $1 < \sigma_1^2/\sigma_0^2 \le 2$, then the bit-error parameter Q and the BER may be deduced from a measurement of γ. For BER $= 10^{-9}$ we have from Eq. (7.15) $Q = 6$ and from Eq. (7.18) $\gamma = 36 \mathrel{\widehat{=}} 15.6\,\mathrm{dB}$.

Assuming a signal-independent noise power P_R, (e. g., electronic receiver noise dominates), the signal quality factor Q is proportional to the *optical* signal power, $Q \sim \overline{P_e}$. A display of $\lg(-\lg \mathrm{BER})$ vs. $\lg \overline{P_e}$ then results approximately in a straight line, see Eq. (7.15) on Page 128 and Eq. (7.9) on Page 128.

Forward error correction (FEC) With increasing computing power, a redundancy transmission scheme became possible, where codes are transmitted which allow an error correction at the receiver side. For an optical communication channel at 40 Gbit/s the bit error ratio (BER) performance of moderate-length nonbinary low-density parity-check (LDPC) codes show that[4]

[4] Djordjevic, I. B.; Vasic, B.: Nonbinary LDPC codes for optical communication systems. IEEE Phot. Technol. Lett. 17 (2005) 2224–2226

- an LDPC(1225, 1088) code of moderate length with 12.59 % redundancy achieves a coding gain of 9.9 dB at BER = 10^{-10}, and
- an LDPC(29136, 27 315) with 6.7 % redundancy achieves a coding gain of 9.7 dB at BER = 10^{-10}.

Without going into coding details, these examples demonstrates the potential of the technique: A coding gain of about 10 dB means that 10 dB less power may be received for a final BER = 10^{-10} than without FEC. This allows a BER of the order of 10^{-4} if FEC of the signal is ignored. The price is an increased gross bit rate of, e. g., 43 Gbit/s for a redundancy of 7 %.

Power penalty Depending on the circumstances, an additional noise source with an additional electrical noise power P_{Rz} may be compensated by increasing the electrical signal power from P_S to P_{S+}. Because of Eq. (7.18) we write

$$\gamma = Q^2 = \frac{P_S}{P_R} = \frac{P_{S+}}{P_R + P_{Rz}} \,. \tag{7.19}$$

We define the quantities γ_+, Q_+, BER_+ by

$$\gamma_+ = Q_+^2 = \frac{P_S+}{P_R} \,, \qquad \mathrm{BER}_+ = \frac{1}{2}\,\mathrm{erfc}\left(\frac{Q_+}{\sqrt{2}}\right). \tag{7.20}$$

The SRV which would have been measured for an increased electrical signal power P_{S+} without additional noise sources is denoted as γ_+.

Additive noise If P_{Rz} stands for signal-independent additive electrical noise power, we calculate from Eqs. (7.19), (7.20)

$$\frac{P_{S+}}{P_S} = \left(\frac{Q_+}{Q}\right)^2 = 1 + \frac{P_{Rz}}{P_R} \,. \tag{7.21}$$

Additional additive noise can always be compensated by an increased optical power $\overline{P_e}$ at the receiver input. For the electrical signal power we have the relation $P_S \sim i_S^2 = S^2\overline{P_e}^2$, an we define a power penalty by

$$p_B = 10\,\lg\left(\frac{\overline{P_{e+}}}{\overline{P_e}}\right) = 5\,\lg\left(\frac{P_{S+}}{P_S}\right) = 10\,\lg\left(\frac{Q_+}{Q}\right) = 5\,\lg\left(1 + \frac{P_{Rz}}{P_R}\right). \tag{7.22}$$

Multiplicative noise If the additional noise power is always in proportion to the signal power $P_S \sim P_{Rz}$ it may be specified by a fixed quantity $\gamma_R = Q_R^2$ ($r$esidual bit error),

$$P_{Rz} = \frac{1}{\gamma_R} \cdot P_S = \frac{1}{Q_R^2} \cdot P_S \,, \qquad \gamma_R = Q_R^2 = \frac{P_S}{P_{Rz}} \,. \tag{7.23}$$

For the case of additive noise (Eq. (7.19)) the bit error parameter Q could be increased and the BER decreased arbitrarily by increasing the signal power P_S. For multiplicative noise (Eq. (7.23)) the BER is limited to a minimum BER_R by Q_R. Substituting the additional noise specified in Eq. (7.23) into Eq. (7.19), one calculates

$$\gamma = Q^2 = \frac{P_S}{P_R} = \frac{P_{S+}}{P_R + P_{Rz}} = \frac{P_{S+}}{P_R + P_{S+}/Q_R^2} = \frac{Q_+^2 Q_R^2}{Q_+^2 + Q_R^2} \,. \tag{7.24}$$

With Q_+/Q from Eq. (7.24) the power penalty Eq. (7.22) becomes

$$p_B = 10\,\lg\left(\frac{\overline{P_{e+}}}{\overline{P_e}}\right) = 10\,\lg\left(\frac{Q_+}{Q}\right) = 5\,\lg\left(\frac{Q_R^2}{Q_R^2 - Q^2}\right). \tag{7.25}$$

Only for $Q_R > Q$ (i. e., $P_{Rz} < P_R$) the additional noise may be compensated by an increased signal power. For $p_B \to \infty$ the bit error parameter reaches the limit $Q \to Q_R$. The BER approaches asymptotically the residual or floor error "rate"

$$\mathrm{BER}_R = \frac{1}{2}\,\mathrm{erfc}\left(\frac{Q_R}{\sqrt{2}}\right). \tag{7.26}$$

Substituting Q from Eq. (7.24) into Eq. (7.15) leads to

$$\mathrm{BER} = \frac{1}{2}\,\mathrm{erfc}\left(\frac{Q}{\sqrt{2}}\right) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\frac{Q_+ Q_R}{\sqrt{Q_+^2 + Q_R^2}}\right). \tag{7.27}$$

In Fig. 7.4 the bit error probability Eq. (7.27) is displayed as a function of Q and of Q_+. For additive

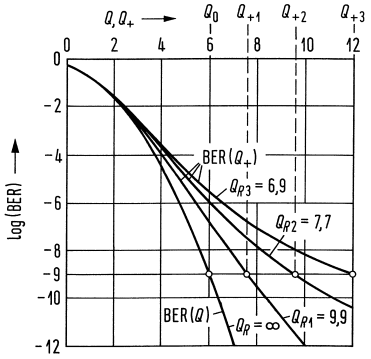


Fig. 7.4. Bit error probability from Eq. (7.27) as a function of the bit error parameters Q (denoted as BER(Q)) and Q_+ (denoted as BER(Q_+)) for various values of the residual bit error parameter Q_R

noise only (i. e., $Q_R \to \infty$, see Eq. (7.23)) the BER decreases monotonically with increasing Q (increasing signal power P_S, see Eq. (7.18)). At $Q = Q_0 = 6$ an value of BER $= 10^{-9}$ is reached. With multiplicative noise as in Eq. (7.23), characterized by a floor error probability (Eq. (7.26)) of $\mathrm{BER}_R = 2.6 \times 10^{-23}$, 5.2×10^{-15}, 2.0×10^{-12} ($Q_R = 9.9$, 7.7, 6.9), a bit error probability of BER $= 10^{-9}$ requires $Q_{+1} = 7.55$, $Q_{+2} = 9.5$, $Q_{+3} = 12$. After Eq. (7.20), Q_+ measures the increase in signal power, $Q_+ \sim \sqrt{P_{S+}} \sim \overline{P_{e+}}$). The power penalties Eq. (7.25) amount to $p_B = 1\,\mathrm{dB}$, $2\,\mathrm{dB}$, $3\,\mathrm{dB}$.

An example for multiplicative noise Eq. (7.23) are uncertainties (jitter) of the sampling time. If the sampling at $t = 0$ takes place with an effective uncertainty of Δt, the residual bit error parameter is[5]

$$Q_R = \frac{\sqrt{128}}{\pi^2 (\Delta t / T_t)^2}\,. \tag{7.28}$$

For $Q_R = 9.9$ ($p_B = 1\,\mathrm{dB}$ at BER $= 10^{-9}$) the maximum jitter is given by $\Delta t/T_t \le 0.34$, i. e., $\Delta t = 85\,\mathrm{ps}$ at a bit rate of $f_t = 4\,\mathrm{Gbit/s}$.

Quantum limit From Eqs. (7.3), (7.18), (7.15) we estimate a "shot-noise limited" bit error ratio

$$\mathrm{BER} = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\eta N_e}\Big/\sqrt{2}\right). \tag{7.29}$$

[5] Shen, T. M.: Power penalty due to decision-time jitter in receivers using avalanche photodiodes. Electron. Lett. 22 (1986) 1043–1045

For an optimum receiver with $\eta = 1$ and BER $= 10^{-9}$ the mean number of photons per 1-bit is $N_e = 36$. As stated after Eqs. (7.3), actual direct receivers without optical amplifiers are thermal-noise limited.

The BER expression Eq. (7.29) is not truly accurate, since its derivation is based on the Gaussian approximation for the receiver noise statistics. For an ideal detector (no electronic noise, no dark current, quantum efficiency $\eta = 1$), we have $\sigma_0 = 0$, and vanishing quantum noise in the absence of incident optical power, so the decision threshold can be set arbitrarily close to the 0-level signal. Indeed, for such an ideal receiver, 1-bits can be identified without error as long as at least one photon is detected. An erroneous detection occurs only if a 1-bit fails to produce an electron-hole pair. For such a small number of photons and electrons, shot-noise statistics cannot be approximated by a Gaussian distribution, and the exact Poisson statistics Eq. (6.6) must be used.[6]

If zero photons arrive for a 0-bit, and $N_e > 0$ is the average number of photons in each 1-bit, the probability that a 0-bit is wrongly taken for a 1-bit is $w(0\mathrm{r}|1\mathrm{d}) = 0$, and the probability of taking wrongly a 1-bit for a 0-bit is $p(0\mathrm{d}|1\mathrm{r}) \neq 0$. For the Poisson distribution the probability of deciding for a 0-bit when actually a 1-bit was received is given by

$$p(0\mathrm{d}|1\mathrm{r}) = p_N(0) = \frac{N_e^0}{0!}\,\mathrm{e}^{-N_e} = \mathrm{e}^{-N_e}\,. \tag{7.30}$$

The probability of a 1-bit be $w(1\mathrm{r}) = 1/2$. The probability of deciding for a 0-bit instead of the received 1-bit is given by Eq. (7.30). The bit error probability according to Eq. (7.7) is

$$\mathrm{BER} = \overbrace{p(1\mathrm{r})}^{1/2}\overbrace{p(0\mathrm{d}|1\mathrm{r})}^{\exp(-N_e)} + \overbrace{p(0\mathrm{r})}^{1/2}\overbrace{p(1\mathrm{d}|0\mathrm{r})}^{0}, \quad \text{therefore:} \quad \boxed{\mathrm{BER} = \tfrac{1}{2}\,\mathrm{e}^{-N_e}} \tag{7.31}$$
$$p(0\mathrm{r}) + p(1\mathrm{r}) = 1\,,$$

For BER $= 10^{-9}$ one needs

$$N_e = -\ln\left(2 \times 10^{-9}\right) = 20 \quad \Longrightarrow \quad \boxed{N_e = 20 \quad \text{for} \quad \mathrm{BER} = 10^{-9}} \tag{7.32}$$

photons for a 1-bit. Since this requirement is a direct result of quantum fluctuations associated with the incoming light, it is referred to as the quantum limit. Each 1-bit must contain at least $N_e = 20$ photons to be detected with a BER $< 10^{-9}$. This requirement can be converted into power by using the relation $\overline{P_e} = N_e h f_S / T_t$. The receiver sensitivity, defined as $\overline{P_{e\,\mathrm{bit}}} = (\overline{P_e} + 0)/2 = \overline{P_e}/2$ is given by (for the choice of the signal bandwidth $B = 1/(2T_t)$ see text after Eq. (7.2))

$$\overline{P_e} = \frac{N_e h f_S}{T_t}\,, \qquad \overline{P_{e\,\mathrm{bit}}} = \frac{\overline{P_e}}{2} = \frac{N_e h f_S}{2T_t} = N_e h f_S B = 2N_{e\,\mathrm{bit}} h f_S B\,,$$
$$N_{e\,\mathrm{bit}} = (N_e + 0)/2 = N_e/2\,. \tag{7.33}$$

Therefore an average number of $N_{e\,\mathrm{bit}} = N_e/2 = 10$ photons per bit (1-bit and 0-bit with equal probability) must be received. This represents the absolute quantum limit for an ideal binary system with direct reception for BER $= 10^{-9}$.

Most receivers operate 20 dB or more above the quantum limit. This is equivalent to saying that N_e typically exceeds 1 000 photons in practical pin-photodiode receivers. With avalanche photodiodes (APD), optical preamplifiers or with heterodyne reception the quantum limit may be approached more closely. Figure 7.5 displays some experimental data. With direct reception the following receivers were realized for BER $= 10^{-9}$:

$N_e = 4\,000$	pin-photodiode
$N_e = \ 150$	APD
$N_e = \ 152$	pin-photodiode and optical preamplifier

(7.34)

[6] See Sect. 4.5.3 Page 175 in reference Footnote 9 on Page 4

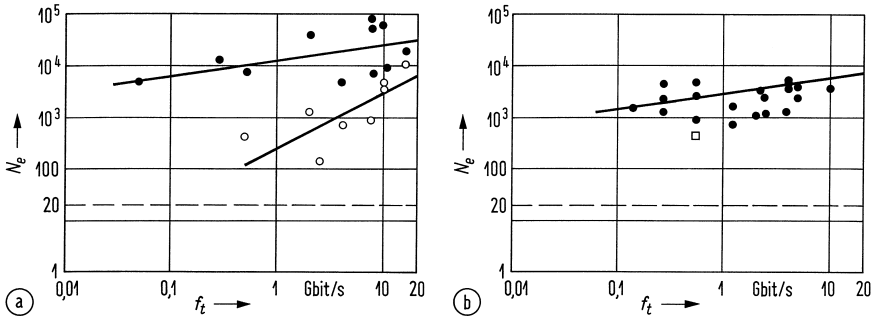


Fig. 7.5. Measured minimum received photon numbers N_e for a 1 bit (η not known, BER $= 10^{-9}$). Quantum limit $N_e = 20$, $\eta = 1$ (– – –). (a) pin-photodiode $\lambda = 1.3$; $1.55\,\mu$m (•), pin-photodiode with optical amplifier (○) (b) avalanche photodiode (APD) $\lambda = 1.55\,\mu$m (•), $\lambda = 0.85\,\mu$m (□)

Summaries, problems and quizzes

1 Introduction

Summary

1. Light may be either viewed as an electromagnetic wave characterized by its frequency $f = c/\lambda$ and described by Maxwell's equations, or as a particle (photon) with energy $W = hf$ and de Broglie momentum $p = h/\lambda$, depending on the experiment.

2. Optical communication systems use lightwaves with a vacuum wavelength $0.6\,\mu\text{m} \ldots 1.2\,\mu\text{m} \leq \lambda \leq 1.6\,\mu\text{m}$ corresponding to carrier frequencies $500\,\text{THz} \ldots 250\,\text{THz} \geq f \geq 190\,\text{THz}$.

3. A point-to-point transmission link is referred to as a *system*. Many interconnected links with multiplexing or switching functions are called a communication *network*.

4. Semiconductor laser diodes or LED serve as transmitters. Switching their electric current $i(t)$ (unit A) on and off with time t modulates the light intensity $I(t)$ (unit W/m^2) in a binary fashion. This type of modulation is called intensity modulation (IM). The most common digital modulation format is pulse code modulation (PCM).

5. Light rays normally go in straight lines, but optical fibres can guide them around corners via total internal reflection. The low refractive index cladding confines light to the core with an effective area F.

6. Group delay dispersion is caused by geometrical ray path differences in multimode fibres, or by different group delays in singlemode fibres depending on the spectral content of the optical signal. This leads to a temporal broadening of transmitted light impulses, see Figs. 1.4 and 1.5 on Pages 3.and 4.

7. A photodetector with sensitivity S (unit A/W) reconverts light to an electric photocurrent $i(t) = SFI(t)$ in proportion to the light intensity.

8. Optical fibres have a much lower loss than copper cables. With $0.2\,\text{dB}/\text{km}$ loss, 40 % of the input power is still available after transmission through a 20 km long quartz glass singlemode fibre (the same attenuation holds for only 20 m of coaxial cable!).

9. Optical fibre loss may be compensated with optical amplifiers.

10. Optical fibres have a much larger transmission capacity than copper cables. This is specially true for singlemode fibres which are therefore employed in long-haul transmission links.

11. The dispersive broadening of transmitted light impulses leads to intersymbol interference and an increased bit error probability, such limiting the maximum transmission rate.

12. Dispersion in singlemode fibres, however, may be compensated with specifically designed fibre sections.

13. The capacity of transmission links can be greatly extended by employing more than one optical carrier ("channel") in a wavelength division multiplexing scheme (WDM). Typical channel spacings are $\Delta f = 100\,\text{GHz}$ in the vicinity of $\lambda = 1.55\,\mu\text{m}$.

14. Bit rates and link lengths of transmission media are:

    (a) Copper twisted pair 6 Mbit/s (6 km)

    (b) Coaxial cable 650 Mbit/s (1.5 km)

    (c) Singlemode glass fibre 40 Gbit/s single channel (1×10^6 km);[7] 1.28 Tbit/s single channel (240 km);[8]

[7] J. Leuthold, G. Raybon, Y. Su, R. Essiambre, S. Cabot, J. Jaques, M. Kauer: 40 Gbit/s transmission and cascaded all-optical wavelength conversion over 1 000 000 km. Electron. Lett. vol. 38 no. 15, July 2002. (Bell / IHQ)

[8] H.G. Weber, S. Ferber, M. Kroh, C. Schmidt-Langhorst, R. Ludwig, V. Marembert, C. Boerner, F. Futami, S. Watanabe, C. Schubert: Single channel 1.28 Tbit/s and 2.56 Tbit/s DQPSK transmission. Electron. Lett. Vol. 42 No. 3, Feb. 2006. (HHI)

fibre+WDM 25 Tbit/s (400×10^6 full ISDN channels, 200 000 full TV channels)

Problems

1. Most of the light lost in going through a glass window is reflected at the surface (4 %). Ignoring this surface reflection loss, suppose that a one-millimeter window absorbs 1 % of the light entering and transmits 99 %. How much light would emerge from a 7 cm thick laminated-glass window at the cash desk of a bank if the absorption of the intermediate polymer layers is neglected?
2. Why need optical amplifiers be distributed along a long-haul transmission distance?
3. If optical fibres transmit signals so much better than metallic wires, why are they not used everywhere?
4. Electromagnetic interference occurs in metallic wires because electromagnetic fields cause electrons to move inside the wires, generating currents that interfere with the signal. This is also how antennas work. Why does electromagnetic interference not affect signals in optical fibres?
5. Most switching in the telephone and data networks is done by special-purpose computers. Why do network operators continue to convert most optical signals to electronic form for switching?

Quiz

1. What is the principal requirement for a cladding in an optical fibre?
   (a) It must have a refractive index lower than the core to produce total internal reflection.
   (b) It must be opaque so light does not leak out.
   (c) It must be made of plastic to keep the fibre flexible.
   (d) It must have a lower refractive index than air.
2. Today's best optical fibres with an attenuation of about 0.2 dB / km transmit light so well that 10 % of the light entering the fibre remains after
   (a) 0.5 km
   (b) 4 km
   (c) 20 km
   (d) 50 km
   (e) 100 km
3. All fibre-optic communication systems must include
   (a) transmitter, receiver and fibre.
   (b) transmitter and cable.
   (c) fibre and receiver.
   (d) fibre only.
   (e) cable only.
4. A switched telecommunication system makes connections between pairs of terminal devices. Which is the best example of a switched network?
   (a) Cable television
   (b) Telephone network
   (c) A link between a single personal computer and a printer
   (d) A network that shares one printer among four personal computers
5. The notion "optical channel" associates
   (a) a television channel transmitted optically.
   (b) all the light transmitted through a single fibre.
   (c) one of several wavelengths carrying separate signals through an optical fibre.
   (d) the electrical input to an optical transmitter.
6. Which of the following systems can transmit with the highest speed over the longest distance?
   (a) An optical fibre carrying a single, pure wavelength
   (b) A coaxial cable carrying many video signals
   (c) An optical fibre carrying signals at many separate wavelengths
   (d) A coaxial cable transmitting light

2 Light waveguides — Maxwell's equations for the curious ones

Summary (and some supplements 2–8)

1. Dielectric constant and magnetic permeability, velocity of light, wavelength in vacuum at frequency f or angular frequency $\omega = 2\pi f$ are ϵ_0, μ_0, $c = 1/\sqrt{\epsilon_0\mu_0}$, $\lambda = c/f$. The wave impedance of vacuum is $Z_0 = \sqrt{\mu_0/\epsilon_0}$, see Eq. (2.1) on Page 7.

2. The divergence theorem relates the total flux (German *Ergiebigkeit*) $Q = \oint_F \vec{D} \cdot \vec{e_n}\,\mathrm{d}F$ of a vector field $\vec{D}$ out of a closed surface F (outward pointing normal $\vec{e}_n$) to the integral $Q = \iiint_V \operatorname{div}\vec{D}\,\mathrm{d}V$ of the divergence over the enclosed volume V (Gauss' theorem).

3. Therefore, the divergence (German *Quellendichte*) of a vector field $\vec{D}$ is the net differential outflow integral per differential volume at a certain point $\vec{r}$, $\operatorname{div}\vec{D} = \lim_{V\to 0}\left(\oint_F \vec{D}\cdot\vec{e_n}\,\mathrm{d}F/V\right)$. A positive (negative) divergence equals the rate of decrease (increase) of the spatially diverging (converging) flow lines per volume (German *Feldliniendichte*).

4. Positive or negative electric charges ϱ are sources or sinks of electric flow lines (German *Feldlinien*), and $\operatorname{div}\vec{D} = \operatorname{div}(\epsilon\vec{E}) = \varrho$. Flow lines begin at a source and end at a sink or may be even open, e. g., for an isolated charge. The total charge Q in a volume V may be computed by either form of Gauss' theorem.

5. Fields with $\operatorname{div}\vec{D} = 0$ are called solenoidal (German *quellenfrei*), and the associated flow lines are closed. A solenoid is a cylindrical coil of wire acting as a magnet when carrying electric current. The naming "solenoidal" reflects the general property $\operatorname{div}\vec{B} = 0$ for the magnetic induction $\vec{B}$ which may be generated by a solenoid: Isolated magnetic charges do not exist.

6. Stokes' theorem is the curl analogue of the divergence theorem and relates the integral $I = \iint_F \operatorname{curl}\vec{H}\cdot\vec{e_n}\;\mathrm{d}F$ of the the vortex density vector $\operatorname{curl}\vec{H}$ (German *Wirbeldichte*) over an open surface F (normal $\vec{e}_n$) to the line integral (circulation integral, German *Zirkulation*) $I = \oint_C \vec{H}\cdot\vec{e_t}\,\mathrm{d}s$ of the vector field around the perimeter C bounding the surface (closed-contour integral, contour C right-handed around $\vec{e}_n$).

7. The flow lines of the vector field $\operatorname{rot}\vec{H}$ are called vorticity lines (German *Wirbellinien*).

8. The flow lines of $\vec{H}$ for $\operatorname{curl}\vec{H} \neq 0$ are always closed, and $\operatorname{div} H = 0$ results. Fields $\vec{H}$ with $\operatorname{curl}\vec{H} = 0$ are called vortex-free (German *wirbelfrei*).

9. Neither currents nor electric space charges ϱ be involved at the moment. The medium is isotropic, linear and non-magnetic, so that its properties are given by scalar, amplitude-independent quantities; the relative magnetic permeability is $\mu_r = 1$.

10. Maxwell's partial differential equations (2.1) on Page 7 connect the magnetic vortex density $\operatorname{curl}\vec{H}$ with the the positive time derivative of the displacement $\vec{D}$, and the electric vortex density $\operatorname{curl}\vec{E}$ with the negative time derivative of the magnetic induction $\vec{B}$.

11. Maxwell's equations are supplemented by the constitutive relations formulating the absence of electric space charges, $\operatorname{div}\vec{D} = \varrho = 0$, and the solenoidal property of $\vec{H}$ in a non-magnetic medium, $\operatorname{div}\vec{B} = \mu_0 \operatorname{div}\vec{H} = 0$.

12. The displacement $\vec{D}$ is in proportion to the electric field $\vec{E}$ *in vacuo* and becomes modified by the polarization $\vec{P}$ of the medium.

13. A glass molecule becomes deformed by the electric field, and the resulting dipole with polarization $\vec{P}$ reradiates secondary fields, which add up to the vacuum field.

14. Due to the mechanical inertia of the dipole there is a phase lag or lead of the secondary radiation. This results in retarding or advancing the originating field, which in turn gives rise to a decrease or an increase of the phase velocity over c, i. e., an increase or decrease of the refractive index n over 1, see Fig. 2.1 on Page 9.

15. In any low-loss frequency region this influence may be expressed by a relative dielectric constant ϵ_r so that $\vec{D} = \epsilon_0 \epsilon_r \vec{E}$ and $n = \sqrt{\epsilon_r}$.

16. The time-frequency Fourier transform relation reads, see Eq. (2.2) on Page 7,

$$\breve{\Psi}(f) = \int_{-\infty}^{+\infty} \Psi(t)\, \mathrm{e}^{-\mathrm{j}\, 2\pi f t}\, \mathrm{d}t\,.$$

For the inverse transform, exchange $\Psi(t) \Longleftrightarrow \breve{\Psi}(f)$ and make the Fourier exponential complex conjugate, $\mathrm{e}^{+\mathrm{j}\, 2\pi f t}$.

Problems

1. Is it possible that a time-dependent electric field $\vec{E}(t)$ exists, while the magnetic field disappears everywhere and at all times, $\vec{H}(t) = 0$?

2. What is the physical meaning of $\operatorname{div} \vec{D} = 0$? Is $\operatorname{div} \vec{B} \neq 0$ observed in nature?

3. Compute the Fourier spectrum $\breve{\Psi}(f)$ of the rectangular time function $\Psi(t)$ which exists only in an interval $-\Delta t/2 \leq t < \Delta t/2$ where it is Ψ_0. How are spectral and temporal half-maximum widths Δf and Δt interrelated? *Hint*: $(\sin x)/x = 1/2$ for $x \approx 0.60 \times \pi \approx \pi/2$

Quiz

1. For an electrostatic field $\vec{E}$ and $\operatorname{div} \vec{E} \neq 0$, Maxwell's equation

   (a) apply.

   (b) do not apply.

   (c) apply approximately.

2. For a magnetostatic field $\vec{H}$ and $\operatorname{div} \vec{B} \neq 0$, Maxwell's equations

   (a) apply.

   (b) do not apply.

   (c) apply approximately.

3. For an electrostatic field $\vec{E}$ to exist its divergence must obey

   (a) $\operatorname{div} \vec{E} = 0$ everywhere.

   (b) $\operatorname{div} \vec{E} \neq 0$ everywhere.

   (c) $\operatorname{div} \vec{E} \neq 0$ where the sources of the field are located.

2.1 Fundamentals of wave propagation

Summary

1. The complex refractive index $\bar{n} = n - \mathrm{j}\, n_i$ in Eq. (2.6) on Page 8 is a crucial property of optical materials. In a spectral region of a medium without loss or gain its real part n equals the vacuum speed of light c divided by the material speed of light $v = c/n$, so we have always $n > 1$ in this region.

2. The refractive index $n(f)$ changes with frequency. This is called (material) dispersion.

3. Following the electrical engineering convention, monochromatic waves vary in time according to $\exp(+\mathrm{j}\, \omega t)$.

4. Waves are classified according to the shape of their phase surfaces. Plane waves $\Psi(t, \vec{r}) = \exp(\mathrm{j}\, \omega t) \exp(-\mathrm{j}\, \vec{k} \cdot \vec{r})$ have a plane phase surface,

$$\vec{k} \cdot \vec{r} = k_x\, x + k_y\, y + k_z\, z = \text{const}\,.$$

The velocity of the phase surface in propagation direction is the phase velocity $v = c/n$ of the wave.

5. For plane waves to be true solutions of Maxwell's equations the condition

$$\vec{k}^{\,2} = \vec{k} \cdot \vec{k} = n^2 k_0^2 \quad \text{(real!)}, \quad k_0 = \frac{\omega}{c}$$

must hold, Eq. (2.18) on Page 11.

6. The amplitude vector $\vec{a}$ points into the direction where the amplitude decreases fastest. The phase vector $\vec{b}$ is normal to constant-phase surfaces.

7. The plane-wave propagation vector may be expressed by $\vec{k} = \vec{b} - \mathrm{j}\, \vec{a}$. For a real modulus of $\vec{k}$ the condition $\vec{a} \cdot \vec{b} = 0$

holds: Either $\vec{a}$ vanishes and the amplitude remains constant everywhere, or amplitude and phase surfaces are perpendicular to each other.

8. Plane waves are called "homogeneous", if phase and amplitude surfaces coincide, which implies $\vec{a} = 0$; the condition $\vec{b} = 0$ means no wave propagation.

9. For homogeneous plane waves, the vectors $\vec{E}$, $\vec{H}$, $\vec{b}$ form a right handed orthogonal system in this sequence. Because *e*lectric and *m*agnetic field are *t*ransverse ($\vec{E} \perp \vec{b}$, $\vec{H} \perp \vec{b}$, normal to the propagation direction $\vec{b}$) we call it a TEM wave. $\vec{E}$ and $\vec{H}$ are in phase, and the wave impedance Z_0/n of the medium (ratio of two orthogonal components of $\vec{E}$ and $\vec{H}$) is real.

10. Information is transmitted by signals which are localized in time, because otherwise there would be no surprise: The less probable (or the less predictable) an event is the more information it carries. In that sense, a purely random signal ("noise") which need not carry any meaning for a specific human being, transports the ultimate amount of information.

11. A time-localized signal is represented by a superposition ("a wave group") of sinusoidals with different frequencies, see Fig. 2.2 on Page 12. If the velocities v of all waves are identical, the wave group propagates unchanged.

12. Waves with differing frequencies propagate at different velocities $v(f) = c/n(f)$ because of the refractive index dispersion. Then, the constituting waves of an impulse as in Fig. 2.2 do not necessarily maintain the same relative position along z, so that the velocity of the wave group ("group velocity v_g") is different from the phase velocity v, and the shape of the envelope may become distorted.

13. Fused silica glass (SiO_2, quartz) is today's material for high quality light waveguides exhibiting low attenuation and high transmission bandwidth. Losses due to scattering and absorption are described by an exponential power decay $P(z) = P_0 \times \exp(-\alpha z)$, Eq. (2.23) on Page 13.

14. Power attenuation constant α (unit km^{-1}) and attenuation a (unit dB) for a transmission length z are simply interrelated by $a = 4.34\,\alpha z$.

15. Unintentionally existing impurities and intentionally applied impurities (doping) as well as the amorphous nature of glass give rise to nearly isotropic Rayleigh scattering. Its power attenuation constant $\alpha_S \sim f^4 \sim \lambda^{-4}$ has the same frequency dependence as the radiation strength of a Hertzian dipole.

16. A larger doping leads to a larger refractive index, but also to increased attenuating because of Rayleigh scattering, Eq. (2.24) on Page 13.

17. In contrast to Rayleigh scattering, Brillouin and Raman scattering are nonlinear phenomena, where a partial power of a propagating field is transferred to a field of a different frequency.

18. Absorption results from exciting bound charges, either molecule resonances of SiO_2 at $8\,\mu$m (37.5 THz), or electron vibrations at $\lambda_2 = 140\,$nm ($f_2 = 2\,140\,$THz), see Fig. 2.1(b) and 2.3 on Pages 9 and 15.

19. The most disrupting loss peak near $\lambda_{\mathrm{OH}} = 1.38\,\mu$m is due to water and may be avoided with dehydrated fibre preforms.

20. The basic quartz-glass attenuation limit of about 0.2 dB / km can be reached routinely at the wavelength $\lambda_\alpha = 1.55\,\mu$m.

21. The group delay difference of two plane waves at optical carriers differing in wavelength by $\Delta\lambda = \lambda - \lambda_1$ is given by the Taylor expansion $\Delta t_g/L = M\,\Delta\lambda + N\,(\Delta\lambda)^2 + \ldots$ with the first and second order material dispersion coefficients M and N, Eq. (2.27) on Page 16.

22. With respect to a plane boundary in Fig. 2.5 on Page 17 separating regions with different refractive index, a wave is called E-polarized when $\vec{E} = E\,\vec{e}_y$ is parallel to the boundary plane (TE- or H-wave). It is called H-polarized if $\vec{H} = H\,\vec{e}_y$ is oriented parallel to the boundary plane (TM- or E-wave).

23. Waves in a medium n_1 are reflected when hitting the boundary to a medium with a different refractive index n_2.

24. The reflected wave superimposes the incident (monochromatic) wave and forms a sinusoidal standing-wave pattern, see Fig. 2.5 on Page 17.

25. For normal incidence the power reflection and transmission factors are $R = \left(\frac{n_1-n_2}{n_1+n_2}\right)^2$ and $T = 1 - R$.

26. For $n_2 < n_1$ and for an incident angle smaller than the limiting angle of total internal reflection $\vartheta_{1T} = \arccos(n_2/n_1)$, no energy is transported perpendicularly to the boundary $x = 0$, and the reflection coefficient for E- and H-polarization has a modulus of $|r_{E,H}| = 1$.

27. The phase $\varphi_{E,H}$ of $r_{E,H}$ depends on the polarization. An H-polarized wave penetrates deeper into the medium $n_2 < n_1$. This may be inferred from Fig. 2.6 on Page 18: The modulus of the reflection coefficient for E-polarization is always larger than the one for H-polarization, due to the zero of r_H at the Brewster angle ϑ_{2B}.

Problems

1. What are the properties of monochromatic homogeneous plane waves?

2. When do phase and group velocity differ?

3. Infrared and ultraviolet absorption in glass limits the useful frequency range of transmission. What is the minimum-dispersion, what the minimum-attenuation wavelength?

4. Light experiences Rayleigh scattering at the air molecules. How come that the sky appears blue? Why does the sun appear red at dusk and dawn? How would the sky look like if you had travelled to our moon?

5. Diamond (glass) has a refractive index of $n_C = 2.4$ ($n_{SiO_2} = 1.5$). What are the respective critical angles of total internal reflection at the boundary to air, and what does that have to do with the beautiful sparkle of a diamond?

6. List the properties of the reflection coefficient for total internal reflection!

7. What is the power reflection factor for normal incidence at an glass-air interface? What is the power transmission?

8. If you look through the windscreen of a car, and your spectacles are not antireflection-coated, what is the minimum total power loss? Why is it called minimum? *Hint*: Usual windscreens are inclined from the normal.

9. If two plane waves propagate in undoped silica glass at wavelengths $1.550\,\mu\text{m} \pm 0.5\,\text{nm}$, respectively, how do their arrival times differ after a propagation length of 1 km?

10. Consider two plane waves propagating at wavelengths $1.2758\,\mu\text{m} \pm 0.5\,\text{nm}$ in undoped silica glass, what is their arrival difference after a propagation length of 1 km?

11. How large is the relative group delay in glass with a refractive index of $n = 1.5$?

Quiz

1. Which of the following *is not* an electromagnetic wave?

   (a) Radio waves
   (b) Light
   (c) Infrared radiation
   (d) X-rays
   (e) Acoustic waves

2. Optical fibres have minimum loss near $1.5\,\mu\text{m}$. What is the frequency that corresponds to this wavelength?

   (a) 200 MHz
   (b) 20 GHz
   (c) 200 GHz
   (d) 20 THz
   (e) 200 THz
   (f) 20 EHz

3. Light at the interface from air to glass is

   (a) reflected.
   (b) refracted.
   (c) absorbed.
   (d) experiencing all of the above effects.

4. The output of a 20 km fibre with attenuation of 0.5 dB / km is 5 μW. What is the input power to the fibre?
   (a) 0.5 mW
   (b) 0.1 mW
   (c) 0.05 mW
   (d) 0.03 mW
   (e) 0.01 mW

5. What fraction of input power remains after light travels through 100 km of fibre with 0.3 dB / km attenuation?
   (a) 0.1 %
   (b) 0.5 %
   (c) 1 %
   (d) 5 %
   (e) 10 %

6. If a 1 cm thick glass plate transmits 90 % of the light that enters, how much light will emerge from a 10 cm slab of the same glass (no surface reflection)?
   (a) 0 %
   (b) 9 %
   (c) 12 %
   (d) 35 %
   (e) 80 %

7. Doping of glass with GeO_2
   (a) leaves its refractive index unchanged.
   (b) decreases its refractive index.
   (c) increases its refractive index.

8. Doping of glass with F
   (a) leaves its refractive index unchanged.
   (b) decreases its refractive index.
   (c) increases its refractive index.

9. Doping of glass with of GeO_2
   (a) decreases the attenuation.
   (b) increases the attenuation.
   (c) changes the dispersion.
   (d) introduces Mie scattering.

10. Doping of glass with of F
    (a) decreases the attenuation.
    (b) increases the attenuation.
    (c) changes the dispersion.
    (d) introduces Rayleigh scattering.

11. What happens to light at communication wavelengths that is scattered in an optical fibre?
    (a) It escapes from the sides of the fibre.
    (b) Glass atoms absorb its energy.
    (c) Glass atoms store the light and release it later.
    (d) It is reflected back toward the light source.
    (e) It excites acoustic waves in the glass.

12. What effect does dispersion cause?
    (a) Scattering of light out the sides of the fibre
    (b) Stretching of signal impulses that increases with distance
    (c) Shrinking of signal impulses that become shorter with distance
    (d) Attenuation of signal impulses

13. If a monochromatic plane wave is incident on a plane boundary one observes a standing wave
    (a) on both sides.
    (b) on no side.
    (c) inside the medium with the larger refractive index.
    (d) inside the medium with the smaller refractive index.
    (e) at the incident side of the boundary.

14. If a monochromatic plane wave is incident on a plane boundary hitting it from the high-refractive index side under an angle ϑ_1 (see Fig. 2.5 on Page 17), which is smaller than the angle of total internal reflection ϑ_{1T}, then
    (a) a transmitted wave exists which increases exponentially.
    (b) a transmitted wave exists which decreases exponentially.
    (c) a transmitted wave exists which oscillates in space.
    (d) a reflected wave exists which decays exponentially.
    (e) a reflected wave exists which oscillates in space.
    (f) a reflected wave exists which is constant in space.

2.2 Principles of waveguiding

Summary

1. A wave is guided if the field propagates along a direction while being laterally confined to a certain spatial region.
2. Wave guidance needs an inhomogeneity of the medium.
3. In a lossless straight waveguide consisting of a high-index core and a low-index cladding, power may be transported along the waveguide axis.
4. Therefore the phase surfaces of these guided fields must be perpendicularly to the waveguide axis.
5. A diffraction-caused divergence of light is counterbalanced by the reduced phase velocity in the core region as compared to the phase velocity in the cladding.
6. This functions much as with a lens where the rays through the centre of the lens propagate a longer geometrical distance than the peripheral rays.
7. In a general, light paths in an inhomogeneous medium are always bent towards the optically denser medium with the larger refractive index.

Problems

1. During sunset and sunrise the sun looks slightly oval, as if being compressed in the vertical direction. Why?
2. Radio waves at 300 . . . 3000 MHz may be transported in an atmospheric waveguide, a so-called duct. Why? *Hint*: The temperature high up in the troposphere may be temporarily larger (inverted) than near the earth surface due to infrared radiation during clear nights.
3. During summer time, the air flickers over a sun-heated stretch of tar. — In looking at stars you see them twinkling. What is the reason for these phenomena? *Hint*: The density of the air influences the refractive index.

Quiz

1. Light is confined within the core of a simple clad optical fibre by
   (a) refraction.
   (b) total internal reflection at the outer edge of the cladding.
   (c) total internal reflection at the core-cladding boundary.
   (d) reflections from the fibre's plastic coating.
2. Guided light concentrates most of its power in
   (a) the plastic coating.
   (b) the fibre cladding.
   (c) the fibre core.
   (d) air.
3. Light is guided because of
   (a) medium inhomogeneities.
   (b) a focussing action counteracting diffraction.
   (c) its potential energy is maximum where the refractive index is minimum.
   (d) its potential energy is minimum where the refractive index is minimum.
   (e) its potential energy is minimum where the refractive index is maximum.
4. In a lossless optical waveguide, the phase surfaces of the fields are
   (a) spherical.
   (b) cylindrical.
   (c) plane.
   (d) more complicated.

2.3 Slab waveguide

Summary

1. A symmetric slab waveguide consists of a high-index layer (n_1, core) extending vertically in the region $h/2 \leq x \leq +h/2$. The slab is infinitely extended in y and z-direction. It is sandwiched between two half-infinitely extended cladding layers with an index $n_2 < n_1$ for $|x| > h/2$, Fig. 2.7 on Page 20.

2. Light rays are guided through total internal reflection. Weak guidance means that the refractive indices of core and cladding differ only by a few percent ($0.1 \ldots 2\,\%$).

3. In air, the maximum exit angle for a guided ray defines the numerical aperture $A_N = \sqrt{n_1^2 - n_2^2} = n_1\sqrt{2\Delta}$. The formula for the relative refractive index difference $\Delta = (n_1^2 - n_2^2)/(2n_1^2) \approx (n_1 - n_2)/n_1$ may be simplified for weak guidance, Eq. (2.36) on Page 21.

4. A light ray in a slab waveguide can be represented by a monochromatic plane wave $\exp\left[\mathrm{j}(\omega t - \vec{k}\cdot\vec{r}\,)\right]$ with propagation vector
$$\vec{k} = k_x\,\vec{e}_x + k_z\,\vec{e}_z = k_x\,\vec{e}_x + \beta\,\vec{e}_z.$$

5. A mode is defined to be a transverse field distribution which does not change its shape when propagating along z. It is called an eigensolution of the problem.

6. For a mode to develop the standing waves originating from the upper and the lower boundary must match.

7. This leads to the eigenvalue equation
$$-2k_{1x}h + 2\varphi_p = -2m\pi \;:$$

   (a) A plane wave $\exp\left[\mathrm{j}(\omega t - k_{1x}x)\right]$ traveling from the lower boundary to the upper one and back again accumulates a phase of two times $-k_{1x}h$.

   (b) Taking the polarization-dependent phase of the reflection coefficient into account (see Eq. (2.34 on Page 2.34), two times the phase φ_p must be added.

   (c) The total phase must equal an integer number m of -2π.

8. The eigenvalue equation (also called dispersion relation) may be formulated in terms of the following normalized quantities (Eq. (2.38) on Page 2.38):

   (a) The core phase constant (German *transversales Phasenmaß*)
   $$u = k_{1x}\tfrac{h}{2} = \tfrac{h}{2}\sqrt{k_1^2 - \beta^2}$$

   (b) The transverse cladding attenuation (German *transversales Dämpfungsmaß*)
   $$w = |k_{2x}|\tfrac{h}{2} = \tfrac{h}{2}\sqrt{\beta^2 - k_2^2}$$

   (c) The normalized frequency (waveguide parameter)
   $$V = \tfrac{h}{2}k_0 A_N = \sqrt{u^2 + w^2}$$

   (d) The normalized propagation constant (capital β; Latin B if you like)
   $$B = (\beta^2 - k_2^2)/(k_1^2 - k_2^2) = w^2/V^2$$
   $$= 1 - u^2/V^2 \approx (\beta - k_2)/(k_1 - k_2)$$

9. In a simplified form and for symmetric (below left) and antisymmetric (below right) TE-waves only (E-polarization, H-wave) the dispersion equation reads:
$$w = u\tan u \qquad \text{and} \qquad w = -u\cot u$$

10. This transcendental equation has to be solved graphically (Fig. 2.9(a) on Page 23) or numerically, and leads to discrete eigenvalues u (or w, or B, or β) for a given normalized frequency V.

11. The graph Fig. 2.9(b) on Page 23 that displays the normalized propagation constant B vs. the normalized frequency V is dubbed "dispersion diagram". It corresponds to the W-k_μ diagrams of solid-state physics, e. g., Fig. 3.7 on Page 59: $V \sim \omega \sim \hbar\omega = W_{\text{photon}}$, $B \sim \beta = k_{\mu\,\text{photon}}$. The only difference is the rotation of the axes by $90\,^\circ$.

12. Weak guidance $n_1 \approx n_2$ makes the differences in polarization disappear in Eq. (2.40) on Page 22, $\varphi_E \approx \varphi_H$. This justifies a scalar-optics approach. Then, the eigenvalues β for E- and H-polarization become virtually identical, and the modes are degenerate with respect to polarization.

13. The transverse modes are symmetric or antisymmetric sinusoidals with exponential tails added at the core-cladding boundaries, Eq. (2.50) on Page 26.

14. The modal index m counts the number of field extrema minus 1 (mode $m = 2$ has 3 extrema, see Fig. 2.8 on Page 21.

15. At cutoff, the relations $w = 0$, $B = 0$ and $\beta = k_2$ hold, i. e., for $V_{mG} = u_m = m\pi/2$, mode m propagates as a plane wave in a virtually homogeneous and infinitely extended medium where the finite-sized core represents only a negligible perturbation. Because the field does not experience total internal reflection, there is no spatial confinement and consequently no waveguiding.

16. Only the fundamental mode $m = 0$ can propagate at arbitrary frequencies, including $V = 0$. All other modes $m \geq 1$ experience a lower cutoff where the field extends all over the cross section $-\infty < x < +\infty$. This sets a practical limit to the single-mode operation range $V_{0G} < V < V_{1G}$, because an infinitely extended cladding cannot be be realized technically.

17. For a given normalized frequency V one could try exciting a mode with a plane wave launched at an angle $\vartheta_1 < \vartheta_{1T}$ chosen such that $k_1 \cos\vartheta_1$ does not equal the eigenvalue $\beta(V)$ of the guide. In this case the standing waves originating from the upper and lower boundary do not match. The field cannot escape laterally, but it cannot propagate either because the upper and lower boundary reflections interfere destructively on average. As a consequence, the field decays exponentially along z and therefore is reflected back to the input.

18. The number $M_g = \frac{4}{\pi}V$ of propagating guided modes in this two-dimensional structure is in proportion to V.

19. The group delay for a propagation length L as discussed in Eq. (2.22d) on Page 13 is defined by the derivative of the propagation constant with respect to angular frequency, Eq. (2.54) on Page 27:
$$t_g/L = \mathrm{d}\beta/\,\mathrm{d}\omega$$

20. The term "intramodal dispersion" applies for the group delay difference of different-wavelength signals inside the same modal structure m.

21. The term "intermodal dispersion" or simply "modal dispersion" refers to the group delay differences of different modes m and $m + \Delta m$.

22. Polarization mode dispersion (PMD) refers to slight group delay differences in orthogonally polarized waveguide modes.

23. At bends, the originally plane modal phasefront bends (because at some point the local speed of light is exceeded), and power spills out. All bent waveguides are leaky.

24. The fields of neighbouring waveguides may couple. Energy is transferred periodically along the coupling zone. This may be described by the superposition of even and odd supermodes of the double-waveguide structure.

25. Waveguide branches — as opposed to directional couplers — are optimized for operation in one direction only.

Problems

1. Consider a glass slab with refractive index 1.5 surrounded by air with index 1.0003. What is the angle of total internal reflection ϑ_T?

2. For the above example, what is the numerical aperture A_N? What meaning could you associate with the angle γ_N from the definition $\sin\gamma_N = A_N$?

3. What happens to fields which are excited by a plane wave under not their proper angle, but under an angle which is still smaller than the critical angle of total internal reflection?

4. Formulate the dispersion relation Eq. (2.41) on Page 22 in words! — Why is it called a "transverse resonance condition"?

5. When do you need to discriminate between the polarization of modes?

6. What is the cutoff frequency $f_{1G} = c/\lambda_{1G}$ of a singlemode slab with core index 1.5, cladding index 1.495 and height $h = 5.7\,\mu$m?
7. How many modes does a weakly guiding symmetric slab waveguide carry (core index 1.5, cladding index 1.495, height $h = 57\,\mu$m) at an operating frequency of 193.4 THz?
8. What is the maximum group delay in the multimode waveguide defined above?
9. For a singlemode waveguide, assume a chromatic dispersion of $C = 17\,\text{ps}/(\text{km nm})$ and a spectral width of an impulse-modulated source of $\Delta\lambda = 1\,$nm ($\Delta f = 125\,$GHz at $\lambda = 1.55\,\mu$m). What is the maximum relative group delay difference of the signal's spectral components?
10. Explain the terms "intramodal dispersion", "intermodal dispersion", "modal dispersion", and "polarization mode dispersion"!
11. How is transmission speed f_t (clock rate) connected to group delay dispersion Δt_g?
12. How are temporal and spectral widths of a signal interrelated?
13. Compute the bandwidth of a Gaussian low pass, if the input (output) impulse is a Gaussian with a width of $\Delta t_{H\,\text{in}} = 20\,$ps ($\Delta t_{H\,\text{out}} = 40\,$ps)!
14. Why are waveguide bends lossy?
15. How can the coupling ratio of a directional coupler be influenced?
16. Why are waveguide Y-branches like the ones on Page 32 optimized for operation in one direction only?

Quiz

1. What is the half-acceptance angle for a symmetric step-index slab waveguide with core and cladding indeces 1.5 and 1.495?
   (a) 4.7°
   (b) 7.0°
   (c) 9.4°
   (d) 11°
2. What is the maximum allowable core height for a symmetric step-index singlemode slab operating at 1.3 μm, with core and cladding indices 1.5 and 1.0003 (air)?
   (a) 0.58 μm
   (b) 0.89 μm
   (c) 3.0 μm
   (d) 4.8 μm
3. The transverse modes in a symmetric slab waveguide are
   (a) sin-functions.
   (b) cos-functions.
   (c) sin- or cos-functions.
   (d) exponentials.
   (e) sin- or cos-functions with exponential tails.
4. The modal index m counts the number of
   (a) guided modes.
   (b) radiated modes.
   (c) zeros of the transverse field.
   (d) maxima of the transverse field.
   (e) extrema of the transverse field.
   (f) maxima of the transverse intensity minus 1.
   (g) extrema of the transverse intensity.
5. At its cutoff frequency, a mode is
   (a) strongly guided.
   (b) weakly guided.
   (c) not guided.
6. The fundamental mode of a symmetric slab waveguide has a
   (a) theoretically non-zero cutoff frequency.
   (b) virtually non-zero cutoff frequency.
   (c) normalized cutoff frequency of $\pi/2$.
7. Modes in a bent waveguide are
   (a) guided.
   (b) radiating.
   (c) radiating if the bend is too large.
   (d) lossy.
   (e) absorbed.
   (f) leaky.

2.4 Strip waveguide

Summary

1. Strip waveguides trap the light in two spatial directions, vertically as in a slab waveguide, and in addition laterally.
2. Consequentially, there are two modal indices m and n in the vertical and in the lateral direction, Fig. 2.15 on Page 33.
3. Both polarizations have longitudinal E- and H-components, and are therefore called hybrid. For EH_{mn}-modes, the longitudinal H-components is less pronounced than the longitudinal E-component, for HE_{mn}-modes it is just the other way round, Fig. 2.16(a) on Page 33.

Problems

1. How many intensity maxima has an EH_{42}-mode, how many has an HE_{00}-mode?
2. For a square "strip" waveguide, how do the EH_{23}-mode and the HE_{23}-mode differ?

Quiz

1. An E_1-mode in a slab waveguide has a
   (a) longitudinal H-component.
   (b) longitudinal E-component.
   (c) no longitudinal components.
2. An EH_{11}-mode in a strip waveguide has a
   (a) longitudinal H-component.
   (b) longitudinal E-component.
   (c) no longitudinal components.

2.5 Fibre waveguides

Summary

1. There are several different types of optical fibres, with distinct properties.
2. Fibre properties depend on the core-cladding structure and the material from which the fibre is made.
3. To guide light, a fibre must have a core with higher refractive index than the cladding.
4. Optical fibres — like all dielectric waveguides — carry light mostly inside the core, with some light in the cladding.
5. Strictly speaking, the fibre fields have to be calculated fully vectorial. For weakly guiding fibres $\Delta \ll 1$ this requirement may be relaxed. Proper superposition of degenerate vector modes with virtually equal phase and group delays leads to $\mathrm{LP}_{\nu\mu}$-modes, which are much easier to handle, Fig. 2.19 on Page 36.
6. However, because fibres are never perfectly circular symmetric, orthogonal polarizations propagate with slightly different group velocities, Eq. (2.67) on Page 37.
7. Both polarizations do couple in real fibres, and hence the group delay differences do not increase linearly with transmission length L, but only in proportion of its square root $\sqrt{L}$, Eq. (2.76) on Page 43.
8. Radial and azimuthal resonance conditions with associated modal numbers $\mu = 1, 2, 3, \ldots$ and $\nu = 0, 1, 2, \ldots$ define the dispersion relation and the dispersion diagram for the $\mathrm{LP}_{\nu\mu}$-modes, Eq. (2.20) on Page 39.
9. Grading the refractive-index differential between core and cladding can strongly reduce modal dispersion in a multimode fibre, Fig. 2.18 on Page 35.
10. Graded-index fibres have standard core diameters of $2a = 50\,\mu$m or $62.5\,\mu$m. They are used for transmission over distances up to a couple of kilometers.
11. Fibres with core diameters only $6 \ldots 10\,\mu$m transmit a single mode of light.
12. The cutoff wavelength is the shortest wavelength at which a singlemode fibre transmits only one mode.

13. For step-index fibres this corresponds to a maximum normalized frequency $V = ak_0A_N$ of $V_{01G} = 2.405$, Fig. 2.20 on Page 39.

14. The number of guided modes $M_g = V^2/2$ carried by a step-index fibre depends on its normalized frequency $V = ak_0A_N$ squared, i. e., on the fibre core radius a, the relative refractive index difference Δ between core and cladding defining a numerical aperture $A_N = n_1\sqrt{2\Delta}$, and on the operating (angular) frequency $\omega/c = k_0$.

15. The number of guided modes for graded-index fibres with the same V is less by a factor of two, $M_g = V^2/4$.

16. Singlemode fibres are used for high-speed communications over distances of more than a couple of kilometers. They may be also used over shorter distances.

17. Chromatic dispersion is the sum of material dispersion and waveguide dispersion, Eq. (2.55) on Page 27. It is the main type of dispersion in singlemode fibre, Table 2.1 on Page 40.

18. Dispersion-shifted fibre has waveguide dispersion increased by an increasead Δ, so waveguide dispersion cancels material dispersion at a wavelength generally longer than $1.3\,\mu$m, Table 2.2 on Page 40. Types now in use have dispersion shifted to wavelengths shorter or longer than the Erbium-fibre amplifier band, usually to about $1.5\,\mu$m (S band) or about $1.64\,\mu$m (L+ band), see Table 1.1 on Page 5.

19. Dispersion-compensating fibres have even higher-doped cores (larger Δ), Table 2.3 on Page 41, and therefore suffer from larger loss, Eq. (2.24) on Page 13. A figure of merit relates the chromatic dispersion coefficient $|C|$ to the attenuation a/L_{DCF} in dB / km.

20. Dispersion flattened fibres with more complicated profiles (e. g., a triple-clad step-index fibre), Fig. 2.22 on Page 42, allow for an adjustment of the dispersion slope D of $C_\lambda(\lambda) = C + D\,\Delta\lambda$.

21. Polarization mode dispersion (PMD), Table 2.4 on Page 43, becomes important for optical fibres where short light impulses (e. g., $\Delta t_H = 17\,$ps for a bit rate of 40 Gbit/s, bit period $T_t = 1/40\,$GHz = 25 ps) travel over long distances.

22. Modes in a parabolic-index fibre decay like a Gaussian along the radius and have a modal structure determined by Laguerre polynomials, Fig. 2.23 on Page 44. As opposed to step-index fibres, the modal fields do not fill the core area fully. Especially the lowest order fundamental LP_{01} mode is strongly concentrated on the fibre axis.

23. The modes of a straight perfect waveguide are orthogonal. This property allows the calulation of mode power coupling coefficients, Eq. (2.81) on Page 45.

Problems

1. Consider a typical multimode fibre with $n_1 = 1.5$ and $\Delta = 1\,\%$. For a step-index fibre, what is the achievable bit rate for a transmission length of $L = 400\,$m?

2. With the same data as above, what is the maximum bit rate for a graded-index fibre with a parabolic profile?

3. Why does polarization mode dispersion not increase linearly with transmission length L, but only in proportion to its square root, $\Delta t_g \sim \sqrt{L}$? *Hint 1*: Consider going to the theater either walking (arrival after $t_{g\,\mathrm{slow}} = 60\,$min), or by taxi (arrival after $t_{g\,\mathrm{fast}} = 10\,$min). For a random initial choice, the expected effective difference in arrival times would be $\Delta t_g = 1\,$h. If you would live double as far from the theater, the arrival time difference would be obviously $\Delta t_g = 2\,$h. But what do you expect if you change your transportation randomly when being on your way? *Hint 2*: Do not despair. Not all of the relevant information has been provided,[9] just read the solution on Page 171.

4. At a bit rate of 40 Gbit/s the bit period is 25 ps and the half-power light impulse

[9] W. Freude: Field propagation and coherence (FPC, formerly Ausbreitung und Kohärenz optischer Felder). See Sect. 1.7.4 „Modenkopplungsmodell (Mode coupling model)“. Compuscript for download at `http://www.ipq.kit.edu`.

width amounts to $\Delta t_H = 17\,\text{ps}$. For a typical fibre the polarization mode dispersion is $D_p = 0.1\,\text{ps}/\sqrt{\text{km}}$. What is the maximum transmission length L if a group delay spread of $\Delta t_{gp} = 1\,\text{ps}$ seems to be permissible?

5. Reducing dispersion slope (Eq. (2.56) on Page 27) tends to decrease the mode field diameter. Why could that be detrimental?

6. Why is it more difficult to compensate for dispersion in a DWDM system than in one transmitting only a single optical channel?

7. Write a formula for how much dispersion-compensating fibre L_{DCF} you need to add to an existing system with L_{exist} of fibre to reduce dispersion to a desired value C_{net}!

8. Nonlinear crosstalk depends on the power density in the fibre core. If you have a fibre with effective mode area $A_{\text{eff}} = 50\,\mu\text{m}^2$, what is the intensity in the fibre if it carries 100 optical channels at 1 mW each?

9. Why are parabolic-index fibres of practical interest?

10. What does the orthogonality of modes mean?

Quiz

1. You transmit a δ-shaped impulse through a 20 km long multimode fibre with a total dispersion of 10 ns / km. Disregard chromatic dispersion. What will the impulse length be at the end?

   (a) 200 ns
   (b) 100 ns
   (c) 50 ns
   (d) 20 ns
   (e) 10 ns

2. You transmit a 100 ns wide impulse through the same fibre as above. What will the impulse length be at the end?

   (a) 300 ns
   (b) 224 ns
   (c) 200 ns
   (d) 150 ns
   (e) 100 ns

3. A singlemode fibre has material dispersion 20 ps /(km nm) and waveguide dispersion of −15 ps /(km nm) at the signal wavelength. What is the total chromatic dispersion?

   (a) 35 ps /(km nm)
   (b) 25 ps /(km nm)
   (c) 5 ps /(km nm)
   (d) 0 ps /(km nm)
   (e) −35 ps /(km nm)

4. You send 100 ps wide impulses through a 100 km length of the above fibre, using a laser with a spectral width of 0.002 nm as measured without modulation at its emission wavelength 1.55 μm. What is the width of the output impulse? *Hint*: What is the spectral width of a 100 ps wide impulse?

   (a) 1 ps
   (b) 108 ps
   (c) 125 ps
   (d) 200 ps
   (e) 250 ps

5. You have to send a signal at the highest data rate possible through 2 500 km of singlemode fibre with polarization mode dispersion of $1\,\text{ps}/\sqrt{\text{km}}$. Neglecting other types of dispersion, what is the best you can do, if PMD should accumulate no more than one tenths of the bit period?

   (a) 10 Gbit/s
   (b) 5 Gbit/s
   (c) 2 Gbit/s
   (d) 1 Gbit/s
   (e) 100 Mbit/s

6. Suppose you only had to transmit signals 400 km through the same fibre. What is the maximum data rate under the same assumption as above?

   (a) 10 Gbit/s
   (b) 5 Gbit/s
   (c) 2 Gbit/s
   (d) 1 Gbit/s
   (e) 100 Mbit/s

3 Light sources

Summary

1. A laser oscillator consists of an active medium that is capable of providing optical amplification. Placed around the amplifying medium there is an optical resonator that provides the necessary optical feedback.

2. The optical resonator consists of a transverse strip waveguide with height d and width b, and two plane mirrors with power reflection coefficients $R_{1,2}$ at $z = 0, L$, see Fig. 3.1 on Page 47.

3. In addition to the two *transverse* field consistency conditions of the strip, $2k_{x,y} \times L_{x,y} = m_{x,y} \times 2\pi$, a third *longitudinal* resonance condition $2k_z L_z = m_z \cdot 2\pi$ fits the modal phase along the z-axis, Eq. (3.1) on Page 48.

4. The number of resonant modes ("states") for this three-dimensional arrangement is in proportion of f^3, the density of states (DOS) per frequency interval is $\sim f^2$.

5. Specific corrections apply for oddly-shaped or singlemoded volumes. Semiconductor laser resonators are usually transversely singlemoded!

Problems

1. Estimate the number of modes for a GaAs laser resonator with $d = 100\,\text{nm}$, $b = 3\,\mu\text{m}$, $L = 400\,\mu\text{m}$, an amplification bandwidth $\Delta f = 12\,\text{THz}$ (see Eq. (3.66) on Page 78) and a refractive index of $n \approx n_g = 3.6$ at $\lambda = 0.85\,\mu\text{m}$! *Hint*: Count the number of modes by using Eq. (3.2) on Page 48.

2. Try using also Eqs. (3.7) and (3.4) on Page 49 and compare to the result above. What is the reason for the discrepancies?

3.1 Luminescence and laser radiation

Summary

1. A microsystem may release or acquire energy by a transition to a lower or higher energy state, respectively, Fig. 3.2 on Page 50.

2. The processes involved are (induced) absorption, spontaneous emission, and induced emission.

3. Induced transitions are enforced by already existing photons. The more there are, the more likely is the transition.

4. Spontaneous emission is a random process and adds noise. Spontaneously emitting (noise) sources are incandescent lamps and luminescent diodes (LED).

5. If there are more induced emissions than absorptions, the system is called inverted. This unnatural state has to be enforced by a so-called pump mechanism.

6. For induced emission the emitted radiation is coherent with the stimulating radiation regarding direction, phase and polarization: The optical field becomes amplified much like by an electronic amplifier, if the net gain (induced emission rate) − (absorption rate) is positive.

7. An excited state has a certain natural lifetime τ_{sp} ($0.1 \ldots 1\,\text{ns}$ for semiconductors) describing the rate $1/\tau_{\text{sp}}$ at which excited systems release their energy spontaneously, Eq. (3.8) on Page 51.

8. By changing the pump rate, the inversion may be modified. If the light source has no optical resonator and therefore emits only spontaneous radiation, the number of excited microsystems cannot decrease faster than their spontaneous lifetime τ_{sp}, and the maximum light intensity modulation frequency is in the order of $1/\tau_{\text{sp}}$

9. If the pump rate is changed through the injection current of a semiconductor laser, this is called direct modulation, Page 52.

10. The effective lifetime $\tau_{2\,\text{eff}}$ becomes strongly reduced over τ_{sp} via induced

emission in the presence of an optical field. For a strong field the transition may occur virtually instantaneously.

11. Coherent laser radiation is emitted in one transverse mode only and may be coupled with very high efficience to a singlemode fibre, Eq. (3.9) on Page 52. The emission can be also spectrally narrow ($\Delta f = 40\,\mathrm{MHz}$). Lasers generate more power than LED and are more efficient.

12. Incoherent LED radiation is emitted in many (typically 10^4) modes and experiences therfore losses of around 40 dB when coupled to a singlemode fibre. Coupling efficiency is of the order of 10 dB or better for a multimode fibre. The spectral emission is typically broad ($\Delta f = 12\,\mathrm{THz}$). LED generate less power than lasers.

13. Laser active materials may be pumped optically or — in the case of semiconductors — also electrically.

14. Inversion may occur in a three or four-level system, Fig. 3.3 on Page 54. Only transparency may be achieved in a two-level system.

15. For inversion to occur in a semiconductor, the energy $hf^{(\mathrm{e})}$ of emitted photons must be larger than the bandgap energy W_G, but smaller than the difference of the quasi Fermi levels $W_{Fn} - W_{Fp} = eU$, which are fixed by the external forward voltage U, Eq. (3.13) on Page 54.

16. Compound materials are used to adjust bandgap and lattice constant independently, Fig. 3.4 on Page 55. This is important because strong lattice mismatch of adjacent layers leads to crystal defects and early failure when operating the devices with high current densities of typically more than $500\,\mathrm{A}\,/\,\mathrm{cm}^2$.

17. The most common light sources for short glass fibre systems are GaAlAs LED or semiconductor lasers (LD, laser diodes) operating at $0.82\,\mu$m or $0.85\,\mu$m. InGaAsP lasers emit from $0.9\,\mu$m to $1.7\,\mu$m. Some InGaAsP LED are used at $1.3\,\mu$m.

18. The most common light sources for long-haul glass fibre systems are InGaAsP semiconductor lasers (LD, laser diodes) operating in the minimum-loss region near $1.55\,\mu$m.

Quiz

1. Induced absorption is
   (a) the process underlying optical amplification.
   (b) the process underlying optical detection.
   (c) nonsensical.
   (d) a source of optical noise.
   (e) a reason for optical loss.

2. Induced emission is
   (a) the process underlying optical amplification.
   (b) the process underlying optical detection.
   (c) nonsensical.
   (d) a source of optical noise.
   (e) a reason for optical loss.

3. Spontaneous absorption is
   (a) the process underlying optical amplification.
   (b) the process underlying optical detection.
   (c) nonsensical.
   (d) a source of optical noise.
   (e) a reason for optical loss.

4. Spontaneous emission is
   (a) the process underlying optical amplification.
   (b) the process underlying optical detection.
   (c) nonsensical.
   (d) a source of optical noise.
   (e) a reason for optical loss.

5. The net emission rate is defined as
   (a) the difference between the total optical emission and absorption rates.
   (b) the number of photons emerging at the receiver point of an optical network.

(c) the difference between induced emission and absorption rates.

6. Population inversion

   (a) is impossible with systems in thermal equilibrium.
   (b) may be achieved with a two-level system.
   (c) may be achieved with a three-level system.
   (d) may be achieved with a four-level system.
   (e) may be achieved in semiconductor structures.
   (f) causes difficulties in paying pensions.
   (g) is natural with systems in thermal equilibrium.
   (h) is something odd in relation with sexual preferences.
   (i) needs a pump mechanism which transfers energy to the system.

7. The lifetime of an excited state

   (a) is a material property.
   (b) may be influenced from outside.
   (c) is basically infinite.
   (d) should be large so that maintenance cost remain reasonable.

8. Laser radiation is

   (a) more powerful than LED radiation.
   (b) possibly harmful for the eye.
   (c) spectrally narrow.
   (d) spatially widely divergent.
   (e) not always visible.
   (f) efficiently coupled to singlemode fibres.
   (g) efficiently coupled to multimode fibres.

9. LED radiation is

   (a) less powerful than LD radiation.
   (b) harmless for the eye.
   (c) spectrally narrow.
   (d) spatially widely divergent.
   (e) always visible.
   (f) efficiently coupled to singlemode fibres.
   (g) efficiently coupled to multimode fibres.

10. In a semiconductor pn-diode laser, population inversion occurs if

    (a) the energy $hf^{(e)}$ of emitted photons is larger than the bandgap energy W_G.
    (b) the energy $hf^{(e)}$ of emitted photons is smaller than the difference of the quasi Fermi levels $W_{Fn} - W_{Fp}$.
    (c) the energy equivalent eU of junction voltage U is larger than the energy $hf^{(e)}$ of emitted photons.
    (d) a sufficiently large forward current is applied.
    (e) a sufficiently large reverse current is applied.

11. Compound materials are used

    (a) for adjusting the bandgap.
    (b) for adjusting the lattice constant.
    (c) for adjusting bandgap and lattice constant independently.
    (d) for creating direct semiconductors.
    (e) for avoiding the development of crystal defects.
    (f) for an improved carrier lifetime.
    (g) for all of the above reasons.

12. The most common light sources for short glass fibre systems are made of

    (a) InP.
    (b) AlAs.
    (c) GaAlAs.
    (d) InGaAs.
    (e) InGaAsP.

13. The most common light sources for long-haul glass fibre systems are made of

    (a) InP.
    (b) AlAs.
    (c) GaAlAs.
    (d) InGaAs.
    (e) InGaAsP

3.2 Semiconductor physics

Summary

1. Indirect semiconductors like Si or Ge are inefficient light emitters because a third interaction particle must supply the crystal momentum for enabling the recombination of an electron and a hole at the lowest possible transition energy, Fig. 3.7 on Page 59.

2. Si or Ge may be successfully used for photodetectors, because the low indirect-semiconductor absorption probability may be effectively increased by a longer absorption region.

3. Direct semiconductors like GaAs or InP show their lowest-energy transition for the same crystal momentum. They are efficient light emitters.

4. The number of electron states Z with two spin directions and a modulus of the crystal momentum up to $p_\mu = \hbar k_\mu$ (Eq. (3.18 on Page 60) is derived in analogy to Eq. (3.6) on Page 49.

5. The density of states $\rho(W)$ measures — according to Eq. (3.20 on Page 60 — the number of electron states per volume and energy interval.

6. The effective density of states N_B gives the number of states in an energy interval kT from the band edge. The larger N_B, the bigger the effective mass is, Table 3.3 on Page 60.

7. The Fermi-Dirac distribution, Fig. 3.8 on Page 61, governs the filling of states in the conduction and valence bands at equilibrium. It may be controlled by an external junction voltage U.

8. Pure semiconductors would have little use by themselves because of their low conductivity (carrier concentration at room temperature $\sim 10^{11}\,\mathrm{cm}^{-3}$) compared to metals ($\sim 10^{21}\,\mathrm{cm}^{-3}$). By introducing impurities (doping) a semiconductor may become p or n-conducting.

9. For a non-equilibrium condition the carriers in the CB and VB states need some time to re-arrange for achieving an intraband stationary state, which may be also characterized with a Fermi distribution. However, the conduction band electrons are not in equilibrium with the valence band holes, and therefore the so-called quasi Fermi levels of both bands differ, $W_{Fn} \neq W_{Fp}$, Eq. (3.28) on Page 63.

10. Heterostructure diodes confine carriers by potential barriers in a low-bandgap region which can become population-inverted. Because smaller bandgaps go along with a higher refractive index, also the electromagnetic field is confined to this active region enabling a strong interaction, Fig. 3.11–3.14 on Pages 65–68.

11. Spontaneous emission is observed for a wide range of photon energies hf above the bandgap energy, Fig. 3.15(a) on Page 71.

12. Induce net emission (amplification) is to be seen for photon energies hf above the bandgap, but below the difference of the quasi Fermi levels, Fig. 3.15(b). Above this limit stimulated absorption dominates.

13. Net gain and amplification bandwidth increase with increasing carrier density n_T, which are controlled by the injection current, Fig. 3.16 on Page 72.

14. Besides radiative transitions with a rate $\sim n_T p$ there are nonradiative recombinations via impurities in the forbidden band $\sim n_T$, or Auger recombinations in InP with a rate $\sim n_T p^2$, Fig. 3.17 on Page 73. Nonradiative transitions reduce the radiation efficiency.

Quiz

1. Indirect semiconductors

   (a) do not support a direct current.

   (b) need phonons for band-to-band transitions with minimum energy difference.

   (c) can be used for detectors.

   (d) can be used for emitters.

(e) like Si or Ge are widely used.

2. Direct semiconductors

   (a) do support a direct current.
   (b) need phonons for band-to-band transitions with minimum energy difference.
   (c) can be used for detectors.
   (d) can be used for emitters.
   (e) like GaAs or InP are rarely used.

3. The density of states describes

   (a) the concentration of carriers.
   (b) the number of impurities per volume.
   (c) the number of electronic wave functions per volume and energy interval.
   (d) the number of different electronic wave functions per volume and energy interval kT, in the case of the effective density of states N_B.
   (e) — near the band edges — the band curvature.

4. The Fermi-Dirac distribution for band states

   (a) equals 1/2 at the Fermi energy W_F for a temperature $T = 0$.
   (b) equals 1/2 at the Fermi energy W_F for all temperatures.
   (c) is step-like at a temperature $T = 0$.
   (d) gives the occupation probabilities of states in the bands.
   (e) may be approximated by a Boltzmann distribution for energies far away from the Fermi level.
   (f) is described by all points above.

5. The quasi Fermi levels of the conduction and of the valence band

   (a) are the same in thermal equilibrium.
   (b) describe the filling of states in the respective bands.
   (c) are always constant throughout the various layers of a semiconductor.

6. Heterojunctions

   (a) confine carriers.
   (b) have a larger refractive index where the bandgap is larger.
   (c) confine light.
   (d) are essential for achieving a high radiation efficieny.

7. Optical emission from a heterojunction diode is observed

   (a) if the diode is forward-biased.
   (b) at photon energies below the bandgap energy.
   (c) only at photon energies smaller than the difference of the quasi Fermi levels.
   (d) basically at all photon energies larger than the bandgap energy.

8. An induced net emission means

   (a) coherent optical amplification.
   (b) that the induced emission rate is larger than the absorption rate.
   (c) amplified spontaneous emission (ASE) is of no consequence.
   (d) that the energy of the emitted photons is smaller than the bandgap.
   (e) that the energy of the emitted photons is smaller than the difference of the quasi Fermi levels.
   (f) that the carrier concentration provided by the injection current is larger than the transparency level.
   (g) that the carrier concentration provided by the injection current is larger than the laser threshold level.

9. Optical net gain and amplification bandwidth

   (a) depend on the carrier concentration.
   (b) decrease with decreasing injection current.
   (c) are constant because of the constant doping.

10. Band-to-band transitions may occur

    (a) radiatively through Auger recombination.

(b) nonradiatively via recombinations over impurities in the forbidden band.

(c) radiatively by recombinations of CB electrons and VB holes.

11. Nonradiative transistions are

(a) detrimental for the radiation efficiency.

(b) advantageous if carrier lifetime in rapidly modulated LED is a concern.

(c) nonexistent in GaAs compounds.

(d) unavoidable for InP compounds.

3.3 Light-emitting diode

Summary

1. LED emit spontaneously generated light (optical noise), and their dynamics are therefore governed by the lifetime τ_{sp} of spontaneous emission, Eq. (3.63) on Page 76.

2. The rate I/e of carriers entering the active region determines the rate $P/(hf)$ of generated photons (optical power P) with a certain internal efficiency $0.5 \leq \eta_{\mathrm{int}} \leq 0.9$, Eq. (3.58) on Page 75. This reflects the proportion of nonradiative recombinations.

3. The generated power P is partially reflected at the semiconductor-air interface, so the light power P_a is radiated into air with a certain optical efficiency $0.015 \leq \eta_{\mathrm{opt}} \leq 0.035$ only, Eq. (3.60).

4. Finally, because of parasitic contact resistances, the external electrical power $P_{\mathrm{el}} = UI$ is not fully available for inverting the medium, and there is a voltage drop between the external contacts and the internal junction. Thus the power P is only a percentage η_{el} of P_{el}.

5. The total wall-plug efficiency for $eU \approx hf$ is $\eta_{\mathrm{tot}} = \frac{P_a/(hf)}{P_{\mathrm{el}}/(eU)} = 0.75 \ldots 3.1\,\%$. With operating currents up to 200 mA the emitted power reaches the 10 mW range.

6. For surface emitters, Fig. 3.19 on Page 77, the emitting area of the junction is confined by oxide isolation, and the contact is usually $15 \ldots 100\,\mu$m in diameter.

7. Edge emitters, Fig. 3.20 on Page 77, have a very thin active layer, $d = 0.05 \ldots 0.1\,\mu$m and emit — because of the strip waveguide structure — radiation in one single transverse mode.

8. The LED spectrum has a width of about $\Delta f_H = 12.1\,\mathrm{THz}$, Eq. (3.66) on Page 78.

Problems

1. Assume an internal LED efficiency of $\eta_{\mathrm{int}} = 0.7$ at a wavelength of $0.6\,\mu$m ($h = 6.6 \times 10^{-34}\,\mathrm{W\,s^2}$) and an injection current of $I = 50\,\mathrm{mA}$ ($e = 1.6 \times 10^{-19}\,\mathrm{A\,s}$) What is the internal optical power P? What is the number of photons in a time interval of 25 ps?

2. With an optical efficieny $\eta_{\mathrm{opt}} = 0.025$, what is the power P_a that may be actually used?

Quiz

1. Operating wavelengths of GaAs LED and lasers are

   (a) $0.82\,\mu$m and $0.85\,\mu$m
   (b) $0.665\,\mu$m
   (c) $1.3\,\mu$m
   (d) $1.55\,\mu$m
   (e) none of the above

2. Light emission from an LED is modulated by

   (a) voltage applied across the diode.
   (b) current passing through the diode.
   (c) illumination of the diode.
   (d) all of the above.

3. Surface emitters are suitable for

   (a) operating a fibre gyro made of a singlemode-fibre Sagnac loop.
   (b) inexpensive indicator lamps.
   (c) easy coupling to multimode fibres.

4. Edge emitters are suitable for

(a) operating a fibre gyro made of a singlemode-fibre Sagnac loop.

(b) inexpensive indicator lamps.

(c) easy coupling to multimode fibres.

5. The emission frequency bandwidth of a LED depends

(a) heavily on the emission wavelength.

(b) on the curvature of the valence band.

(c) on the energetic position of the quasi Fermi levels.

(d) on temperature.

(e) on all of the above.

3.4 Laser diode

Summary

1. Basic laser diodes (LD) with a rectangular cavity are equivalent to a Fabry-Perot (FP) resonator, Fig. 3.1 on Page 47 and Fig. 3.6 on Page 58, and are thus called Fabry-Perot laser diodes (FP LD).

2. Part of the field is propagating outside the active region. The percentage of power concentrated inside the active region is given by the field confinement factor Γ, Eq. (3.69) on Page 79.

3. LD are usually laterally singlemoded, but they have longitudinal resonances spaced $\Delta f_z = \frac{c}{2n_g L} = \frac{v_g}{2L} = \frac{1}{\tau_U}$ apart. The quantity τ_U is called photon round-trip time, Eq. (3.68) on Page 79.

4. The longitudinal guided-wave propagation is described by equivalent plane waves $\exp(\mathrm{j}\,\omega_0 t)\exp(-\mathrm{j}\,\bar{k}z)$ with an (effective) complex refractive index $\bar{n} = n - \mathrm{j}\,n_i$ and an (effective) complex propagation constant $\bar{k} = k_0\bar{n} = k + \frac{1}{2}\mathrm{j}\,(g - \alpha_V)$, Eq. (3.73) on Page 80.

5. The optical power P, the electric field E and the photon number N_P of the field inside the resonator are related by $P(t) \sim |E(t)|^2 \sim N_P(t)$, see Eq. (3.94) on Page 85.

6. The quantities g, α_V are the modal power gain and loss constants corresponding to the net effective photon gain rate ΓG and a photon loss rate $1/\tau_P$ specified by the photon lifetime τ_P.

7. This lifetime τ_P is basically limited by the transmission loss through partially-transparent mirrors. A lifetime $\tau_P \to \infty$ makes no sense, because we want to extract some radiation!

8. Net photon gain rate G and modal power gain g are related through the group velocity, $G = v_g g$. The gain rate $G(n_T)$ increases with the carrier concentration n_T and therefore also with the injection current I.

9. At the transparency carrier concentration n_t the net gain rate G becomes zero. The material is transparent (no gain, no loss).

10. The difference of net effective gain rate and loss rate specifies the net increase in relative photon number per second, $G - \frac{1}{\tau_P} = \frac{1}{N_P}\frac{\mathrm{d}N_P}{\mathrm{d}t}$.

11. If $G(n_{TS}) - 1/\tau_P = 0$, i. e., if the gain rate G just compensates the loss rate $1/\tau_P$, the threshold carrier density n_{TS} is reached (threshold current I_S).

12. For $G(n_T) - 1/\tau_P > 0$ the photon number increases with time, i. e., the electric field would start oscillating according to $E(t) \sim \sqrt{N_P(t)}\exp(\mathrm{j}\,\omega_0 t)$ at the frequency f_0 of that longitudinal mode which is closest to the spectral gain maximum which can be seen in Fig. 3.16 on Page 72.

13. If the gain rate would remain constant, the photon number would grow exponentially according to $N_P(t) = N_P(0)\exp\left[\left(G - 1/\tau_P\right)t\right]$, Eq. (3.74) on Page 80. The photon number $N_P(0)$ at time $t = 0$ is provided by spontaneously emitted photons, i. e., by noise.

14. With increasing photon number the gain rate $G(n_T > n_{TS})$ becomes clamped at the threshold level $G(n_{TS})$, because all carriers in excess of the threshold concentration n_{TS} recombine immediately.

15. The gain rate reduces further by additional mechanisms like hot carrier effects

and spectral hole burning. This quickly reacting gain compression $G(n_T, N_P)$ for stronger optical field is modelled phenomenologically by Eq. (3.80) on Page 81.

16. The dynamics of photon number and carrier concentration are formulated heuristically in Eq. (3.83) on Page 82 (and may be derived from first principles also).

17. Because of the interaction of two energy reservoirs (charge carriers and photons, like capacitor and inductor in a tank circuit), resonance phenomena are to be expected.

18. The lasing threshold current $I_S/e = n_{TS}V/\tau_{\text{eff}}$ is defined at $\Gamma G(n_{TS}, 0) = \Gamma G_S = 1/\tau_P$ where for this carrier concentration n_{TS} the gain rate ΓG_S just compensates the loss rate $1/\tau_P$. Here, spontaneous emission is neglected, Eq. (3.85) on Page 83, and the field confinement factor Γ is taken into account (Summary 2 above).

19. The threshold current density J_S increases with the thickness d of the active region for maintaining a fixed n_{TS}, but also decreases in proportion to $1/d$ because the field becomes more confined to the active region (Eq. (3.70) on Page 79), which makes the gain rate ΓG larger. An optimum layer thickness $d \approx 100\,\text{nm}$ then leads to a minimum threshold current density J_S, Eq. (3.88) on Page 84.

20. In normalized form and neglecting spontaneous emission, the static rate equations are represented by straight lines, Fig. 3.21 on Page 85. These characteristics are "softened" when spontaneous emission is taken into account.

21. The power stored inside the laser resonator is given by the total photon energy $N_P hf$ per photon lifetime τ_P. The sum power emitted through both mirrors is the total stored photon energy $N_P hf$ per photon lifetime τ_R associated with the average mirror transmittivities.

22. The small-signal response $N_{P1}(\omega)$ of the photon number with respect to an injection current perturbation $I_1(\omega)$ depends an the (angular) frequency ω by which the current is modulated. The transfer function $N_{P1}(\omega)/N_{P1}(0)$ exhibits a resonance characteristic near the small-signal resonance angular frequency ω_R, see Fig. 3.22 on Page 88. The strength of the resonance depends on the static injection current I_0 which defines the operating point.

23. For the large-signal case strong relaxation oscillations are to be seen, Fig. 3.24 on Page 89.

24. The switch-on delay, i. e., the charging of the junction's diffusion capacity, may be overcome by biasing the device slightly above threshold. This is the usual mode of operation for directly modulated laser diodes.

25. The carrier concentration n_T inside the active region influences its refractive index n: The larger n_T grows, the smaller n becomes, Fig. 2.1(a) on Page 9. This is a consequence of the causality requirement.

26. From Summary 3 on Page 149 or Eq. (3.67) on Page 79 the product of oscillation frequency and refractive index must be constant, $\mathrm{d}(\omega n) = 0$.

27. If n_T increases, the gain constant g (and therefore the negative imaginary part of the refractive index n_i) will grow, and because of causality n becomes smaller, and ω increases.

28. This causes a chirp of the oscillation frequency of a directly modulated laser diode, and a broadening of the spectrum, Eq. (3.129) on Page 93.

29. The ratio of $\partial n/\partial n_T$ and $\partial n_i/\partial n_T$ defines the quantity α in Eq. (3.123) on Page 92 (α-factor, line broadening factor, Henry factor).

30. Semiconductor lasers may be fine-tuned in oscillation frequency by modifying the injection current. A coarser tuning is possible by changing the temperature.

31. Semiconductor lasers with external mirrors may be tuned over a wide wavelength range, e. g., 120 nm. They can also be designed to generate a series of very short impulses.

32. Gain-guided lasers are cheaply produced but are longitudinally multimoded, Fig. 3.27(a) on Page 94.

33. Index-guided lasers are technologically more demanding but perform better, Fig. 3.27(b).

34. Stronger spectral selectivity can be achieved if either the cleaved facets are replaced by distributed Bragg mirrors (DBR laser), Fig. 3.29(a) on Page 96, or if the feedback is distributed all over the length of the active region (DFB laser), Fig. 3.29(b).

35. Vertical cavity surface emitting lasers (VCSEL) emit perpendicularly to their p-n junction plane like surface-emitting LED, Fig. 3.31 on Page 98. Their structure is that of a DBR laser. Because of their low threshold in the sub-mA region they may be easily modulated directly, even without a bias current.

Problems

1. Formulate the laser diode rate equations!

2. Below threshold a laser diode emits some light by spontaneous emission. Why does it behave like an LED?

3. A diode laser has a threshold current of $I_S = 12\,\mathrm{mA}$. That current passes through a strip $b = 5\,\mu\mathrm{m}$ wide and $L = 500\,\mu\mathrm{m}$ long. What is the threshold current density J_S?

4. The laser diode in the above problem emits $P_a/2 = 5\,\mathrm{mW}$ of light. The junction layer is $d = 0.5\,\mu\mathrm{m}$ thick. If the light is evenly distributed across the end of the active layer, what is the optical power intensity I in this cross-section?

5. InGaAsP has a refractive index of about $n = 3.5$. A Fabry-Perot laser emitting multiple longitudial modes at a nominal wavelength of $\lambda = 1.55\,\mu\mathrm{m}$ has a cavity $L = 430\,\mu\mathrm{m}$ long. What is the wavelength difference $\Delta\lambda_z$ between two adjacent longitudinal modes?

6. Suppose the laser in the above problem was a VCSEL with cavity length only $10\,\mu\mathrm{m}$, and for the time being forget about the difficulty of making InGaAsP VCSEL. Calculate the separation between two modes using the same technique! What is the consequence?

Quiz

1. The net gain rate ΓG can be
   (a) smaller than zero.
   (b) greater than zero.
   (c) zero.

2. The smallest possible loss rate $1/\tau_P$ is determined by
   (a) material loss.
   (b) the transparency of the semiconductor.
   (c) induced absoption.
   (d) spectral hole burning
   (e) the mirror transparency.
   (f) intraband relaxation.

3. At the transparency point, the difference of net gain rate ΓG and loss rate $1/\tau_P$ is
   (a) smaller than zero.
   (b) greater than zero.
   (c) zero.

4. At threshold, the difference of net gain rate ΓG_S and loss rate $1/\tau_P$ is
   (a) smaller than zero.
   (b) greater than zero.
   (c) zero.

5. Above threshold and ignoring spontaneous emission, the difference of net gain rate ΓG and loss rate $1/\tau_P$ is
   (a) smaller than zero.
   (b) greater than zero.
   (c) zero.

6. Above threshold and *not* ignoring spontaneous emission, the difference of gain rate ΓG and loss rate $1/\tau_P$ is
   (a) smaller than zero.
   (b) greater than zero.
   (c) zero.

7. The threshold current of a semiconductor laser is minimum if the active region thickness d is

   (a) as small as possible.
   (b) as large as possible.
   (c) optimized with respect to gain constant and confinement factor.

8. Neglecting spontaneous emission, the characteristic curves of photon number and carrier number versus injection current are

   (a) weird curves with kinks.
   (b) straight lines.
   (c) softly bent curves.

9. *Not* neglecting spontaneous emission, the characteristic curves of photon number and carrier number versus injection current are

   (a) weird curves with kinks.
   (b) straight lines.
   (c) softly bent curves.

10. The small-signal response of the laser output power due to a modulation of the injection current can exhibit a resonance peak because

    (a) the laser cavity rings when the electric field amplitude changes rapidly.
    (b) electrons need some time to reestablish intraband equilibrium having experienced a perturbation.
    (c) the energy reservoirs for photons and carriers interact much like the field energies, which are stored in the capacitor and the inductor of a tank circuit.
    (d) the operation point for the injection current has been chosen to exhibit this feature.

11. Relaxation oscillations and switch-on delay of a laser may be avoided by a bias current

    (a) way above threshold.
    (b) at threshold.
    (c) way below threshold.

12. Amplitude-phase coupling

    (a) broadens the laser spectrum.
    (b) makes an oscillation frequency variation influence the field amplitude.
    (c) makes a field amplitude variation influence the oscillation frequency.
    (d) makes a field amplitude variation influence its phase.
    (e) is based on the Kramers-Kronig relation.

13. Which of the following statements about the difference between semiconductor lasers and LED are true?

    (a) Lasers emit higher power at the same drive current.
    (b) Lasers emit light only if drive current is above a threshold value.
    (c) Output from LED spreads out a broader angle.
    (d) LED do not have reflective end facets.
    (e) All of the above.

14. Laser light is produced dominantly by

    (a) stimulated emission.
    (b) spontaneous emission.
    (c) black magic.
    (d) electricity.

15. A Fabry-Perot diode laser operating at $1.3\,\mu$m has a cavity length of $500\,\mu$m and a refractive index of 3.2. How far apart are its longitudinal modes?

    (a) 0.013 nm
    (b) 0.053 nm
    (c) 0.53 nm
    (d) 5.3 nm
    (e) 0.13 nm

16. Assume the Fabry-Perot semiconductor laser above operates at a power of 1 mW per facet. Its spectral half-power width Δf_H at a total output power P_a (see Eq. (3.129) on Page 93) be $\Delta f_H\, P_a = 2\,\text{THz}\,\text{mW}$. How many longitudinal modes do you expect oscillating simultanesously?

(a) 1
(b) 2
(c) 10
(d) 20
(e) 100
(f) 200

17. A distributed feedback laser is

(a) a laser that emits multiple longitudinal modes from a narrow stripe.
(b) a laser with a corrugated layer acting as a grating, which enables single-longitudinal mode operation.
(c) a laser made of two segments that are opticall coupled but electrically separated.
(d) a laser that requires liquid-nitrogen cooling to operate.

18. Which of the following is an important advantage of external modulation of lasers?

(a) Operation is simpler.
(b) Does not require electrical power.
(c) No extra devices are needed.
(d) Avoids wavelength chirp that could cause dispersion.

19. What guides light in a narrow-strip edge-emitting laser?

(a) Reflective layers on the edges of the laser wafer
(b) The stripe has higher refractive index than surrounding material, so it functions as a waveguide.
(c) Coatings applied above and below the junction
(d) Light entering it from an external optical fibre

20. Which of the following is *not* true for a VCSEL?

(a) VSEL emit light from their surfaces.
(b) VSEL beams are rounder than those from edge-emitting lasers.
(c) VSEL can be made easily from GaAs or InGaAsP compounds.
(d) VSEL have low-threshold currents.
(e) VSEL have multilayer coatings as their resonator mirrors.

4 Optical amplifiers

Summary

1. Semiconductor amplifiers are laser oscillators without feedback, i. e., Fabry-Perot lasers with anti-reflection coated facets.

2. A residual reflectivity leads to a ripple in the spectral gain dependency, Eq. (4.2) on Page 101.

3. Angled facets help in reducing the effective facet reflection factors.

4. Semiconductor optical amplifiers (SOA) exhibit strong nonlinearities and may be employed as modulators, switches, wavelength converters and logical optical gates.

5. Erbium-doped fibre amplifiers (EDFA) are the "workhorses" for amplification in WDM systems. EDFA are pumped by diode lasers emitting radiation at 0.98 μm or 1.48 μm, see Fig. 4.2 on Page 104 and Fig. 3.3(b) on Page 54.

6. EDFA can amplify signals between 1.53 μm and 1.62 μm (lower C to upper L band, Table 1.1 on Page 5).

7. Erbium-doped fibres with mirrors on their ends can make lasers, which emit in the same range as the EDFA. They can be tuned in wavelength, or designed to generate a series of very short impulses.

Problems

1. A semiconductor amplifier is 1 mm long and produces a small-signal gain of 30 dB. Assuming that the amplifier is not saturated, and neglecting complicating factors like pump current, how long an amplifier would you need for a 20 dB gain? If the input power was 1 μW, how much does this

reduce the output power?

2. A fibre amplifier is 10 m long and produces a small-signal gain of 30 dB .Neglecting again complicating factors like pump power, answer the same questions as above!

Quiz

1. A semiconductor laser oscillator may be turned into an optical amplifier by
   (a) cutting it in half.
   (b) removing one mirror.
   (c) removing both mirrors.
   (d) removing all optical feedback.
   (e) reducing the operation current.
2. A real-world semiconductor optical amplifier may be turned into a laser oscillator by
   (a) parasitic reflections.
   (b) increasing the injection current sufficiently.
   (c) decreasing the injection current sufficiently.
   (d) modulating the injection current with a 10 GHz microwave signal.
   (e) antireflection-coating the facets.
3. A semiconductor optical amplifier may be used as
   (a) a limiter.
   (b) an on-off switch.
   (c) a 40 GHz modulator.
   (d) a wavelength converter.
   (e) a logical gate.
4. Semiconductor optical amplifiers
   (a) are insensitive to WDM channel crosstalk.
   (b) can be designed for many wavelengths from 1.2 μm to 1.7 μm.
   (c) need little electrical power.
   (d) have low gain.
5. Which of the following is *not* an advantage of Erbium-doped fibre amplifiers?
   (a) High gain
   (b) Insensitivity to signal speed
   (c) Operation near 1.3 μm
   (d) Broad amplification bandwidth at 1.55 μm
6. What is the power source for Erbium-doped amplifiers?
   (a) Electric current passing through the fibre
   (b) They require no power.
   (c) Diode pump lasers emitting at 0.98 μm or 1.48 μm
   (d) Power is drawn from the optical signal.

5 Pin photodiode

Summary

1. An optical receiver detects light signals and converts them into electronic form. It includes an optical detector (photodetector, PD), an electronic amplifier, and thresholding and retiming circuits.
2. Any optical amplifiers and wavelength division multiplexing precede the receiver.
3. Detectors are colour-blind, so optical channels must be separated and routed to different receivers, one per channel, see Fig. 1.6 on Page 4.
4. PD for communication purposes are usually reverse-biased semiconductor diodes. Light striking the carrier-depleted region near the junction generates free electrons-hole pairs, so a current flows through the diode and becomes the electronic output signal.
5. Pin-photodiodes have a high-resistivity intrinsic layer between p and n layer, which improves sensitivity. They are widely used for their high speed and sensitivity, Fig. 5.1 on Page 106.

6. Avalanche photodiodes (APD) have a high reverse-bias voltage that creates an internal cascade of electrons, which multiplies their electrical output power above that of pin-photodiodes. They are not as fast and not as linear as pin-photodiodes, and have higher noise.

7. The wavelength at which detectors respond depend on their composition, Fig. 5.4 on Page 110. Silicon is used at $0.4\,\mu\text{m} \ldots 1\,\mu\text{m}$, InGaAs at $0.8\,\mu\text{m} \ldots 1.7\,\mu\text{m}$.

8. The impulse response or the bandwidth of a detector determines how fast its output current rises when the input optical signal jumps, Fig. 5.6 on Page 113. Loss of high frequencies in the signal decreases rise time, and rounds sharp-edged signals.

9. For pin-photodiodes there is a trade-off between quantum efficiency η (sensitivity) and bandwidth $f_{3\,\text{dB}}$, Eq. (5.36) on Page 114. Edge coupling photodiodes overcome this constraint.

10. Photodiode current i and light external power P_e are proportional, $i = SP_e$. The sensitivity (responsivity) at light frequency f_S is $S = \eta e/(hf_S)$, Eq. (5.39) on Page 115. The sensitivity increases with wavelength λ_S simply because more photons are present for the same optical power.

Problems

1. If a detector has a sensitivity of $1\,\mu\text{A} / \mu\text{W}$ and its input is $-30\,\text{dBm}$, what is the output current? How many electrons per second does this correspond to, recalling that 1 A transports 0.624×10^{19} elementary charges $e = 1.602 \times 10^{-19}$ As per second? How many electrons would be contained in a 1 ns wide impulse?

2. A data sheet claims a record sensitivity of $S = 0.7$ for a silicon pin-photodetector at $0.8\,\mu\text{m}$. Comment!

3. A silicon pin-photodetector has peak sensitivity of $S = 0.5\,\text{A} / \text{W}$ at $0.8\,\mu\text{m}$. If the input signal is $-20\,\text{dBm}$, how many electrons N_n does a 1 ns wide impulse produce?

4. The risetime of an InGaAs pin-photodiode is 5 ps. The risetime of an InGaAs avalanche photodiode at the same wavelength is 100 ps. All other things being equal, how much larger is the bandwidth of the pin-photodiode? If the transit-time is all that limits the bandwidth of the two devices, what are their bandwidths, assuming weak absoption?

5. The dark current in a Germanium pin-photodiode is 100 nA, compared to 1 nA in an InGaAs pin-photodiode. Suppose that all other things are equal, and the Germanium detector has a sensitivity of $-15\,\text{dBm}$, limited by dark current. What would be the sensitivity of the InGaAs detector?

6. How does the impulse response of a pin-photodiode look like for the cases of weak and strong absorption? What is the reason for the spike at $t = 0$ for the case of strong absoption?

Quiz

1. Photodiodes used as fibre-optic detectors normally are
   (a) reverse biased.
   (b) thermo-electrically cooled.
   (c) forward biased.
   (d) unbiased to generate a voltage like asolar cell.
   (e) none of the above.

2. Silicon detectors are usable at wavelengths
   (a) $800 \ldots 900\,\text{nm}$
   (b) $1\,300\,\text{nm}$
   (c) $1\,550\,\text{nm}$
   (d) all of the above.

3. Which detector material is most often used in the $1.55\,\mu\text{m}$ window?
   (a) Silicon
   (b) InGaAs
   (c) GaAs
   (d) Germanium
   (e) All of the above

4. A pin-photodiode is a

(a) point-contact diode detector in which a pin makes contact with the semiconductor.

(b) semiconductor detector with an undoped intrinsic region between p and n-materials.

(c) circuit element used in receiver amplification.

(d) hybrid detector-amplifier.

5. An avalanche photodiode

(a) has an internal amplification stage based on avalanche multiplication of electrons.

(b) has a external amplification stage containing a single transistor.

(c) generates a photocurrent in the base of a transistor, which amplifies the signal.

(d) is an ordinary transistor that generates an optical signal under bright lights.

6. What is the maximum value of the quantum efficiency possible in a pin-photodiode?

(a) 0

(b) 0.5

(c) 0.9

(d) 1.0

(e) 100

7. What type of photodetector could have a sensitivity of $S = 20\,\mathrm{A}\,/\,\mathrm{W}$?

(a) Silicon pin-photodiode

(b) Silicon avalanche photodiode

(c) InGaAs pin-photodiode

(d) InGaAs avalanche photodiode

(e) No detector on earth wll meet this requirement

6 Noise

Summary

1. The classical light power $P_e(t)$ results from an average over a few optical cycles. Fluctuations in $P_e(t)$ are transferred to the photocurrent $i(t) = SP_e(t)$.

2. The ideal classical signal has constant amplitude and phase. Quantum mechanics tells that such an ideal laser signal consists of a stream of independent photons which are Poisson distributed in time.

3. Each photon generates an electron-hole pair with a quantum efficiency η, Eqs. (5.14), (5.36) on Pages 108, 114. Thus, the photocurrent consists of a stream of statistically independent elementary charges which are Poisson distributed in time. This type of quantum noise is known as shot noise.

4. In fact, each time an electron transverses a potential barrier at random points of time, depending on its instantaneous thermal energy, the quantized nature of electronic charge becomes visible, and the associated current shows shot noise fluctuations. This is true for any pn-junction.

5. The equivalent short-circuit photocurrent noise current (shot noise, quantum noise) in a differential frequency interval $\mathrm{d}f$ and for a mean current $\bar{i}$ is given by (Eq. (6.4) on Page 120)
$$\overline{|i_{RD}|^2} = 2e\bar{i}\,\mathrm{d}f.$$

6. At finite temperatures electrons move randomly, which manifests as a fluctuating current even in the absence of an applied voltage. This additional noise component is referred to as thermal noise. (Johnson noise, Nyquist noise).

7. The equivalent short-circuit noise current of the conductance G_Q in a differential frequency interval $\mathrm{d}f$ and at a temperature T_0 is given by (Eq. (6.15) on Page 122)
$$\overline{|i_Q|^2} = 4kT_0G_Q\,\mathrm{d}f, \quad T_0 = 293\,\mathrm{K}\,.$$

8. The noise figure F is specified by the quotient of the available signal-to-noise ratio $\mathrm{SNR}_{\mathrm{input}}^{(\mathrm{available})}$ at the device's input (referring to the available input powers) and the signal-to-noise ratio $\mathrm{SNR}_{\mathrm{output}}$ at the out-

put (referring to the actual output powers), Eq. (6.17) on Page 123.

9. The output SNR of any linear device is always worse, $\mathrm{SNR}_{\mathrm{output}} < \mathrm{SNR}_{\mathrm{input}}^{(\mathrm{available})}$.

10. Alternatively, the noise figure $F = 1 + T_R/T_0$ specifies the equivalent fictitious temperature increase T_R which must be attributed to the source conductance G_Q if the amplifier is regarded noiseless, Eq. (6.17) on Page 123.

11. Optical amplifier noise is defined for the detected electric signals (it must be measured somehow!), and therefore depends on the detection process.

12. If the amplifier is regarded noiseless, an equivalent, fictitious noise power $P_{R\,\mathrm{eq}}$ must be thought entering the amplifier input. In an electrical bandwidth B and depending on the receiver type (homodyne, heterodyne, or, here, direct receiver) there are $\kappa = \frac{1}{2}$, 1, or 2 equivalent (unpolarized) noise photons of energy hf_S (Eq. (6.18) on Page 123),
$$P_{R\,\mathrm{eq}} = \kappa h f_S B, \quad \kappa = \tfrac{1}{2}, 1, 2\,.$$

13. Because $P_{R\,\mathrm{eq}}$ represents minimum-uncertainty quantum fluctuations this power cannot be extracted from the system.

14. After an optical amplifier having a single-pass gain $\mathcal{G}_s$, only the (unpolarized) optically amplified spontaneous noise power $P_R = 2\,(\mathcal{G}_s - 1)\,hf_S B$ may be directly detected. For a transparent (but not noise-free!) amplifier $\mathcal{G}_s = 1$ no output power can be extracted.

15. For direct detection of an optical signal power P_S in an electrical bandwidth B the optimum electrical SNR is obtained if thermal and amplifier noise may be neglected, and quantum (shot) noise of the photocurrent is the only source of uncertainty. For a quantum efficiency $\eta = 1$ this quantum-noise limited SNR is (Eqs. (6.19) and (7.2) on Pages 123 and 125)
$$\gamma_{\mathrm{dir}}(\eta = 1) = \gamma_1 = \eta P_S/(2hf_S B).$$

16. Practically this is of no use, because P_S should be in the mW-region so that electronic noise might be neglected. An optical pre-amplifier in front of the photodetector, however, makes shot-noise limited reception feasible.

17. To this end the direct detection (quadratic rectification) of a coherent signal with narrowband noise must be studied, Fig. 6.3 on Page 123. As it turns out, if we employ an optical filter with a bandwidth $B_O = 2B$ in front of the photodetector, the dominant noise component results from the mixing process between both the lower and upper noise sidebands and the coherent carrier.

18. For an ideal optical amplifier without excess noise (ideally population-inverted), but with large gain *and direct detection* the minimum noise figure is (Eq. (6.21) on Page 124)
$$F_{\mathrm{dir,\,min}} = 2\,.$$
This is in contrast to coherent detection (measuring the amplitude and possibly the phase of the electric field) where basically $F < 2$ may be reached as with an electrical amplifier.

Problems

1. A pin-photodetector delivers an average current of $\bar{i} = 100\,\mathrm{nA}$ for an NRZ signal of 10 Gbit/s. Calculate the shot-noise limited signal-to-noise ratio $\gamma = \bar{i}^{\,2}/\overline{|i_{RD}|^2}$!

2. A thermally fluctuating conductance of $G_Q = 1/50\,\Omega$ may be described at room temperature by a fluctuating current source $\overline{|i_Q|^2}$. What is the effective current?

Quiz

1. An electromagnetic signal, which is ideally stable in the classical sense, has
   (a) shot noise.
   (b) constant amplitude and phase.
   (c) finite temperature.
   (d) Nyquist noise.

2. Quantum mechanics tells that an ideal laser signal
   (a) produces shot noise.
   (b) fluctuates in amplitude.
   (c) fluctuates in phase.

(d) consists of a stream of independent photons which are Poisson-distributed in time.

(e) has constant amplitude and phase.

3. When an ideal pin-photodetector is illuminated with an ideal laser signal, the photocurrent

   (a) shows quantum noise

   (b) is constant.

   (c) exhibits shot noise.

   (d) consists of a stream of independent electrons which are Poisson distributed in time.

   (e) has all the above properties.

4. An electronic amplifier is usually employed for amplitude- and phase-sensitive (= coherent) reception. For this case, it has a minimum noise figure

   (a) smaller than 2.

   (b) smaller than 1.

   (c) smaller than 0.

   (d) larger than 1.

   (e) larger than 2.

5. An optical amplifier has a minimum noise figure

   (a) smaller than 2 for coherent detection.

   (b) smaller than 1 for coherent detection.

   (c) smaller than 0 for any type of detection.

   (d) larger than 1 for direct detection.

   (e) larger than 2 for direct detection.

6. The equivalent noise power $P_{R\,\mathrm{eq}}$

   (a) may be used for frying eggs.

   (b) cannot be extracted from the system.

   (c) represents minimum-uncertainty quantum fluctuations.

7 Receivers and detection errors

Summary

1. Binary electrical signals must be thresholded to find out their logical value. Noise can introduce errors in this decision process.

2. The bit error probability (bit error "rate" BER) is the likelihood for detecting erroneously a logical 0 if actually a 1 was received plus the probability for detecting erroneously a logical 1 if actually a 0 was received.

3. The optimum decision threshold voltage u_S minimizes the bit error probability, $\partial\,\mathrm{BER}\,/\partial u_S = 0$. For Gaussian noise the bit-error parameter (quality factor) Q relates the mean voltage u_1 of a logical 1 to the sum of the standard deviations $\sigma_0 + \sigma_1$ for a logical 0 and a logical 1, i. e., to the sum of the effective noise voltages.

4. If logical 1 and 0 are approximately evenly distributed, the minimum bit error probability is (Eq. (7.15) on Page 128)
$$\mathrm{BER} = \tfrac{1}{2}\,\mathrm{erfc}(Q/\sqrt{2}\,).$$

5. For many practical cases where the main noise stems from optical amplifiers the square Q^2 of the bit-error parameter equals the electrical signal-to-noise ratio $\gamma \approx Q^2$.

6. A minimum of $N_{e\,\mathrm{bit}} = N_e/2 = 10$ photons per bit (1-bit and 0-bit with equal probability) must be received. This is the absolute quantum limit for an ideal binary system with direct reception for BER $= 10^{-9}$. With a high-quality optical amplifier this limit may be approached closely.

Quiz

1. How many separate receivers are required for a 32-channel WDM system with an optical preamplifier that provides 15 dB gain on all channels?

   (a) 1

   (b) 4

   (c) 8

   (d) 32

   (e) 100

2. When would an optical preamplifier be used at the receiver end of a system?

   (a) Always
   (b) When input power is below 1 mW
   (c) When the input signal contains multiple optical channels
   (d) When input power is below receiver sensitivity
   (e) Never

3. What is present in a digital receiver that is not present in an analogue receiver?

   (a) Nothing
   (b) A detector
   (c) Thresholding and retiming circuits
   (d) Amplification circuits
   (e) Wavelength division multiplexing

4. Employing a high-quality optical preamplifier allows

   (a) shot-noise limited detection with a pin-photodiode and a real-world electronic post-amplifier.
   (b) introduces additional noise into the system.
   (c) reduces the maximum signal-to-noise ratio of a pin-photodetector, disregarding the influence of an electronic post-amplifier.
   (d) all of the above.

5. The eye diagram is a means to visualize

   (a) the results of iris diagnostics.
   (b) the certainty of biometrical identification.
   (c) signal quality.
   (d) feedback circuits by graphical means.

6. The bit error parameter (quality factor) Q may be related to the

   (a) bit error probability in general.
   (b) signal-to-noise ratio in general.
   (c) signal-to-noise ratio, if Gaussian statistics apply for the noise.
   (d) bit error probability, if Gaussian statistics apply for the noise.

7. What bit error probability is most often specified for digital communication systems?

   (a) 40 dB
   (b) 10^{-4}
   (c) 10^{-9}
   (d) 10^{-12}
   (e) 10^{-18}

8. The minimum number of photons per bit at the absolute quantum limit for an ideal binary system with direct reception

   (a) cannot be approached in practice.
   (b) is $N_{e\,\mathrm{bit}} = N_e/2 = 10$ for BER $= 10^{-9}$.
   (c) is closely approached in reality by employing a high-quality optical amplifier.
   (d) may be computed assuming Poisson noise for high optical powers and zero noise for no optical power.

9. A bit error floor

   (a) is connected with additive noise.
   (b) is none of my concern.
   (c) is connected with multiplicative noise.

Solutions to problems and quizzes

1 Introduction

Solutions to problems

1. 99 % transmission per mm means a loss of $(-10 \lg 0.99 = 0.044)\,\mathrm{dB}\,/\,\mathrm{mm}$. Propagation through a 7 cm thick window results in a loss of $70\,\mathrm{mm} \times 0.044\,\mathrm{dB}\,/\,\mathrm{mm} = 3\,\mathrm{dB}$, i. e., only half of the light power is transmitted.

2. Distributed amplification is necessary because the power reduces exponentially.[10] — Attenuation is cumulative and usually uniform throughout the fibre length. Assume that 99 .995% of the respective input power emerges from the 1st, 2nd, ..., 1 000th meter, then for a $z = 1\,\mathrm{km}$ long fibre the output power $P(z)$ is the fraction $0.99995^{1\,000} = 0.95$ of the input power P_0, namely

$$P(z) = P_0 \times \left(\begin{array}{c}\text{unit-length}\\ \text{transmission}\end{array}\right)^{\frac{z}{\text{unit length}}},$$

$$P(1\,\mathrm{km}) = P_0 \times 0.99995^{1\,000} = 0.95\,P_0\,.$$

   The unit-length transmission, which is smaller than 1, may be expressed by a positive unit-length factor α through an exponential $\exp(-\alpha)$,

$$P(z) = P_0\,[\exp(-\alpha)]^z = P_0 \exp(-\alpha z)\,.$$

   This is the same relation as Eq. (2.23) on Page 13. Because α describes the transmission per unit length and z is measured as a multiple of the same length, their dimensions cancel and may be chosen arbitrarily. It is customary to express α in km^{-1}. The power attenuation a measured in dB (decibel, a tenth of the measure bel[11] which is virtually never used) is

$$a = 10 \lg \frac{P_0}{P(z)} \quad (\text{here: } \frac{a}{z} = 0.2\,\mathrm{dB}\,/\,\mathrm{km}).$$

3. While tens of millions of fibre kilometers already exist in the underground ducts, subscribers are mostly connected by existing telephone wires which represent an immense investment, roughly 30 billion € in Germany. Recent trends are to connect providers and subscribers via *f*ibre *t*o *t*he *h*ome (FTTH) technology.[12,13,14]

4. The operating frequency range of optical fibres (193 THz) is far beyond the frequency range of natural or man-made terrestrial electromagnetic interferers (< 100 GHz). Light cannot easily enter a fibre through the cladding: If no light can leak out because of total internal reflection, then no light can "leak in" because of the same reason. Further, there is a plastic protection coating which light cannot penetrate.

5. A "golden rule of thumb" says: For logic operations nothing equals electrons, for transport of information you are much better off with photons. Meanwhile some progress has been made in all-optical processing of signals,[15] however, we are way off from the optical computer, so exchange and switching still relies on conventional technology.

Solutions to quiz

1(a), 2(d), 3(a), 4(b), 5(c), 6(c)

[10] See Footnotes 20 and 21 on Page 6

[11] Alexander Graham Bell, British-American inventor, ⋆Edinburgh 03.03.1847, †in Baddeck (Nova Scotia, Kanada) 01.08.1922; constructed the first practical telephone (patent 1876), which was improved by T. A. Edison by employing a carbon microphone; basic arrangement still in use today.

[12] N. J. Frigo, K. C. Reichmann, P. P. Iannone: Fiber to the home: Niche market or universal broadband access? Technical Digest Optical Fiber Communication Conference (OFC'04), Los Angeles (CA), USA, 22.–27.02.2004. Tutorial WO1

[13] Th. Pfeiffer, J. Hehmann, H. Schmuck, W. Freude, J. Vandewege, Hiroki Yanagisawa: Monitoring and protecting the optical layer in FTTH networks. FTTH Conference & Expo, Las Vegas (NV), USA, October 3–6, 2005

[14] DT Flings Billions at Fiber Access. — German incumbent Deutsche Telekom AG plans to spend up to 3 billion € on an ambitious fiber-to-the-curb (FTTC) project that is set to deliver up to 50 Mbit/s of broadband access to homes in 50 German cities by the end of 2007. Light Reading, Sept. 1, 2005. http://www.lightreading.com

[15] T. Morioka: Recent progress in ultra high-speed and wideband OTDM / WDM technologies. 30th European Conf. Opt. Commun. (ECOC'04), Stockholm, September 5–9, 2004, Paper Th1.2 (Tutorial)

2 Light waveguides

Solutions to problems

1. No, because the time derivatives of $\vec{E}$ and $\vec{H}$ are always connected to non-vanishing spatial derivatives of $\vec{H}$ and $\vec{E}$, respectively, at the same point of space $\vec{r}$.

2. $\operatorname{div}\vec{D} = \varrho$, so space charges, e. g., electrons, may exist, but are assumed absent in Eq. (2.1). — $\operatorname{div}\vec{B} = 0$ implies that local magnetic charges are either absent or only appear in pairs in form of infinitesimally short dipoles with a North pole and an associated South pole. Magnetic monopoles (*either* a South *or* a North pole) for which $\operatorname{div}\vec{B} \neq 0$ holds are not observed in nature.

3. From Eq. (2.2) on Page 7 we see

$$\breve{\Psi}(f) = \int_{-\Delta t/2}^{+\Delta t/2} \Psi_0 \, \mathrm{e}^{-\mathrm{j}\,2\pi f t} \, \mathrm{d}t$$

which leads to

$$\breve{\Psi}(f) = \Psi_0 \Delta t \, \frac{\sin(\pi f \Delta t)}{\pi f \Delta t} \, .$$

The spectral half-maximum width Δf is defined by $\breve{\Psi}(\Delta f/2)/\breve{\Psi}(0) = 1/2$. This leads to $\pi\,(\Delta f/2)\,\Delta t \approx \pi/2$ resulting in

$$\Delta t \, \Delta f \approx 1 \, .$$

The product of temporal width and spectral width is about 1. This is a general property of Fourier-transform related functions.

Solutions to quiz

1(a), 2(b), 3(c)

2.1 Fundamentals of wave propagation

Solutions to problems

1. Homogeneous plane waves vary in time and space according to $\exp\left[\mathrm{j}\left(\omega t - \vec{k}\cdot\vec{r}\right)\right]$, propagate with a real propagation vector $\vec{k}$, and therefore have constant amplitude everywhere. The phase fronts are plane and propagate with a phase velocity $v = c/n$.

2. In a dispersive medium where the refractive index depends on frequency.

3. Minimum dispersion near $\lambda_0 = 1.3\,\mu\mathrm{m}$, minimum loss near $\lambda_\alpha = 1.5\,\mu\mathrm{m}$

4. Rayleigh scattering increases with f^4, so mostly blue light is scattered away from the incident sunlight path. For the same reason the sun appears red when it propagates along a longer air path at sunrise and sunset: The blue frequency components have been scattered away. On the moon there is no scattering atmosphere. Either we are blinded by the sun, or the sky appears pitch dark.

5. $\vartheta_{1T\,\mathrm{C}} = \arccos(1/2.4) = 65\,^\circ$, $\vartheta_{1T\,\mathrm{SiO_2}} = \arctan(1/1.5) = 48\,^\circ$. In a diamond,[16] total reflection takes place in a cone with a much larger semi-angle $\vartheta_{1T\,\mathrm{C}} = 65\,^\circ$ than in glass, and therefore a much larger amount of light is reflected if the cut has been made appropriately.[17] In addition, strong dispersion of the refractive index gives "colour" and "fire" to the light which is backscattered from the diamond.

6. General relation: $r_p = \exp(\mathrm{j}\,\varphi_p)$. E-polarization: $\varphi_E = 2\arctan(|k_{2x}^{(i)}|/k_{1x})$, $\varphi_H = 2\arctan(n_1^2/n_2^2 \cdot |k_{2x}^{(i)}|/k_{1x})$. Penetration depth of wave in medium $n_2 < n_1$ with lesser optical density is $1/|k_{2x}^{(i)}| = \left(k_0\sqrt{n_1^2\cos^2\vartheta_1 - n_2^2}\,\right)^{-1} \sim \lambda$.

[16] Diamond, from Greek ἀδάμαστος, invincible, unmarried(!); also ἀδαμάντινος, made of steel, steely

[17] For the Tolkowsky cut, a ray which enters the diamond parallel to its axis hits the lower facets at an angle of $\vartheta_{1\,\mathrm{lo}} = 49\,^\circ\ 15'$ to the facet plane, well below the TIR angle $\vartheta_{1T\,C} = 65\,^\circ$. Total reflection from this surface hits the upper inclined facet at an angle of $\vartheta_{1\,\mathrm{up}} = 43\,^\circ$ to the plane of the diamond-air interface, and this is again smaller than $\vartheta_{1T\,C}$.

7. The power reflection factor for one glass-air interface is $R = \left(\frac{1-1.5}{1+1.5}\right)^2 = 4\,\%$. The corresponding power transmission factor is $T = 1 - R = 96\,\%$.

8. The configuration has four air-glass interfaces. Each glass plate is so thick that for natural light interference does not play a role. The minimum reflection loss is $R = 4\,\%$ per interface. The total power loss is therefore $R_{\text{total}} = 1 - (1-R)^4 = 15\,\%$ (a bit less than the simple estimate $R_{\text{total}} \approx 4 \times 4\,\% = 16\,\%$). For an inclined windscreen the reflection factor becomes larger, see Fig. 2.6 on Page 18 for the transition from glass to air.

9. At $\lambda_\alpha = 1.55\,\mu\text{m}$ we read from Fig. 2.4 on Page 16 $M_\alpha = 20\,\text{ps}\,/(\text{km}\,\text{nm})$. If two plane waves propagate at wavelengths $1.550\,\mu\text{m} \pm 0.5\,\text{nm}$, respectively, their arrival times differ by $20\,\text{ps}\,/(\text{km}\,\text{nm}) \times 1\,\text{km} \times (2 \times 0.5\,\text{nm}) = 20\,\text{ps}$ after a propagation length of 1 km.

10. At $\lambda_0 = 1.2758\,\mu\text{m}$ we read from Fig. 2.4 on Page 16 $M_0 = 0\,\text{ps}\,/(\text{km}\,\text{nm})$ and $N_0 = 0.053\,\text{ps}\,/(\text{km}\,\text{nm}^2)$. This is the zero of material dispersion of first order. Two plane waves propagating at wavelengths $1.2758\,\mu\text{m} \pm 0.5\,\text{nm}$ experience delay differences of only 0.053 ps after a propagation length of 1 km.

11. The relative group delay in glass with a refractive index of $n = 1.5$ is $n/c = 1.5/(3 \times 10^8\,\text{m}\,/\,\text{s}) = 5\,\mu\text{s}\,/\,\text{km} = 5\,\text{ns}\,/\,\text{m} = 5\,\text{ps}\,/\,\text{mm}$.

Solutions to quiz

1(e), 2(e), 3(d), 4(c), 5(a), 6(d), 7(c), 8(b), 9(b)(c), 10(b)(c)(d), 11(a)(c)(d), 12(b)(c), 13(e), 14(b)(e)

2.2 Principles of waveguiding

Solutions to problems

1. The rays from the upper rim of the sun run through regions of the atmosphere where the air is thinner, the refractive is index smaller, and the light propagates faster. Rays from the lower rim propagate slower. This adds to a focussing effect as with a cylindrical lens.

2. Given a temperature inversion in the troposphere, then a region of larger refractive index exists, and a wave may be guided.

3. Heated air is thinner and has a smaller refractive index. The temperature over a heated area of tar fluctuates, and so does the refractive index. In such an inhomogeneous medium rays do not follow straight lines. — Turbulence in the air causes pressure fluctuations and associated refractive index fluctuations. Therefore stars seem to twinkle.

Solutions to quiz

1(c), 2(c), 3(a)(b)(c)(e), 4(c)

2.3 Slab waveguide

Solutions to problems

1. $\vartheta_T = \arcsin\Big(\sqrt{n_1^2 - n_2^2}\Big/\, n_1\Big) = 47\,^\circ$

2. $A_N = \sqrt{n_1^2 - n_2^2} = 1.1$, $\sin\gamma_N = n_1 \sin\vartheta_T = \sqrt{n_1^2 - n_2^2}$ has no physical meaning in this case.

3. They die away exponentially and are reflected back.

4. A plane wave $\exp\big[\mathrm{j}(\omega t - k_{1x}x)\big]$ traveling from the lower boundary to the upper one and back again accumulates a phase of two times $-k_{1x}h$. Taking the polarization-dependent phase of the reflection coefficient into account, two times the phase φ_p must be added. The total phase must equal an integer number m of -2π. — A roundtrip phase equal to an integer multiple of -2π defines a situation where a plane wave incident perpendicularly to the slab boundaries (along the x-axis) would be transmitted without any reflection. A roundtrip phase of $-\pi$ would lead to a maximum reflection.

5. Whenever the weak guidance assumption $\Delta \ll 1$ does not apply.

6. From Eq. (2.52) on Page 26: $\lambda_{1G} = 1.4\,\mu\mathrm{m}$, $f_{1G} = 215$ THz

7. From Eq. (2.52) on Page 26: $M_g \approx \big\lfloor \frac{4}{\pi} V \big\rfloor = 17$. More accurate solution: $M_g = 2\big(1 + \lfloor V/(\pi/2) \rfloor\big) = 2\big(1 + \lfloor 8.99 \rfloor\big) = 18$

8. From Eq. (2.54) on Page 27: The factor $\mathrm{d}(VB)/\,\mathrm{d}V$ is called group delay factor. At the cutoff frequency of the highest order mode $V = V_{mG}$, the mode m fills the infinite cladding cross-section, therefore $t_g/L = n_{2g}/c$ and $\mathrm{d}(VB)/\,\mathrm{d}V = 0$. On the other hand, if the frequency becomes very high, $V \to \infty$, the field is totally concentrated to the core region, and $t_g/L = n_{1g}/c$, $\mathrm{d}(VB)/\,\mathrm{d}V = 1$. The maximum relative group delay difference is therefore $\Delta t_g/L = (n_{1g} - n_{2g})/c \approx (n_1 - n_2)/c = \Delta/(c/n_1) = 15\,\mathrm{fs}\,/\,\mathrm{mm} = 17\,\mathrm{ps}\,/\,\mathrm{m} = 17\,\mathrm{ns}\,/\,\mathrm{km}$.

9. From Eq. (2.56a) on Page 27: $\Delta t_g/L = C\,\Delta\lambda = 17\,\mathrm{as}\,/\,\mathrm{mm} = 17\,\mathrm{fs}\,/\,\mathrm{m} = 17\,\mathrm{ps}\,/\,\mathrm{km}$.

10. See Pages 28 and 43.

11. $f_t L = (4\Delta t_g/L)^{-1}$, see Eq. (2.58) on Page 28

12. $\Delta t_H\,\Delta f_H \approx 1$, see Eq. (2.60) on Page 28

13. $\Delta t_H = \sqrt{(\Delta t_{H\,\mathrm{out}})^2 - (\Delta t_{H\,\mathrm{in}})^2} = 34\,\mathrm{ps}$, $f_H \approx 0.441/\Delta t_H = 13\,\mathrm{GHz}$, see Eq. (2.61) on Page 29

14. The plane phasefront of a straight-waveguide mode bends, because at some radius beyond the waveguide core the local speed of light is exceeded. The phasefront lags, and power spills out perpendicularly to the bend radius.

15. By the size and by the proximity of the waveguides, by the length of the coupling region, and by the operating wavelength.

16. Explain by looking at Fig. 2.13 and 2.14 on Page 32.

Solutions to quiz

1(b), 2(a), 3(e), 4(c)(f), 5(c), 6(b), 7(b)(d)(f)

2.4 Strip waveguide

Solutions to problems

1. EH_{42}-mode: $(4+1)\times(2+1) = 15$ intensity maxima. EH_{00}-mode: $(0+1)\times(0+1) = 1$ intensity maximum.

2. EH-modes and HE-modes differ by the direction of the electric field in the cross-section: The dominant E-component is along the x- (EH-mode) and the y-axis (HE-mode). Further, there is a jump of the electric field normal to the lateral boundaries, along the x-axis for an EH-mode, along the y-axis for an HE-mode, see Figs. 2.16(a) and 2.19 on Pages 33 and 36.

Solutions to quiz

1(b), 2(a)(b)

2.5 Fibre waveguides

Solutions to problems

1. Substituting Eq. (2.74) on Page 39 in Eq. (2.58) on Page 28, a bit rate of $f_t = c/(4Ln_1\Delta) = 12.5\,\text{MHz}$ results.

2. Substituting Eq. (2.78) on Page 44 in Eq. (2.58) on Page 28, a bit rate of $f_t = c/(4Ln_1\Delta^2/2) = 2.5\,\text{GHz}$ results.

3. The expected arrival time when randomly changing the "transport medium" (propagating an impulse randomly in mode $\text{pol}_{\text{slow}} \mathrel{\widehat{=}}$ "by foot" as well as in mode $\text{pol}_{\text{fast}} \mathrel{\widehat{=}}$ "by taxi") would be limited to the interval $t_{g\,\text{fast}} \le t_g \le t_{g\,\text{slow}}$. If the distance to the theater would be short enough, you had no chance in changing the transport medium: Whatever transport you choose first would decide on t_g, and the arrival time spread would be $\Delta t_g = t_{g\,\text{fast}} - t_{g\,\text{slow}} = 1\,\text{h}$. If L becomes larger the probability of a transport change increases. If you change midway, $t_g = t_{g\,\text{slow}}/2 + t_{g\,\text{fast}}/2 = 35\,\text{min}$ would result. If you increase the distance then you could change your mind much more frequently. The average arrival time would still increase $\sim L$, but — employing the central limit theorem[18,19] (German *zentraler Grenzwertsatz*) — the arrival time *spread* would increase less than linearly, namely by $\Delta t_g \sim \sqrt{L}$.

4. $L = [1\,\text{ps}\,/(0.1\,\text{ps}\,/\sqrt{\text{km}})]^2 = 100\,\text{km}$

5. Reduced dispersion-slope designs tend to have smaller mode-field diameters so that more optical power is concentrated in a smaller fibre cross-section. This increases the strength of nonlinear effects, which can cause crosstalk between WDM channels.

6. A broadband-compensation of dispersion would increase nonlinear crosstalk. It is better to have alternating sections of positive dispersion (CSF) and negative dispersion (DCF) resulting in zero dispersion at the end of the transmission link. To achieve such a "dispersion map" over a broad frequency range is a difficult task.

7. $C_{\text{net}} \cdot (L_{\text{exist}} + L_{\text{DCF}}) = C_{\text{exist}} L_{\text{exist}} + C_{\text{DCF}} L_{\text{DCF}}$, therefore:
$$L_{\text{DCF}} = L_{\text{exist}}\,\frac{C_{\text{exist}} - C_{\text{net}}}{C_{\text{net}} - C_{\text{DCF}}}$$

8. $I = 100 \times 1\,\text{mW}\,/50\,\mu\text{m}^2 = 200\,\text{kW}\,/\,\text{cm}^2$

9. Given the price advantage of multimode fibres, the bandwidth of parabolic-index fibres is significantly larger than the bandwidth of equivalent step-index fibres, compare Solutions 1 and 2 above.

10. If you integrate the field of a multimode fibre all over the infinite cross-section, then the total power equals the sum of the modal powers. Interference of modes may only be observed if the intensity in a selected part of the fibre end is measured.

[18] Papoulis, A.: Probability, random variables, and stochastic processes. 3rd ed. New York: McGraw-Hill 1991. S. 49 Gl. (3-27)

[19] Feller, W.: An introduction to probability theory and its applications, Vol. I 2nd ed. New York: John Wiley 1964. S. 168 ff.

Solutions to quiz

1(a), 2(b), 3(c), 4(b),[20] 5(c), 6(b)

3 Light sources

Solutions to problems

1. $m_{q\max} = \max(1, \Delta f/\delta f_q)$, $\delta f_q = c/(2nL_q)$; mode spacings $\delta f_x = 420\,\text{THz}$, $\delta f_x = 14\,\text{THz}$, $\delta f_x = 104\,\text{GHz}$; amplification bandwidth $\Delta f = 12\,\text{THz}$; modal counts $m_{x\max} = 1$, $m_{y\max} = 1$, $m_{z\max} = 115$, $M_{\text{tot}} = 115$ (Eq. (3.2) on Page 48).
2. $M_{\text{tot}} = \varrho_{\text{tot}}(f)\,V\,\Delta f = 8\pi n^2 n_g/(c\lambda^2) \times V\,\Delta f = 7\,970$ (Eq. (3.7) on Page 49). The such computed modal density $\varrho_{\text{tot}}(f)$ is way too large because the modal number is greatly overestimated when disregarding the requirement $m_{q\max} \geq 1$.

3.1 Luminescence and laser radiation

Solutions to quiz

1(b)(e), 2(a), 3(c), 4(d), 5(c), 6(a)(c)(d)(e)(i), 7(a)(b), 8(b)(c)(e)(f)(g), 9(d)(g), 10(a)(b)(c)(d), 11(a)(b)(c)(d)(e), 12(c), 13(e)

3.2 Semiconductor physics

Solutions to quiz

1(b)(c)(e), 2(c)(d), 3(c)(d)(e), 4(f), 5(a)(b), 6(a)(c)(d), 7(a)(d), 8(a)(b)(e)(f), 9(a)(b), 10(b)(c), 11(a)(b)(d)

3.3 Light-emitting diode

Solutions to problems

1. $P/(hf) = \eta_{\text{int}} I/e$, $I/e = 3.1 \times 10^{15}/\,\text{s}$, $N_P = P/(hf) \times 25\,\text{ps} = 5.4 \times 10^6$, $P = 72\,\text{mW}$
2. $P_a = \eta_{\text{opt}} P = 1.8\,\text{mW}$

Solutions to quiz

1(a), 2(d), 3(b)(c), 4(a)(c), 5(b)(c)(d)

3.4 Laser diode

Solutions to problems

1. See Eq. (3.83) on Page 82.
2. Because of the current and the associated small photon number, induced emission does not yet dominate the recombination mechanism. To start an oscillation, induced emission must overcome all losses at threshold.
3. $J_S = 480\,\text{A}\,/\,\text{cm}^2$
4. $I = 200\,\text{kW}\,/\,\text{cm}^2$

[20] *Hint*: The impulse width is $\Delta t = 100\,\text{ps}$. The impulse's spectral width is $\Delta f = 1/\Delta t = 10\,\text{GHz}$ according to Eq. (2.60) on Page 28. Making use of the differential $\Delta f = -(c/\lambda^2)\Delta\lambda$, this corresponds to a wavelength span $\Delta\lambda = 0.080\,\text{nm} \gg 0.002\,\text{nm}$. The chromatic dispersion of the fibre is therefore not given by the unmodulated laser line width ($5\,\text{ps}\,/(\text{km nm}) \times 100\,\text{km} \times 0.002\,\text{nm} = 1\,\text{ps}$), but by the linewidth of the modulated field, namely $5\,\text{ps}\,/(\text{km nm}) \times 100\,\text{km} \times 0.080\,\text{nm} = 40\,\text{ps}$. From Eq. (2.61) on Page 29 the output impulse width $\sqrt{(100\,\text{ps})^2 + (40\,\text{ps})^2} = 107.7\,\text{ps}$ results.

5. $\Delta f_z = c/(2nL) = 100\,\text{GHz}$, $\Delta\lambda_z \approx (\lambda^2/c)\Delta f_z = 0.800\,\text{nm}$

6. $\Delta f_z = 4.3\,\text{THz}$, $\Delta\lambda_z = 34\,\text{nm}$. The typical gain bandwidth for semiconductor lasers is $\Delta f_H = 12\,\text{THz}$ (Eq. (3.66) on Page 78), so only one longitudinal mode can oscillate.

Solutions to quiz

1(a)(b)(c), 2(e), 3(a), 4(c), 5(c), 6(a), 7(c), 8(b), 9(c), 10(c)(d), 11(a),(b), 12(a)(c)(d)(e), 13(b)(c), 14(a), 15(c), 16(c), 17(b), 18(d), 19(b), 20(c)

4 Optical amplifiers

Solutions to problems

1. 1 mm for a 30 dB, 670 μm for 20 dB gain. Output power reduced from 1 mW to 100 μW.

2. 10 m for a 30 dB, 6.70 m for 20 dB gain. Output power reduced from 1 mW to 100 μW.

Solutions to quiz

1(a)(b)(c)(d), 2(a)(b), 3(a)(b)(d)(e), 4(b)(c), 5(c), 6(c)

5 Pin photodiode

Solutions to problems

1. $W = 1\,\mu\text{W} \times 1\,\text{ns} = 1\,\text{fJ}$, $N_P = P/(hf) \times 1\,\text{ns} = 7600$

2. $S = \eta e/(hf)$, $\eta = Shf/e = 1.08$. A quantum efficiency of $\eta > 1$ is unlikely!

3. $N_P = P/(hf) \times 1\,\text{ns} = 40\,404$, $i = SP$, $\eta = 0.773$, $N_n = i \times 1\,\text{ns}/e = 31\,250$, $N_n/N_P = 0.773$

4. $f_{\text{pin}}/f_{\text{APD}} = 100\,\text{ps}/5\,\text{ps} = 20$, $f_{\text{pin}} = 0.55/5\,\text{ps} = 110\,\text{GHz}$

5. Dark current smaller by a factor of 100 means the same reduction (by 20 dB inpower) in minimum optical input power, resulting in $-15\,\text{dBm} - 20\,\text{dBm} = -35\,\text{dBm}$.

6. See Fig. 5.6 on Page 113. For strong absorption all carriers are generated in an infinitely thin layer at the n-region side (left). The electrons disappear immediatly after their generation into the left contact.

Solutions to quiz

1(a), 2(a), 3(b), 4(b), 5(a), 6(d), 7(b)(d)

6 Noise

Solutions to problems

1. $\overline{|i_{RD}|^2} = 2e\bar{i}/100\,\text{ps} = (13\,\text{nA})^2$, $\gamma = (100\,\text{nA}/13\,\text{nA})^2 = 7.7^2 = 59$

2. $\overline{|i_Q|^2} = 4kT_0G_Q/100\,\text{ps} = (1.8\,\mu\text{A})^2$

Solutions to quiz

1(b), 2(a)(b)(c)(d), 3(a)(c)(d), 4(a)(d), 5(a)(d)(e), 6(b)(c)

7 Receivers and detection errors

Quiz

1(d), 2(d), 3(c), 4(d), 5(c), 6(c)(d), 7(b)(c), 8(b)(c)(d), 9(c)