Concepts of Optoelectronic Engineering

J. Mähnß

November 2, 2001

Contents

Part I

Elements of Optoelectronic Systems

Chapter 1

Light and Photons

For the description of optoelectronic devices and systems its convenient to have a mathematical representation of light. Unfortunately there exist three different physical models for the description namely the light beam or ray, the optical wave and photons. Each model covers a specific characteristic of the light but neglects other. So therefore we have to deal with all three models dependent on the situation. The basics of optical emission and absorption are discussed in the next section of this chapter followed by specific types of optical emission found in optoelectronics. Special emphasis is made on optical amplification due to its great importance.

1.1 Optical Beam, Wave, Particle

Optical Beam

The light description as an optical beam or so called ray-optics reflects our natural knowledge of light in reflection, refraction and imaging. Nothing can be concluded for the spatial resolution or interference and diffraction nor for interaction between light and matter. As an example for ray-optical treatment one may look at the Snellius reflection law. An optical ray travels with velocity $v = c_0/\eta$ through matter with refractive index η. At an interface between two materials with different η the wave

is refracted in such a way that the time for traveling from A to C is minimized as is shown in figure 1.1.

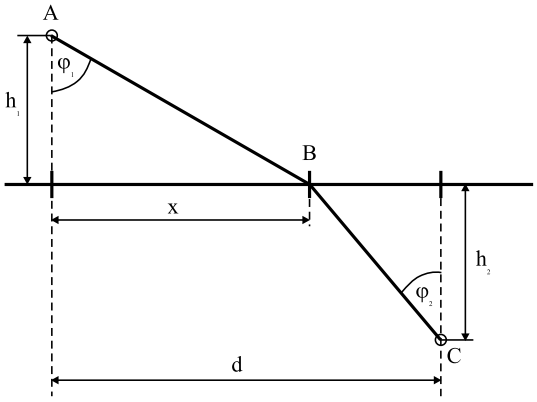

Figure 1.1: Optical path from point A to C for calculation of the minimal optical path length.

$$t = \frac{\overline{AB}}{v_1} + \frac{\overline{BC}}{v_2} = \eta_1 \frac{\sqrt{h_1^2 + x^2}}{c_0} + \eta_2 \frac{\sqrt{h_2^2 + (d-x)^2}}{c_0} \tag{1.1}$$

Minimal time ($\frac{\partial}{\partial x} t = 0$) gives with

$$\frac{\partial}{\partial x} t = \frac{\eta_1 x}{\sqrt{h_1^2 + x^2} \cdot c_0} + \frac{-(d-x) \cdot \eta_2}{\sqrt{h_2^2 + (d+x)^2} \cdot c_0} = \frac{1}{c_0} (\eta_1 \sin\{\varphi_1\} - \eta_2 \sin\{\varphi_2\}) = 0 \tag{1.2}$$

the Snellius law $\eta_1 \sin\{\varphi_1\} = \eta_2 \sin\{\varphi_2\}$.

Generally spoken the law of the shortest way can be expressed as the Fermats principle

$$t = \int \frac{1}{v\{s\}} ds = \frac{1}{c_0} \int \eta\{s\} ds = \frac{\mathrm{OPL}}{c_0} \tag{1.3}$$

with the optical path length

$$\mathrm{OPL} = \int \eta\{s\} ds \tag{1.4}$$

that must be minimal for a given ray or in other words:

the optical path length must be stationary with respect to variations of the path.

Optical Wave

When light is regarded as an optical wave one has to use the Maxwell-equations to solve the problems associated with the waves. Nearly everything like reflection and refraction, imaging, spatial resolution and interference - and diffraction - effects can be described. Only the interaction with matter for the understanding of absorption and emission is not covered. Two examples for useful application of the wave model are the Young's double - slit experiment and the spatial resolution of an optical instrument.

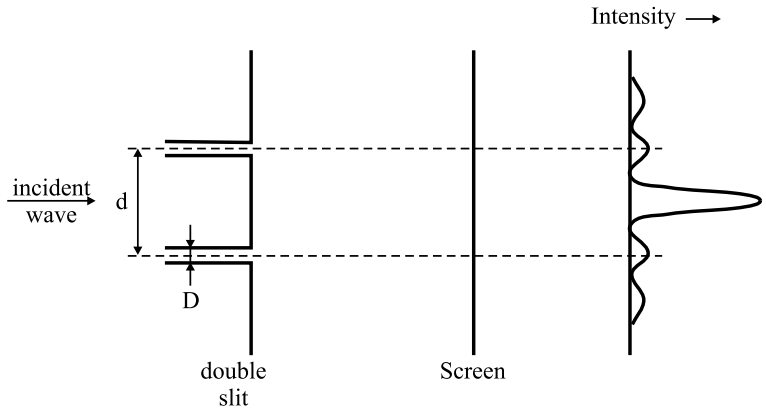

Figure 1.2: Young's double slit experiment. An incident wave does not produce two bright spots behind the slits as expected from ray optics but one just in the middle between them.

In the Young's experiment maximum intensity occurred not in a straight line after the slits as one would argue from ray - optics but just in the middle between them as sketched in figure 1.2.

The spatial resolution of a coherent illuminated object was first given by E. Abbe who said that it is necessary to take not only the zero but as well the first diffraction order of the object to see first details in a microscope. For a grid with period Λ the first order diffraction occurs under an angle of $\sin\{\varphi\} = \frac{\lambda}{\Lambda}$ and for small lens diameter d it follows $\sin\{\varphi\} \simeq \tan\{\varphi\} \leq \frac{d}{2f}$ resulting into the possibility to resolve two objects with distance $\Lambda \geq \lambda \cdot \frac{2f}{d}$.

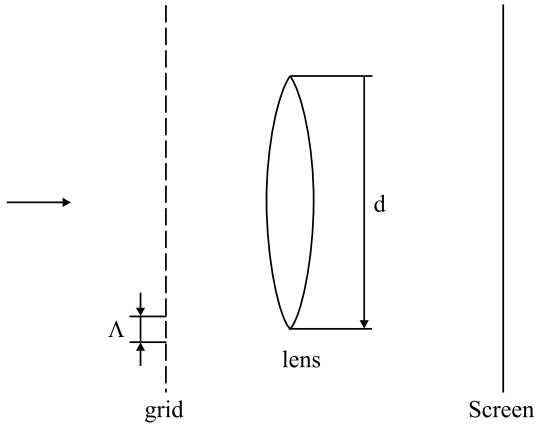

Figure 1.3: Abbe's experiment for spatial resolution of an optical instrument.

Photons

Both the ray and the wave-optics give no information about interaction with matter. The particle model of light can provide this where this is only the aspect it covers. The light particles are called photons. They carry energy $W = \hbar\omega = h\nu = h \cdot \frac{c_0}{\lambda}$ and have impulse $p = \hbar k = \frac{h}{\lambda}$. Everytime when light is absorbed or emitted it is regarded as a particle with energy and impulse that interacts with the matter surrounding it by exchanging energy and impulse.

1.2 Emission and Absorption of Light

For the use of material in a laser it must provide gain for the light to be produced. Basic information about the amount of possible gain and its spectral distribution can be achieved from its absorption due to the fact that absorption can be understood as inverse light emission.

The processes of emission and absorption can basically be modeled by a black body in thermal equilibrium. Planck found a description for the irradiation from a small hole in that body based on physical measurements. Einstein gave an energy -level description and the comparison with Planck's formula results into the well known laws for emission and absorption.

According to Einstein one can understand the processes of absorption and emission as

transition between two energy levels W_1 and $W_2 > W_1$ as sketched in figure 1.4.

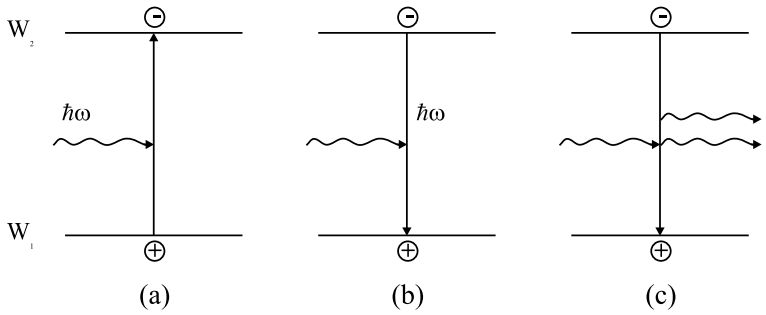

Figure 1.4: (a) Incident energy $\hbar\omega$ cause carrier separation into energy levels W_1 and $W_2 = W_1 + \hbar\omega$ under absorption of the incident energy . (b) Spontaneous recombination of excited carriers leaves Energy $\hbar\omega = W_2 - W_1$ as emission of radiation . (c) under stimulation with energy $\hbar\omega$ excited carriers recombine earlier as they would have done. Their energy $\hbar\omega$ is emitted and added to the incident stimulation. Not all recombination processes lead to an emission of radiation (d), i.e. sidewall recombinations.

The absorption process is depicted in figure 1.4(a) where free carriers are generated under absorption of incident energy $\hbar\omega$. The free carriers recombine eventually under emission of radiation with energy $\hbar\omega$ giving the name spontaneous emission. When the carriers are caused to recombine earlier under an incident energy $\hbar\omega$ adding their emission to the exitation the process is called stimulated emission.

1.2.1 Blackbody Irradiation

The radiation of a hollow black body through a small hole is carried by photons. The photons itself are placed in resonator modes of the black body. This means that the number of modes equals the number of states that photons can occupy. The number of photons in a modes depends at a given temperature on the photon energy which determines the probability that the mode is occupied. The total number of photons therefore is determined by the number of photon states (number of modes) times probability of state (mode) occupation. The energy flux through the hole is directly

proportional to the energy density inside the black body. Therefore it is convenient to use densities instead of numbers to describe the general behaviour.

The modes in a resonator are well described by their k- vectors. The number of modes in a resonator is easily found as a ratio between occupied wavenumbers and total available wavenumbers. Both are expressed in the form of k- volumes. Considering the black body to be a rectangular resonator with ideal metallic walls the total k- volume at circular frequency ω is given by the spherical shell of radius $k = \frac{\eta}{c_0}\omega$ in the k -space resulting into a volume of $V_\mathrm{k} = \frac{4}{3}\pi k^3 = \frac{4\pi}{3}(\frac{\eta\omega}{c_0})^3\frac{32\pi^4}{3}(\frac{\eta\nu}{c_0})^3$.

The volume occupied by resonator modes can be estimated by the possible propagation vector components that follow for a rectangular resonator of thickness d, width b and length L to $\vec{k} = \pi\left(\frac{m_\mathrm{x}}{d}, \frac{m_\mathrm{y}}{b}, \frac{m_\mathrm{z}}{L}\right)^T$ where m_x, m_y and m_z are defined as positive integers and give the mode order. The spacing between two adjacent modes with $\Delta m_\mathrm{x} = \Delta m_\mathrm{y} = \Delta m_\mathrm{z} = 1$ is $\Delta\vec{k} = \pi\left(\frac{1}{d}, \frac{1}{b}, \frac{1}{L}\right)^T$. The distance between neighboring modes in k-space is very small and we assume that all k-values between the modes are also acceptable. With these assumptions one mode occupies the volume

$$V'_\mathrm{k,mode} = \frac{\pi^3}{bdL} = \frac{\pi^3}{V_\mathrm{Res}}$$

in the positive k-space as sketched in figure 1.5.

oe2f6816.eps

The number of modes follows from the ratio of occupied k volume to total k volume. The occupied k volume is defined only in the positive octand and therefore it has to be expanded to the total space by multiplication with a factor of 8. Generally the modes are present in two polarizations that have to share the occupied k- volume resulting into an occupied volume of

$$V_\mathrm{k,mode} = \frac{8}{2}V'_\mathrm{k,mode} = \frac{4\pi^3}{V_\mathrm{Res}} = \frac{(2\pi)^3}{2V_\mathrm{Res}} \quad . \tag{1.5}$$

The number of modes follows as

Figure 1.5: Discrete values of the propagation constant for modes of a rectangular resonator. The whole k-space is filled. It can be assumed that the whole k- space is filled because the modes are closely neighbored. Therefoe a volume with dimensions $\Delta k_{\mathrm{x}}, \Delta k_{\mathrm{y}}, \Delta k_{\mathrm{z}}$ centered around the discrete values is occupied as sketched on the right side.

$$M = \frac{V_\mathrm{k}}{V_\mathrm{k,mode}} = \frac{V_\mathrm{Res}}{3\pi^2}(\frac{\eta\omega}{c_0})^3 = \frac{8\pi\eta^3\nu^3}{3c_0^3}V_\mathrm{Res}$$

resulting into a density of modes $N_\mathrm{mode} = \frac{M}{V_\mathrm{Res}} = \frac{8\pi\eta^3\nu^3}{3c_0^2}$ which leads to the spectral mode density $D_\mathrm{mode}\,\mathrm{d}\nu = \frac{\partial}{\partial\nu}N_\mathrm{mode}\,\mathrm{d}\nu$.

The mode density equals directly the density of photon states

$$D_\mathrm{phot}\{\nu\}\,\mathrm{d}\nu = D_\mathrm{mode}\{\nu\}\,\mathrm{d}\nu = \frac{8\pi\eta^3\nu^2}{c_0^3}\,\mathrm{d}\nu$$

or in terms of energy with $\mathrm{d}\nu = \frac{1}{2\pi}\,\mathrm{d}\omega = \frac{1}{2\pi\hbar}\,\mathrm{d}(\hbar\omega)$

$$D_\mathrm{phot}\{\hbar\omega\}d(\hbar\omega) = \frac{\eta^3(\hbar\omega)^2}{\pi^2\hbar^3c_0^3}\,\mathrm{d}(\hbar\omega) \quad .$$

Regarding that η is frequency dependent the spectral photon density follows from

$$D_\mathrm{phot}\{\nu\}\,\mathrm{d}\nu = \frac{8\pi\eta^2\nu^2}{c_0^3}\eta_\mathrm{gr}\,\mathrm{d}\nu \tag{1.6}$$

with the group index $\eta_\mathrm{gr} = \frac{\mathrm{d}}{\mathrm{d}\nu}(\eta\nu) = \eta + \nu\frac{\mathrm{d}}{\mathrm{d}\nu}\eta$.

Often the mode (= photon state) density is required in terms of wavenumbers instead of frequency as given in (1.6). The number of modes that fill a volume $V_\mathrm{k} = \mathrm{d}^3k$ is $\mathrm{d}^3M = V_\mathrm{k}/V_\mathrm{k,mode}$ and the mode density is expressed by

$$D_\mathrm{k}\,\mathrm{d}^3k = \frac{\mathrm{d}^3M}{V_\mathrm{Res}} = \frac{2}{(2\pi)^3}\,\mathrm{d}^3k \quad . \tag{1.7}$$

An ensemble of m photons carrys the energy $W_m = m\cdot\hbar\omega$. The probability that they occupy the same state follows the Maxwell-Boltzmann distribution $p_\mathrm{B} \sim \exp\{-\frac{W_m}{k_\mathrm{B}T}\} = \exp\{-m\frac{\hbar\omega}{k_\mathrm{B}T}\}$ resulting into an average number of photons per mode

$$f_{\text{phot}}\{\hbar\omega, T\} = \frac{\sum_{m=0}^{\infty} m \cdot p_{\text{B}}}{\sum_{m=0}^{\infty} p_{\text{B}}} = \frac{\sum_{m=0}^{\infty} m \cdot \exp\{-m\frac{\hbar\omega}{k_{\text{B}}T}\}}{\sum_{m=0}^{\infty} \exp\{-m\frac{\hbar\omega}{k_{\text{B}}T}\}} \quad .$$

Substituting $\exp\{-\frac{\hbar\omega}{k_{\text{B}}T}\} = b$ it can easily be seen that the denominator is a simple geometrical series with result $\sum_{m=0}^{\infty} b^m = \frac{1}{1-b}$. The nominator can be written as

$$\sum_{m=0}^{\infty} m \cdot b^m = b \cdot \sum_{m=0}^{\infty} \frac{\partial}{\partial b} b^m = b\frac{\partial}{\partial b} \sum_{m=0}^{\infty} b^m = b \cdot \frac{\partial}{\partial b} \frac{1}{1-b} = \frac{b}{(1-b)^2}$$

resulting into

$$f_{\text{phot}}\{\hbar\omega, T\} = \frac{b}{1-b} = \frac{1}{b^{-1}-1} = \frac{1}{\exp\{\frac{\hbar\omega}{k_{\text{B}}T}\}-1} \tag{1.8}$$

which is the well known Bose-Einstein-Distribution leading to the classification of photons as bosons. The spectral and spacial density of photons ϱ_{phot} at temperature T in an energy interval of $d(\hbar\omega)$ around $\hbar\omega$ results from the density of states and the probability of their occupation with

$$\begin{aligned} \mathrm{d}N_{\text{phot}} = \varrho_{\text{phot}}\{\hbar\omega\}\,\mathrm{d}(\hbar\omega) &= D_{\text{phot}}\{\hbar\omega\} \cdot f_{\text{phot}}\{\hbar\omega, T\}\,\mathrm{d}(\hbar\omega) \\ &= 8\pi\frac{\eta_{\text{gr}}\eta^2\nu^2}{c_0^3}\frac{1}{\exp\{\frac{\hbar\omega}{k_{\text{B}}T}\}-1}\,\mathrm{d}(\hbar\omega) \\ &= 2\pi\frac{\eta_{\text{gr}}}{c_0}\left(\frac{\eta}{\hbar c}\right)^2\left(\frac{\hbar\omega}{\pi}\right)^2\frac{1}{\exp\{\frac{\hbar\omega}{k_{\text{B}}T}\}-1}\,\mathrm{d}(\hbar\omega) \quad . \end{aligned}$$

Regarding the energy $h \cdot \nu$ of each photon the energy-density is $u_{\text{phot}}\{\nu\}\,\mathrm{d}\nu = h\nu\varrho_{\text{phot}}\{h\nu\}\,\mathrm{d}\nu$

$$u_{\text{phot}}\{\nu\}\,\mathrm{d}\nu = 8\pi h\nu\frac{\eta_{\text{gr}}}{c_0}(\frac{\eta\nu}{c_0})^2\frac{1}{\exp\{\frac{\hbar\omega}{k_{\text{B}}T}\}-1}\,\mathrm{d}\nu \quad . \tag{1.9}$$

This is the spectral energy density that is irradiated through the small hole in the black body.

1.2.2 Einstein Relations

Einstein treated the act of spontaneous emission of light as transition of carriers with rate r_{spon} from energy level W_2 to $W_1 = W_2 - \hbar\omega$. This transition is only possible if there are states $D_2 = D_{\text{carr}}\{W_2\}$ that are occupied with probability $f_2 = f_{\text{carr}}\{W_2\}$ at level W_2 and free states $D_1(1 - f_1)$ with $D_1 = D_{\text{carr}}\{W_1\}$ and $f_1 = f_{\text{carr}}\{W_1\}$ at level W_1 resulting into a spontaneous emission rate of $r_{\text{spon}} = \frac{\partial}{\partial t}\varrho_{\text{phot,spon}}$ that is proportional to the spontaneous transition rate of carriers from W_2 to W_1

$$\begin{aligned} r_{\text{spon}}\,\mathrm{d}\hbar\omega &= AD_{\text{carr}}\{W_2\}f_{\text{carr}}\{W_2\}D_{\text{carr}}\{W_1\}(1 - f_{\text{carr}}\{W_1\}) \\ &= AD_2 f_2 D_1(1 - f_1)\,\mathrm{d}\hbar\omega \quad . \end{aligned}$$

Absorption and stimulated emission as well need filled states at W_1 (resp. W_2) and empty states at W_1 (W_2) and additionally only take place in the presence of photons. Therefore the photon density has to be regarded with

$$\begin{aligned} r_{\text{stim}}\,\mathrm{d}\hbar\omega &= B_{21}D_2 f_2 D_1(1 - f_1)\varrho_{\text{phot}}\,\mathrm{d}\hbar\omega \\ r_{\text{abs}}\,\mathrm{d}\hbar\omega &= B_{12}D_1 f_1 D_2(1 - f_2)\varrho_{\text{phot}}\,\mathrm{d}\hbar\omega \quad . \end{aligned} \tag{1.10}$$

The constants A, B_{12} and B_{12} are called Einstein coefficients and indicate the proportionality between crrier transition and photon generation. Under thermal equilibrium the number of transitions with subsequent photon generation and absorption equal each other $r_{\text{abs}} = r_{\text{spon}} + r_{\text{stim}}$ leading to a photon density of

$$\frac{1}{\varrho_{\text{phot}}\,\mathrm{d}\hbar\omega} = \frac{B_{12}D_1 D_2 f_1(1 - f_2)}{AD_1 D_2 f_2(1 - f_2)\,\mathrm{d}\hbar\omega} - \frac{B_{21}}{A\,\mathrm{d}\hbar\omega} \quad .$$

The probability that a carrier occupies a state at energy W is described by the Fermi-distribution

$$f\text{carr} = \frac{1}{\exp\{\frac{W - W_{\text{F}}}{k_{\text{B}}T}\} + 1}$$

Taking $W_2 = W_1 + \hbar\omega$ one finds with the Fermi- levels W_{F1} and W_{F2} for the energy levels W_1 and W_2 respectively

$$\varrho_{\mathrm{phot}}\,\mathrm{d}\hbar\omega = \frac{A}{B_{12}\exp\{\frac{\hbar\omega-(W_{\mathrm{F2}}-W_{\mathrm{F1}})}{k_{\mathrm{B}}T}\} - B_{21}}\,\mathrm{d}\hbar\omega \quad .$$

In the black body all Fermi levels are equal $W_{\mathrm{F1}} = W_{\mathrm{F2}} = W_{\mathrm{F}}$ and comparison with the photon density of a black body (1.9) gives

$$B_{21} = B_{12} \tag{1.11}$$

and

$$A = \left(\frac{\eta}{\hbar c}\right)^3 \left(\frac{\hbar\omega}{\pi}\right)^2 B_{21} \tag{1.12}$$

which says that the transition rates for absorption and stimulated emission equal each other and the probability of spontaneous emission increases quadratic with the emitted energy compared to the stimulated emission indicating the problems associated with the realization of high- photon- energy lasers.
The maximal absorption rate is

$$\max\{r_{\mathrm{abs}}\,\mathrm{d}\hbar\omega\} = B_{12}D_1D_2\varrho_{\mathrm{phot}}\,\mathrm{d}\hbar\omega$$

in the case $f_1 = 1$ and $f_2 = 0$ mich means that all states at level W_1 are occupied and no at W_2. The maximum value for stimulated emission in turn is

$$\max\{r_{\mathrm{abs}}\,\mathrm{d}\hbar\omega\} = B_{12}D_1D_2\varrho_{\mathrm{phot}}\,\mathrm{d}\hbar\omega$$

for the case that all states at level W_2 are occupied $f_2 = 1$ and no state $W_1(f_1 = 0)$. With (1.12) both values are the same. The gain in a medium is the net rate $r_{\mathrm{stim}} - r_{\mathrm{abs}}$

in the case $r_{\mathrm{stim}} \gg r_{\mathrm{spon}}$. Due to the fact that the maximal values of r_{stim} and r_{abs} are the same it can be concluded that high gain can only be expected from materials with strong absorption. Additionally it can be concluded that the maximum gain has the same value as the maximum absorption giving the possibility to evaluate materials by a passive absorption measurement for the use as active material in a laser.

When one of the Einstein coefficients is known the other ones can be calculated. A good estimation for the spontaneous emission factor A can be obtained by the investigation of weakly pumped material. Pumping means that energy is transferred to carriers such that they are lifted into a higher energy level. When pumping is switched off some excess carriers exist and recombine spontaneous under emission of light. The amount of emitted light is therefore a direct measure of the excess carriers. The decay follows $N_{\mathrm{phot}}\{t\} = N_{\mathrm{phot}}\{t=0\}\exp\{-\frac{t}{\tau_{\mathrm{s}}}\}$ resulting into $\frac{\mathrm{d}}{\mathrm{d}t}N_{\mathrm{phot}} = N_{\mathrm{phot}}/\tau_{\mathrm{s}}$ with recombination- lifetime τ_{s}. On the other hand the spontaneous recombination rate can be written as $r_{\mathrm{spon}} = AN_{\mathrm{hole},1}N_{\mathrm{el},2}$ with the steady state density of electrons $n_2 = N_{\mathrm{el},2} = D_2 f_2$ and of holes $p_1 = N_{\mathrm{hole},1} = D_1(1-f_1)$ taking holes as usual as missing electrons. The exitation generates equal amounts of excess carriers $\Delta p_1 = \Delta n_2 = \Delta n$ at both energy levels W_1 and $W_2 = W_1 + \hbar\omega$. The increased spontaneous recombination rate is then $r_{\mathrm{spon}} = A(p_1+\Delta n)(n_2+\Delta n) = A[p_1 n_2 + \Delta n(p_1+n_2+\Delta n)]$. The steady state spontaneous emission rate is $r_{\mathrm{spon}} = Ap_1 n_2$. The emission decay is directly proportional to the spontaneous recombination of excess carriers $A\Delta n(p_1 + n_2 + \Delta n)$.

The comparison

$$\frac{\partial}{\partial t}N_{\mathrm{phot}} = -\frac{1}{\tau_{\mathrm{s}}}N_{\mathrm{phot}} \propto r_{\mathrm{spon}} \simeq A\Delta n(p_1 + n_2 + \Delta n)$$

results into the fact that the lifetime corresponds directly to A with

$$A = \frac{1}{\tau_{\mathrm{s}}(p_1 + n_2 + \Delta n)} \simeq \frac{1}{\tau_{\mathrm{s}}(p_1 + n_2)} \quad . \tag{1.13}$$

1.2.3 Absorption and Spontaneous Emission of Light

For the use as gain material it is of some interest in which spectral range emission can be expected. This information is provided by the spectral absorption distribution. When an optical beam travels in z- direction through a medium with absorption it undergoes a loss in irradiance according to

$$\frac{\mathrm{d}}{\mathrm{d}z}I = -\alpha I \tag{1.14}$$

or

$$I = I\{z=0\}\exp\{-\alpha z\}$$

where α is the absorption constant. Remembering that the irradiance is spectral energy density times speed (group velocity in the case of time harmonic fields) of the wave $I = h\nu\varrho_{\mathrm{phot}}c_{\mathrm{gr}}$ the space derivative of I can be transformed into a time derivative of the spectral photon-density

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\varrho_{\mathrm{phot}} &= \frac{\eta_{\mathrm{gr}}}{\hbar\omega c_0}\frac{\mathrm{d}}{\mathrm{d}t}I = \frac{\eta_{\mathrm{gr}}}{\hbar\omega c_0}\frac{\mathrm{d}}{\mathrm{d}t}z\frac{\partial}{\partial z}I \\ &= -\alpha\frac{I}{\hbar\omega} = -\alpha\frac{c_0}{\eta_{\mathrm{eff}}}\varrho_{\mathrm{phot}}\end{aligned}$$

where we used $\frac{\mathrm{d}}{\mathrm{d}t}z = \frac{c}{\eta_{\mathrm{eff}}}$ and (1.14). This must be the same as the difference between the transition rates of absorption and stimulated emission. Using Einstein relations this means

$$\begin{aligned}r_{\mathrm{stim}} + r_{\mathrm{spon}} - r_{\mathrm{abs}} &= \frac{\mathrm{d}}{\mathrm{d}t}\varrho_{\mathrm{phot}} = -\alpha\frac{c_0}{\eta_{\mathrm{eff}}}\varrho_{\mathrm{phot}} \\ &= B_{12}D_1D_2(f_2-f_1)\varrho_{\mathrm{phot}} + AD_1D_2f_2(1-f_1)\end{aligned} \tag{1.15}$$

or

$$\alpha = B_{12}D_1D_2(f_2 - f_1)\frac{\eta_{\text{eff}}}{c_0} - AD_1D_2f_2(1 - f_1)\frac{\eta_{\text{eff}}}{c_0}\frac{1}{\varrho_{\text{phot}}} \quad .$$

The contribution from spontaneous emission to the absorption coefficient decreases with growing photon density and is neglected therefore giving

$$\alpha \simeq B_{12}D_1D_2(f_2 - f_1)\frac{\eta_{\text{eff}}}{c_0} \quad .$$

With 1.12 B_{12} can be replaced by the spontaneous emission coefficient A and

$$\begin{aligned}\alpha\{\hbar\omega\} &= Acc_{\text{gr}}\frac{\hbar^2}{2\pi}\left(\frac{\pi}{\eta\hbar\omega}\right)^2 D_1D_2(f_1 - f_2) \\ &= Acc_{\text{gr}}\frac{\hbar^2}{2\pi}\left(\frac{\pi}{\eta\hbar\omega}\right)^2 D_{\text{el}}\{W_1\}D_{\text{el}}\{W_1 + \hbar\omega\}(f_{\text{el}}\{W_1\} - f_{\text{el}}\{W_1 + \hbar\omega\}) \quad (1.16)\end{aligned}$$

follows. When f_1 is less than f_2 at a given photon energy $\hbar\omega$ then α is negative and amplification occurs. This can be understood in that way that more stimulated emission than absorption is present. The power of an incident wave increases as it travels through the medium indicating gain in the material.

The case that the population probability of states in the upper level is bigger than that in the lower level is called **population inversion**. Population inversion is the key to gain as has been shown above.

For the required relation between absorption and spontaneous emission A in (1.16) is substituted by the spontaneous emission rate which leads to

$$\begin{aligned}r_{\text{spon}}\{\hbar\omega\} &= \alpha\{\hbar\omega\}\frac{2\pi}{\hbar^2}\left(\frac{\hbar\omega}{\pi}\right)^2\frac{1}{cc_{\text{gr}}}\frac{(1 - f_{\text{el}}\{W_1\})f_{\text{el}}\{W_1 + \hbar\omega\}}{f_{\text{el}}\{W_1\} - f_{\text{el}}\{W_1 + \hbar\omega\}} \\ &= \alpha\{\hbar\omega\}\frac{2\pi}{\hbar^2}\left(\frac{\hbar\omega}{\pi}\right)^2\frac{1}{cc_{\text{gr}}}\frac{1}{\exp\left\{\frac{\hbar\omega + W_{\text{F2}} - W_{\text{F1}}}{k_{\text{B}}T}\right\} - 1} \quad .\end{aligned}$$

In thermal equilibrium ($W_{\text{F1}} = W_{\text{F2}}$) this simplifies to the van Roosboek-Shockley relation

$$r_{\text{spon}}\{\hbar\omega\} = \frac{2\pi}{\hbar^2}\left(\frac{\hbar\omega}{\pi}\right)^2 \frac{1}{cc_{\text{gr}}}\alpha\{\hbar\omega\}\frac{1}{\exp\{\frac{\hbar\omega}{k_{\text{B}}T}\}-1}$$

saying that strong spontaneous emission only can be expected at photon energies where strong absorption exists and vice versa. This is not the complete truth as we will see in section 1.3.1 because we have to take into account the so called Stokes-shift but it gives a good idea what can be expected.

1.3 Luminescence

Luminescence is a general term used to describe the emission of radiation from matter when it is supplied with some form of energy thus generating exited states. Dependent on the form of energy supply it is common to distinguish between three types of luminescence namely

- Photoluminescence: exitation arises from the absorption of photons
- Cathodoluminescence: bombardement with electrons i.e. an electron beam delivers the energy for exitation
- Electroluminescence: application of an electric field (a.c. or d.c.) causes carriers to change their energy state.

Whatever form of energy input W_{pump} is present the final stage in the process is an electronic transition between two energy levels W_1 and W_2 ($W_2 > W_1$) with emission of radiation $\hbar\omega = \frac{hc}{\lambda} = W_2 - W_1 \leq W_{\text{pump}}$. The energy levels normally belong to a band of energies and therefore a band of wavelengths instead of a single wavelength is emitted or in other words instead of a single line a certain spectrum is emitted. As we saw in the section before this is strongly dependent on the special material where the exitation occurred. Normally one would expect the luminescence to diminish in a short time after the energy supply has been switched off. In this case we speak of **fluorescence**.

If the decay time is much longer than expected the phenomenon is called **phosphorescence**. This behaviour was first found in some phosphor and therefore all matters exhibiting phosphorescence are commonly called phosphors. To understand this process we may look at the energy level in figure 1.6.

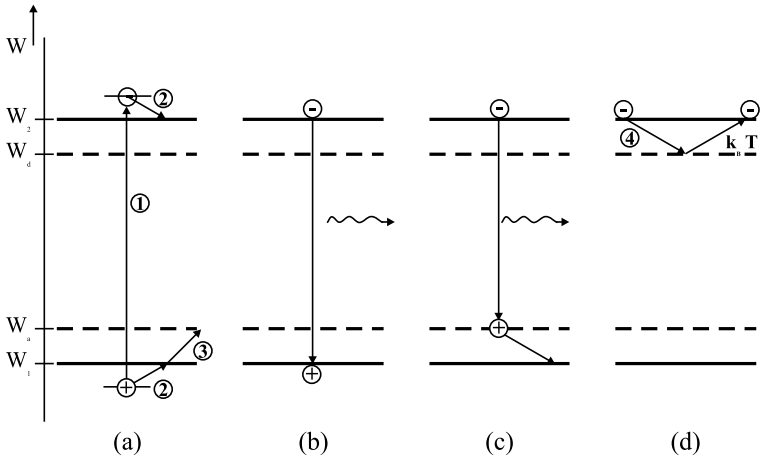

Figure 1.6: (a): Energy is transferred to carriers causing them to separate (1). The carriers relax into the levels W_1 and W_2 by interaction with the surrounding matter(2). Eventually holes are trapped into states W_a just above W_1 (3). (b) and (c): The exited carriers recombine under emission of radiation with radiation-energy $\hbar\omega = W_2 - W_1$ or $\hbar\omega = W_2 - W_\mathrm{a}$. (d): The exited electron may be trapped for a while in the state W_d just below W_2 and then thermally reactivated (4). Emission occurs as in (b) or (c). This process is characteristic for phosphorescence.

The energy supply takes electrons from their bounds and shifts them to a higher level leaving a hole (ionized rest-molecule or -atom). In figure 1.6 this is indicated as path (1). Holes and electrons interact with the surrounding matter and relax under energy-transfer to the matter into the states W_1 and W_2 (path 2). In some cases there are impurities present generating levels W_a and W_d near W_1 and W_2. Both electrons and holes can be trapped into the impurity- states. Only free carriers in W_1 or W_2 can travel and when coming into the vicinity of the corresponding carrier they recombine under emission of energy $\hbar\omega$. Due to the higher mass of the holes they are more

likely to be trapped and therefore only electrons determine the time of recombination. Normally this occurs in the region of about 10^{-8} s. If the electrons are trapped they have to be reactivated thermally (path (4)) before recombination occurs. This is a very slow process leading to a long decay time of luminescence. In this case we speak of phosphorescence. The decay- times are temperature-dependent and can range from minutes to days.

1.3.1 Photoluminescence

In photoluminescence exited carriers are generated by absorption of photons. Remembering the Roosboek-Shockley relation one would argue the emission of the spectrum to be nearly the same as the absorption spectrum, especially there should be no emission where no absorption occurs. When we look at the spectra we nearly always find a shift to lower energies (bigger wavelengths), a so called red-shift. This can be explained by different energy levels of exited and non-exited media. Let the non-exited medium have energy-levels W_1 and W_2 as depicted in figure 1.7. After exitation the structure is changed. In a simple model one can think of an electron lifted to an outer shell of an atom. Now the atom tries to reach the state of minimum energy which is achieved through some vibrations (phonon generation). In the final state the electrons are somewhat nearer to the atom-core resulting into energies W_1' and W_2' separated $\hbar\omega' < \hbar\omega$. This red-shift between absorption and emission is called Stokes-shift.
A typical application for photoluminescence is the mercury-vapor bulb. An electrical discharge causes the mixture of mercury vapor and noble gas to emit radiation. The ultra-violet (UV) part is changed into the visible spectrum by the bulb-coating that exhibits a remarkable Stokes-shift.

1.3.2 Cathodoluminescence

In cathodoluminescence the carrier-exitation is achieved by an electron-beam. The emission process is the same as described for photoluminescence. Empirically has been

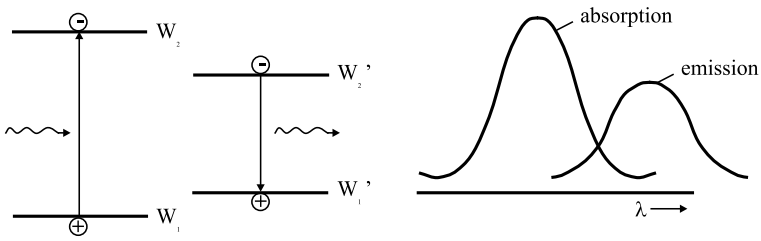

Figure 1.7: (a) and (b): The non-exited medium exhibits Energy levels W_1 and W_2 due to its structure. After exitation the structure is changed and the medium relaxes to new energy-levels W_1' and W_2' separated somewhat closer (c): Absorption and emission spectra due to the Stokes-shift.

found that only about 30% of the electron-beam energy leads to electron-hole pair generation when the beam-energy is not too high and the exiting electrons do not penetrate too much into the luminescent material. At higher electron-beam energies the generation efficiency decreases remarkably. Although the cathodoluminescence has that low efficiency it is widely used because electron beams can be generated and manipulated very easy. The by far most known example for cathodoluminescence occurs in cathode-ray tubes that are installed in most TV- and computer monitors today.

1.3.3 Electroluminescence

Electroluminescence can be dealt as the 'classical electroluminescence' and the modern type called 'injection electroluminescence' as it is found in LED's and Lasers.

1.3.3.1 Classical Electroluminescence

If an electroluminescent material is placed between two electrodes and a.c. or d.c. voltage is turned on some luminescence can be observed. As shown in figure 1.8 carrier separation takes place by tunneling through the bandgap (1) of that material. The tunneling is enhanced under microwave activation due to phonon generation that helps

tunneling. Also field emission, i.e. from sharp edges or spikes on electrodes, enhance the carrier exitation. The free carriers travel (2) in the field and when they find a partner they recombine (3) under emission of radiation.

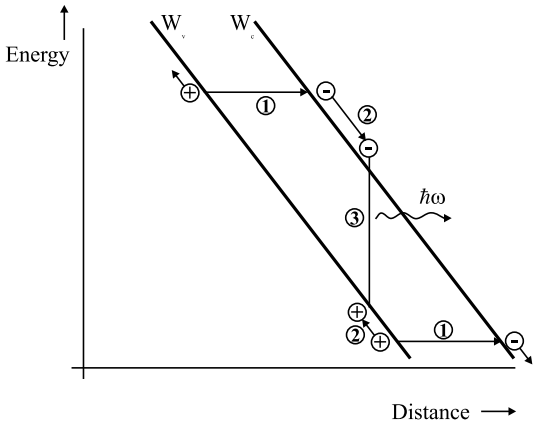

Figure 1.8: Generation of free carriers in an electrical field by means of spatial tunneling through the energy gap.

Another possible explanation is the Avalanche-effect where already free carriers gain that much energy by movement in the field that they are able to create new free carriers by collision which can recombine again under radiation of emission as sketched in figure 1.9. The output power of electroluminescent devices is not too high. For possible commercial applications some research has be done to improve the performance of such devices.

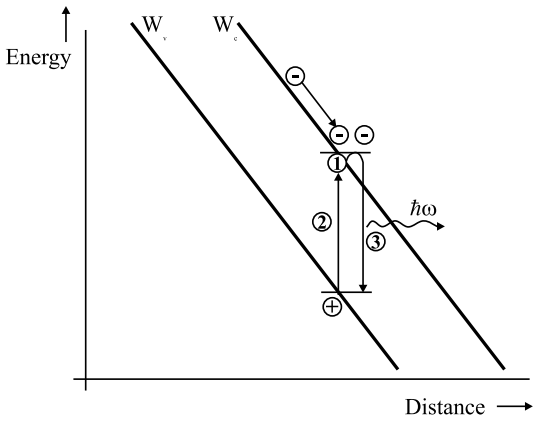

Figure 1.9: A free electron gathers that much energy under acceleration (1) in the electric field that at a collision with a bound electron a free carrier pair is generated (2). This pair recombines (3) under emission of radiation.

1.3.3.2 Injection Luminescence

In modern optoelectronic devices electroluminescence is generated by injection of carriers into a hetero-pn-junction as sketched in figures 1.10 and 1.11 for a simple (homo) pn-junction and a hetero-pn-junction in thermal equilibrium and under forward bias.

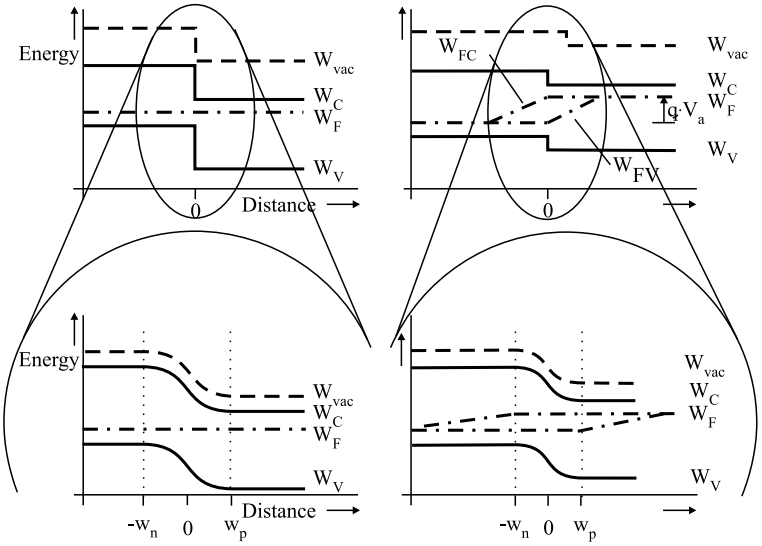

Figure 1.10: Energy diagram of a pn-junction made in homogeneous material. On the left hand side the junction is under equilibrium conditions, on the right hand side a forward bias voltage V_a is applied leading to the reduced depletion region width and nearly flat bands.

We assume the metallurgical interface and the pn- doping interface to be at the same position. The spatial energy distribution calculation follows the same concepts as is known from classical textbooks about semiconductors. Only the dielectric displacement changes at the metallurgical interface due to different dielectric constants. This leads to a sharp bend in the diffusion potential V_d at the metallurgical interface. As usual the vacuum potential (the sum of flat vacuum level and build in diffusion potential) is assumed to be continuous and in equilibrium the Fermi-level W_F is flat over the whole junction. Proposing material dependent electron affinities this leads to the band-bending as sketched in figure 1.11 with the band- discontinuities in the

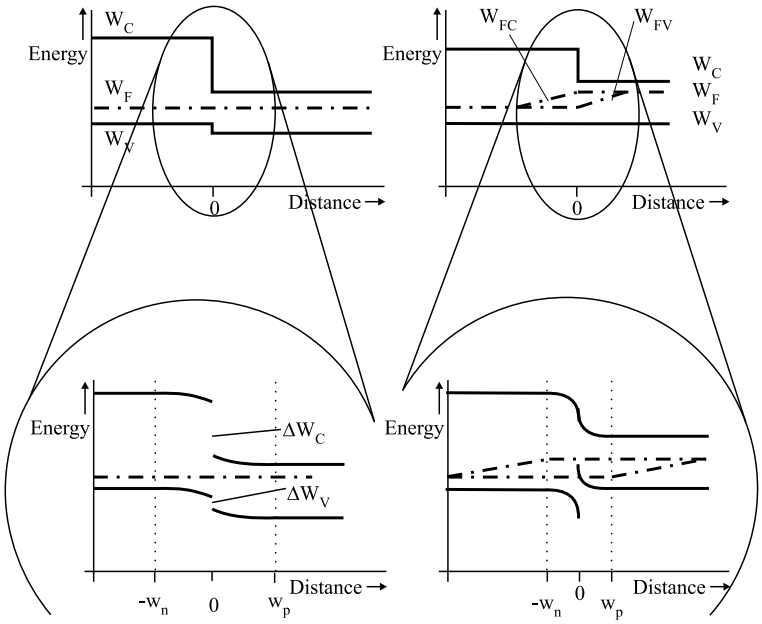

Figure 1.11: Energy diagram for a pn-heterojunction without and with applied forward biased voltage. The general shape is the same as in a homojunction except the discontinuities in valence- and conduction- bandedges ΔW_{v} and ΔW_{c} that remain independently of the applied voltage.

heterostructure. Under forward bias V_a the diffusion potential is partly compensated to $V_d - V_a$ changing the carrier distribution and therefore the band-bending. Calculating the carrier densities at the edges of the depletion layer it is found for the p- side $N_{hole}\{-w_p\} = N_A^+$ and $N_{electron}\{-w_p\} = \frac{n_{ip}^2}{N_A^+}\exp\{\frac{qV_a}{k_BT}\}$ and on the n- side $N_{electron}\{w_n\} = N_D^-$ and $N_{hole}\{w_n\} = \frac{n_{in}^2}{N_D^-}\exp\{\frac{qV_a}{k_BT}\}$. In both cases the concentration of minority carriers is enhanced exponentially with the applied voltage. Taking into account that the intrinsic carrier densities are directly related to the bandgap by $n_i^2 = D_C D_V \exp\{-\frac{W_g}{k_BT}\}$ in heterojunctions is found that in the material with smaller energy gap the minority carrier density is enhanced compared to the level found in on the other side. This looks like an injection of minority carriers from the wide-gap material into the low-gap material. From the edges of the depletion zone the carrier density relaxes to its normal value in the neutral region following a diffusion and recombination process. The complicated dependence of free minority and majority carriers can be modeled by the assumption that the Fermi-level splits into two quasi Fermi-levels W_{FV} for the valence band and W_{FC} for the conduction band as already sketched in figures 1.10 and 1.11. Remembering the formula for absorption we find that in the depletion region and the diffusion region adjacent negative absorption can occur due to the fact that the Fermi-levels are separated by qV_a for the two bands under forward bias. In other words: under carrier injection emission of radiation is possible leading to the name injection luminescence.

The increase of minority carriers in the depletion zone makes it possible that more free majority carriers can recombine because they find more partners.This leads to an enhanced recombination current as sketched in figure 1.12. With a heterostructure the minority carrier density is further increased on the small-gap side of the junction leading to enhanced recombination. This leads to the idea of a double heterostructure where an undoped small-gap material is sandwiched between two highly doped wide-gap materials as sketched in figure 1.13. In this case from both sides of the junction minority carriers are injected into the small-gap material leading to a further enhancement of the recombination process.

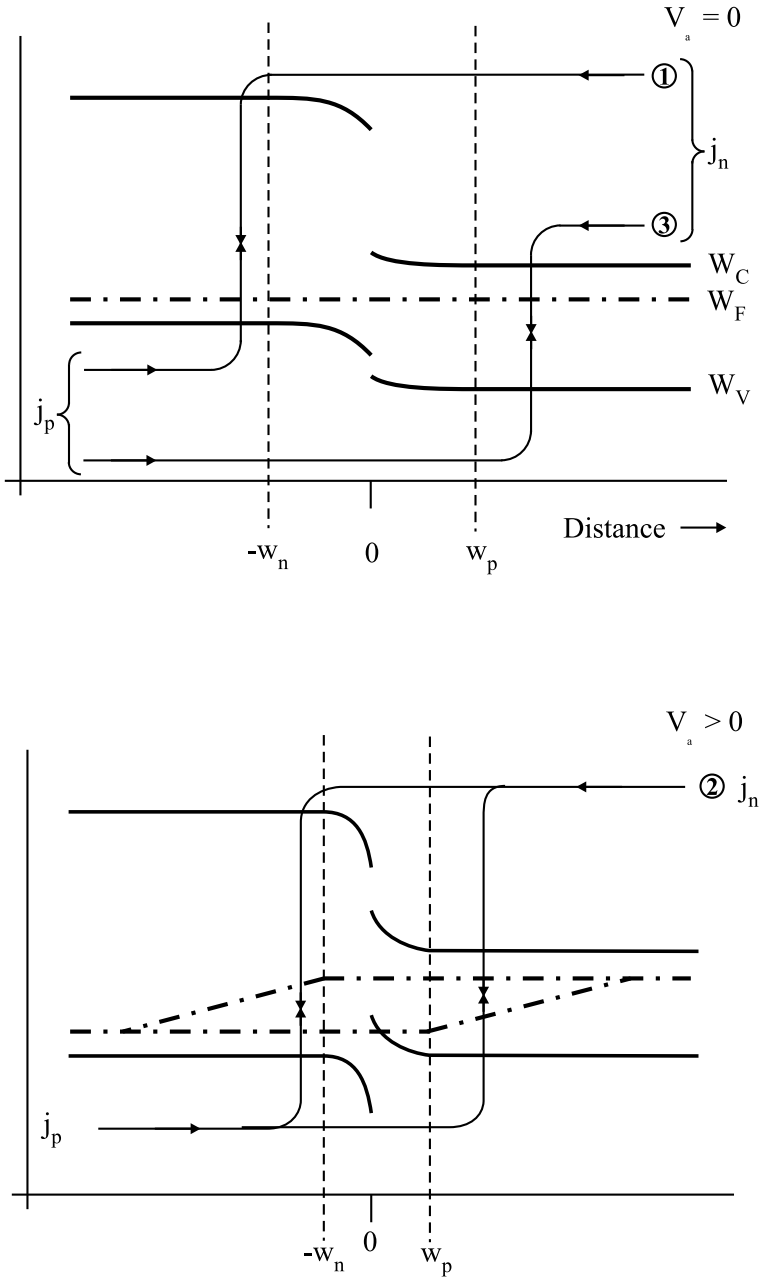

Figure 1.12: Current flow through a pn- heterostructure junction. Under equilibrium conditions recombination in the neutral zone is much more than in the depletion zone and the current is carried through passes (1) and (3). Under forward bias the minority carrier density in the depletion zone is exponentially increased with the applied bias voltage V_{a} and the current is mainly carried over path (2).

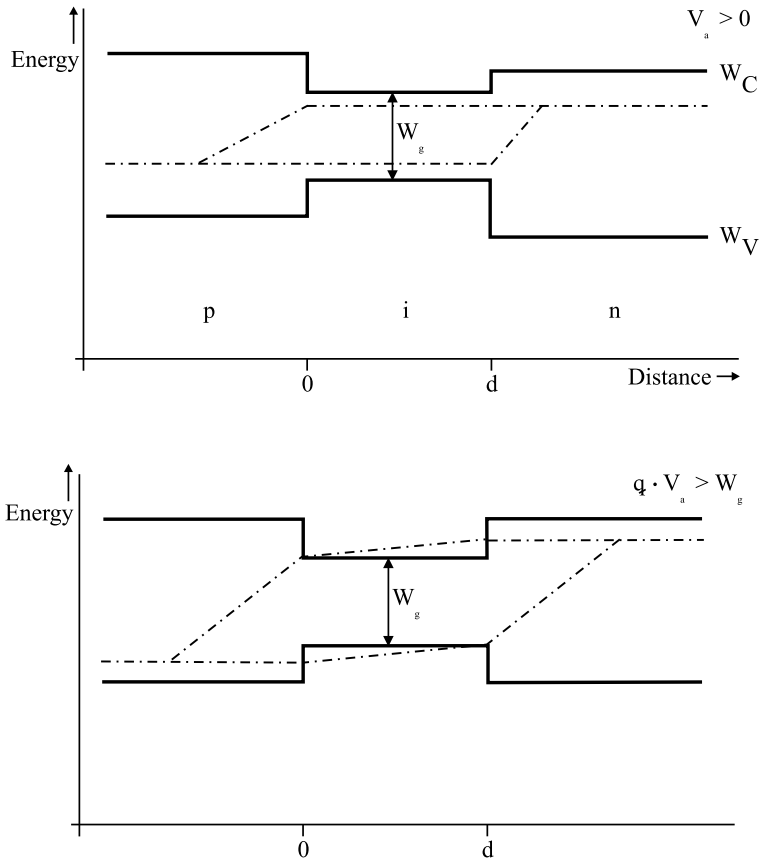

Figure 1.13: Double heterostructure pin- diode under forward bias. Due to carrier injection from the p- and n- side into the undoped small-gap region the luminescence efficiency is remarkably enhanced.

1.4 Optical Amplification

Optical amplification occurs when incident photons of energy $\hbar\omega$ cause stimulated emission that adds photons of the same energy to the incident ones. At the same time absorption occurs. The net rate of absorption and emission under equilibrium results into absorption as we saw already in section 1.2.3. Recalling the expression for $\alpha\{\hbar\omega\}$ we find that the absorption coefficient is proportional to the difference in the Fermi - distributions of the lower and the higher energy level

$$\alpha \propto f_{\mathrm{el}}\{W_1\} - f_{\mathrm{el}}\{W_1 + \hbar\omega\} \quad . \tag{1.17}$$

For optical amplification α must become negative and therefore $f_2 \geq f_1$ resulting into $\exp\left\{\frac{W_2 - W_{\mathrm{F2}}}{k_{\mathrm{B}}T}\right\} \leq \exp\left\{\frac{W_1 - W_{\mathrm{F1}}}{k_{\mathrm{B}}T}\right\}$ or

$$\exp\left\{\frac{W_{\mathrm{F2}} - W_{\mathrm{F1}} - \hbar\omega}{k_{\mathrm{B}}T}\right\} \geq 1 \tag{1.18}$$

indicating that amplification can only occur when the quasi-Fermi-level separation is bigger than the energy difference between the contributing levels

$$W_{\mathrm{F2}} - W_{\mathrm{F1}} \geq \hbar\omega \quad . \tag{1.19}$$

In this case population inversion occurs. Simplified that is more carriers are present at level W_2 than at level W_1 or precisely as has been pointed out before (see 1.2.3) the population probability of states in the upper level is bigger than that in the lower level.

1.4.1 Population Inversion

Population inversion is achieved by separation of the quasi Fermi levels. In a semiconductor pn-diode this can be achieved by application of forward voltage to a double heterostructure as we saw in section 1.3.3.2. The applied voltage must be as high or

higher as the corresponding bandgap of the small-gap material $qV_\mathrm{a} \geq W_\mathrm{gi}$. Under this conditions a strong current flows that is carried by recombination in the small-gap material. The supplied electrical energy flux density jV_a converts into a photon flux density $\frac{jV_\mathrm{a}}{\hbar\omega} \cdot \eta_i$ where η_i denotes the internal conversion efficiency saying how much photons are generated in average by a number of electrons taking into account that some electrons recombine without generation of photons. For non-semiconductors, energy has to be applied to the electrons for transition to a higher energy level. This is the so called pumping where at minimum the later emission energy of the photon $\hbar\omega$ has to be pumped into the corresponding electron.

1.4.2 Lifetime Model for Population Inversion

Pumping of electrons into excited states is used to generate population inversion. The pumping must occur between levels that efficiently links to the states W_1 and W_2 where the emission shall occur. Thinking of lifetimes in the different states we can explain population inversion as sketched in figure 1.14.

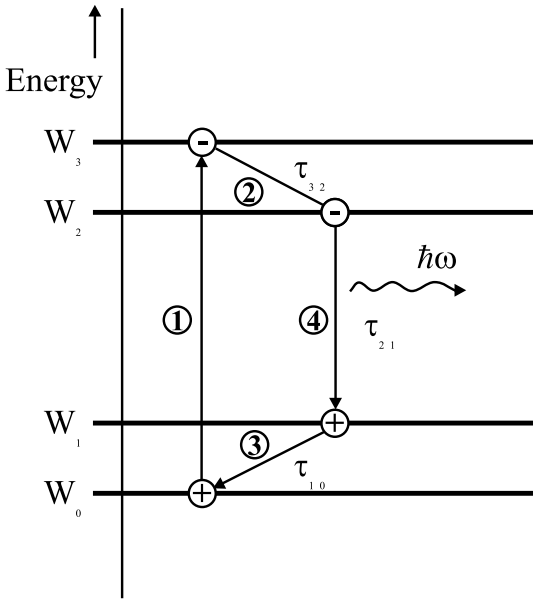

Figure 1.14: After pumping an electron from W_0 into the excited state W_3 (1) the electron relaxes to level W_2 (2) and the hole to W_1 (3). Recombination (4) takes place when an electron falls down to W_1. Population inversion is achieved when $\tau_{32} < \tau_{21}$ and $\tau_{10} < \tau_{21}$.

After generation of an electron-hole pair in levels W_3 and W_0 respectively by pumping the electron in average stays for a time τ_{32} in that state before it relaxes to W_2. This

average time is called the lifetime of the transition $3 \rightarrow 2$. The same applies to the hole in level W_0 that relaxes to W_1. The transition of a hole from the level W_0 to W_1 is the same as the transition of an electron from W_1 to W_0. The later is called emptying of W_1 when only electrons are observed. The lifetime here is τ_{10}. If the lifetime τ_{21} of the transition $W_2 \rightarrow W_1$ is greater than the both τ_{32} and τ_{21} electrons are accumulated in state W_2 under pumping. The model presented here origins from the early gas and solid state lasers with isolated energy levels. In semiconductors one may think of a picture as sketched in figure 1.15. Pumping from the valence band into conduction band can occur by biasing a pn junction or by optical pumping (optical exitation).

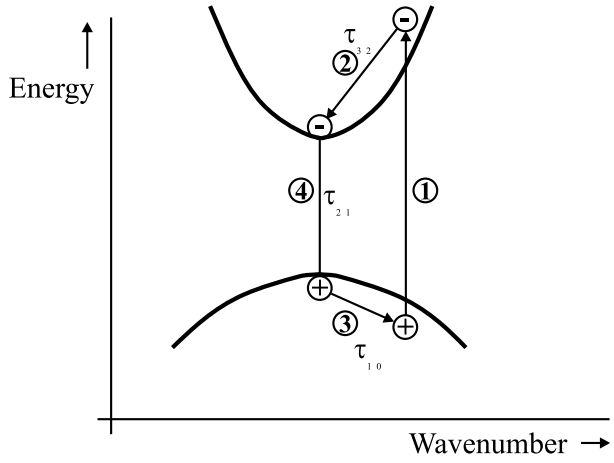

Figure 1.15: Lifetime model for a semiconductor. The lifetimes τ_{32} and τ_{23} are the so called intraband-relaxation lifetimes.

The carriers generated relax to the minimum of the conduction band and the maximum of the valence band. If both relaxation times τ_{10} and τ_{32} are shorter than the recombination time τ_{21} population inversion occurs.

Chapter 2

Optical Resonators and Mirrors

For laser action not only optical amplification is necessary, but also a resonant system where modes can reach such high intensity that remarkable output can be obtained. A well known resonant system is the acoustic feedback from a loudspeaker into a microphone. With sufficient amplification an acoustic wave travels several times from the microphone trough the amplifier into the loudspeaker and back into the microphone. Although there is some loss on the way from the speaker to the microphone with each roundtrip the wave gathers more intensity. In case of proper phase matching, that is constructive interference of the wave in each roundtrip, an eigenmode gets into resonance that one can hear as a high intensity tone. Normally this occurs at high frequencies where the phase matching is easier to obtain (each eigenmode is connected to a discrete frequency). The intensity of the system is limited due to some nonlinear behaviour of the amplifier at high intensities. An optical equivalent of this system is a so called ring resonator as sketched in figure 2.1

Another type of optical resonator is the Fabry-Perot resonator, where only two mirrors are needed and the light travels back and forth through the amplifying region as sketched in figure 2.2. In both cases one can see easily that laser action is easier obtained when the mirrors show little losses, that is high reflectivity. In some cases only one mode is desired. This can be obtained using reflectors that show a strong wave-

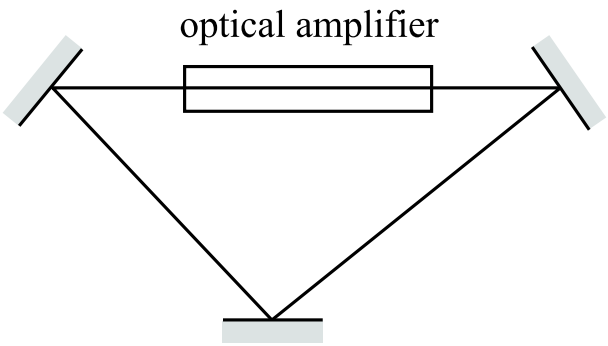

Figure 2.1: Optical ring-resonator. Light travels through the optical amplifier and gathers intensity at each roundtrip. With sufficient amplification the losses through the mirrors are compensated and laser action occurs.

length dependence. Normally only in a small region around the desired wavelength the resonator exhibits high reflectivity thus supporting modes with that wavelength.

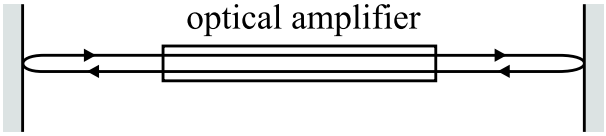

Figure 2.2: Fabry Perot resonator. The light travels forth and back through the amplifying region gathering intensity on both passes resulting into a shorter length of the amplifying region compared to the ring-resonator.

To prevent parasitic modes that can rise at high amplification levels one has to ensure that there are no parasitic reflections. There are three possible ways: First the mirror are placed on the ends of the amplifying region. Second the output facets of the amplifier are anti-reflection coated. The last way is to tilt the facets to the Brewster angle supporting one polarization. The other polarization is traveling out of the resonator as is shown in figure 2.3 for such a so called Brewster-window. Due to the fact that only one polarization is reflected the other one exhibits lower losses and therefore that polarization is dominant in lasers with Brewster windows inside the resonator.

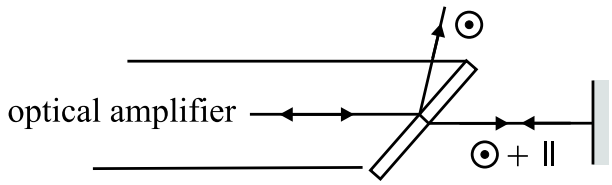

Figure 2.3: Elimination of one polarization in the laser with a Brewster-window. The signs illustrate parallel ‖ and orthogonal ⊙ polarization of the beams.

2.1 Fabry-Perot Resonator

In most lasers a Fabry-Perot resonator is used. The ring resonator is much more complicated in terms of proper justage because at least normally three reflectors have to be adjusted. The ring resonator is used in so called laser gyrometers. Here the Sagnac-effect is used which says that if the laser is rotated around his axis the clockwise and counterclockwise traveling modes are delayed compared to each other. The delay is proportional to the speed of the rotation.

The analysis of both types of resonators follows the same ideas and is detailed here for the Fabry-Perot resonator. In figure 2.4 a well known representation is sketched. A input wave from left hand side is multiply reflected and amplified. The output on the right hand side is collected coherently, this means, we take not only amplitudes but also the phase of each output wave into account.

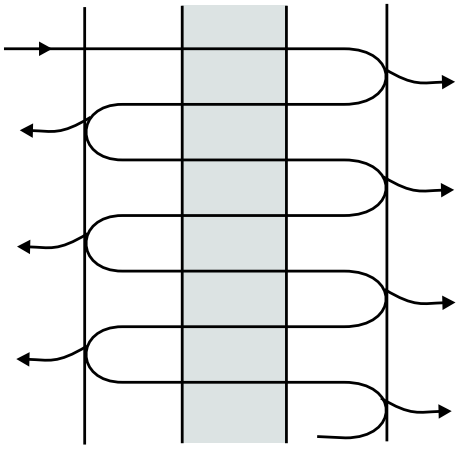

Figure 2.4: Simple model of a Fabry-Perot resonator with incorporated amplification.

In a laser there is no input wave from outside. The action is stimulated by a spontaneous emission that travels around in the laser as depicted in figure 2.5. Collecting the right hand output coherently in this case one finds the same resonance conditions as in the case of light input from outside. Starting form the emission

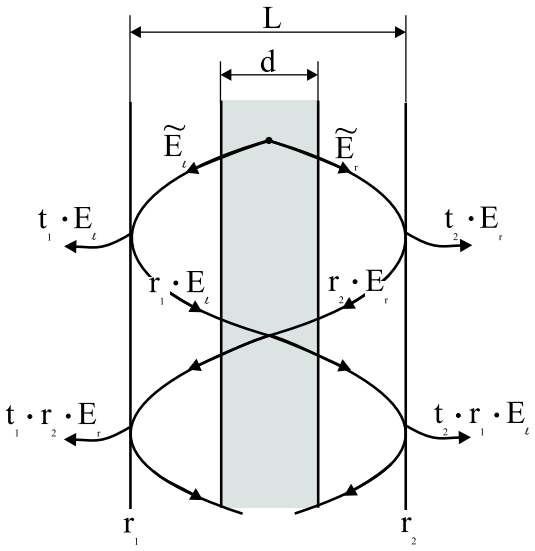

Figure 2.5: Spontaneous emission inside a Fabry-Perot resonator with amplification.

point the wave is reflected at the right facet with factor r_2. It then travels through the etalon between the reflectors gathering some phase $\Psi = \int_0^{\mathrm{L}} k \, \mathrm{d}z$. On the whole way absorption occurs $a_\alpha = \frac{1}{2} \int_0^{\mathrm{L}} \alpha_\mathrm{i} \, \mathrm{d}z$ and inside the amplification zone some gain $a_\mathrm{g} = \frac{1}{2} \int_0^{\mathrm{L}} g \, \mathrm{d}z$ exists. Taking all values k, α_i and g as constants and saying that there is an effective length $\Gamma \cdot L$ [1] of the gain region the amplitude on the left hand side is changed by a factor of $\exp\left\{(\Gamma g - \alpha_\mathrm{i})\frac{L}{2} + ikL\right\}$ compared to the other side. After reflection with r_1 it travels back and is changed by the same factor again. The output is always the amplitude of the right direction traveling wave multiplied by the transmission for $t = 1 + r$. Assuming there is a wave traveling into right direction with amplitude E_r inside the resonator before the first reflection, the total output can be written as

[1] For a ring resonator Γ can be set to $\Gamma = \frac{d}{2L}$ and all the following equations apply well.

$$\begin{aligned} E &= t_1 E_\mathrm{r} + t_1 E_\mathrm{r} r_1 r_2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)L\} \\ &\quad + t_1 E_\mathrm{r} r_1^2 r_2^2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)2L\} t_1 + \cdots \\ &= E_\mathrm{r} t_1 \sum_{m=0}^{M} \left(r_1 r_2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)L\}\right)^m \quad . \end{aligned}$$

this is a geometrical series with result

$$E = E_\mathrm{r} t_1 \frac{1 - \left(r_1 r_2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)L\}\right)^M}{1 - r_1 r_2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)L\}} \quad . \tag{2.1}$$

Resonance is obtained in case that

$$x = r_1 r_2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)L\}$$

equals one. For this situation we have to take a closer look at (2.1) and see that in resonance the output

$$E = \lim_{x \to 1} E_\mathrm{r} t_1 \frac{1 - x^M}{1 - x} = M E_\mathrm{r} t_1$$

is proportional to the number of roundtrips M that coherently superimpose. Coming back to the acoustic analogy this takes into account that the amplifier shows nonlinear behaviour at a certain amplitude ΓE_r where coherence is lost or no more amplification is present. One may regard this as the point where the pumping rate for exited carriers equals the recombination rate mainly due to stimulated recombinations. This is a very simple and rigid model that by far does not describe the real physical effects but explains very well that there is no infinite output of a Fabry Perot with incorporated gain as one would believe from (2.1) under resonance conditions.

Let us now take a closer look at the resonance condition

$$r_1 r_2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)L\} = 1 \quad . \tag{2.2}$$

First of all this can be divided into two equations for amplitude

$$|r_1 r_2 \exp\{(\Gamma g - \alpha_\mathrm{i} + i2k)L\}|\Big|_{\mathrm{g=g_{th}}} = 1 \tag{2.3}$$

and phase

$$\arg\{r_1 r_2 \exp\{(\Gamma_\mathrm{g} - \alpha_\mathrm{i} + i2k)L\}\}\Big|_{\mathrm{resonance}} = m2\pi \quad . \tag{2.4}$$

From the amplitude condition (2.3) follows the so called threshold gain

$$g_\mathrm{th} = \frac{1}{\Gamma}\left(\alpha_\mathrm{i} - \frac{1}{L}\ln\{r_1 r_2\}\right) \quad .$$

Taking intensity reflection coefficients $R = r^2$ we find

$$g_\mathrm{th} = \frac{1}{\Gamma}\left(\alpha_\mathrm{i} - \frac{1}{2L}\ln\{R_1 R_2\}\right) \tag{2.5}$$

that is minimal when both reflecivities R_1 and R_2 are close to $R_1 R_2 = 1$. In this case we rewrite $\sqrt{R_1 R_2} = 1 - \Delta R$ and find

$$g_\mathrm{th} = \frac{1}{\Gamma}\left(\alpha_\mathrm{i} - \frac{1}{L}\ln\{1 - \Delta R\}\right)\Big|_{\Delta R \to 0} \simeq \frac{1}{\Gamma}\left(\alpha_\mathrm{i} + \frac{\Delta R}{L}\right) \tag{2.6}$$

showing the impact of the reflectivity on the threshold gain, especially for low internal absorption or short resonator length. From the phase condition (2.4) the wavelength distance between two modes can be deduced. The phase condition (2.4) decomposes with $k = 2\pi\frac{\eta}{\lambda}$ into

$$\arg\{r_1\} + \arg\{r_2\} + 2\pi\frac{2\eta L}{\lambda}\Big|_{\lambda_m} = m2\pi \tag{2.7}$$

and therefore the resonance-wavelength

$$\lambda_m = \frac{2\eta L}{m - \frac{1}{2\pi}\left(\arg\{r_1\} + \arg\{r_2\}\right)} \quad . \tag{2.8}$$

Due to the fact that the amplitude condition can only be fulfilled in a narrow region around the bandgap wavelength λ_g of the gain material the order of the mode is near

$$m_\mathrm{g} = \frac{2nL}{\lambda_\mathrm{g}} \quad .$$

In case of a "long" resonator, that is $L \gg \lambda_\mathrm{g}$, the phase contribution of the mirrors are negligible and (2.8) simplifies to
$m \gg 1$ ($L \gg \lambda$):

$$m \simeq \frac{2\eta L}{\lambda}$$

and therefore the wavelength distance between two modes is

$$\frac{\Delta m}{\Delta \lambda} = -\frac{1}{\Delta \lambda} \simeq \frac{\mathrm{d}m}{\mathrm{d}\lambda} = -\frac{2\eta L}{\lambda^2}$$

and with $L \gg \lambda$:

$$\frac{\Delta \lambda}{\lambda} \simeq \frac{\lambda}{2\eta L} \quad . \tag{2.9}$$

In the "short"-resonator case $\frac{\Delta m}{\Delta \lambda} \simeq \frac{\mathrm{d}m}{\mathrm{d}\lambda}$ is not completely true but we can use (2.9) in the alternative form

$$\frac{\Delta \lambda}{\lambda} = \frac{\lambda}{2\eta L_\mathrm{eff}} \tag{2.10}$$

with the effective resonator length that a "long"-resonator would have. A good estimation for L_eff can be derived from (2.7) with

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\arg\{r_1\}+\arg\{r_2\}\right)-2\pi\frac{2\eta L}{\lambda^2}=2\pi\frac{\mathrm{d}}{\mathrm{d}\lambda}m\simeq-\frac{2\pi}{\Delta\lambda}$$

resulting into

$$\frac{\Delta\lambda}{\lambda}\simeq\frac{\lambda}{2\eta L-\lambda^2\frac{1}{2\pi}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\arg\{r_1\}+\arg\{r_2\}\right)}$$

and by comparison with (2.10)

$$L_{\mathrm{eff}}\simeq L-\lambda^2\frac{1}{4\pi}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\arg\{r_1\}+\arg\{r_2\}\right)\quad . \tag{2.11}$$

When the gain is switched off the light power decreases exponentially to zero due to the losses α_{i} and outcoupling through the mirrors. The light power is proportional to $\|E\|^2$ respectively the photon density N_{phot}. The decrese for each roundtrip is $N_{\mathrm{phot}}\{2L\}=N_{\mathrm{phot}}\{0\}R_1R_2\exp\{-\alpha_{\mathrm{i}}2L\}=N_{\mathrm{phot}}\{0\}\exp\{(\ln\{R_1R_2\}/(2L)-\alpha_{\mathrm{i}})2L\}$ which can be transformed into the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}z}N_{\mathrm{phot}}=-(\alpha_{\mathrm{i}}-\frac{1}{2L}\ln\{R_1R_2\})N_{\mathrm{phot}}\quad . \tag{2.12}$$

Photons are carrying energy and therefore they are traveling with group velocity (see app. A). With $\frac{\mathrm{d}}{\mathrm{d}t}N_{\mathrm{phot}}=\frac{\mathrm{d}z}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}z}N_{\mathrm{phot}}=c_{\mathrm{gr}}\,\mathrm{d}\,\mathrm{d}zN_{\mathrm{phot}}$ (2.12) is transformed into

$$\frac{\mathrm{d}}{\mathrm{d}t}N_{\mathrm{phot}}=-\frac{c}{\eta_{\mathrm{gr}}}(\alpha_{\mathrm{i}}-\frac{1}{2L}\ln\{R_1R_2\})N_{\mathrm{phot}}=-\frac{1}{\tau}N_{\mathrm{phot}}$$

resulting into a photon lifetime

$$\frac{1}{\tau}=\frac{c}{\eta_{\mathrm{gr}}}(\alpha_{\mathrm{i}}-\frac{1}{2L}\ln\{R_1R_2\})$$

analog to the lifetime model in section 1.4.2.

2.2 Transfer-Matrix Model

As we saw before it would be nice to have high-reflectivity mirrors to obtain laser action already at low amplification. This can be obtained by the usage of stacks of layers with alternating refractive indices.

The fields of a wave in a multilayer stack are concatenated by the boundaries. It is convenient to change to a description of the whole structure in the form of matrices. Each matrix describes the field behaviour inside the individual layers and at their boundaries. With matrix multiplication a description of the whole structure is then easily obtained.

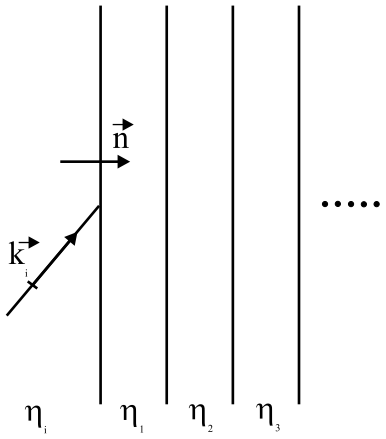

Figure 2.6: Stack of layers with different refractive indexes. A plane wave is incident with direction $\frac{\vec{k}_\mathrm{i}}{k_\mathrm{i}}$.

We suppose a plane wave incident to the stack as sketched in figure 2.6. The vector normal to the parallel-layer stack is $\vec{n}$ ($|\vec{n}| = 1$), the traveling plane wave has a wavenumber vector $\vec{k}_\mathrm{i}$. The situation inside the stack is magnified in figure 2.7. Waves traveling in $z-$direction carry the second index 1, opposite direction has index 2. The first index indicates the layer where the wave travels. The fields of forward and backward travelling waves inside the layer follow

$$E_{m,1}\{z\} = E_{m,1}\{z_m\} \exp\left\{i(\vec{n} \circ \vec{k}_{m,1})(z - z_m)\right\} \tag{2.13}$$

and

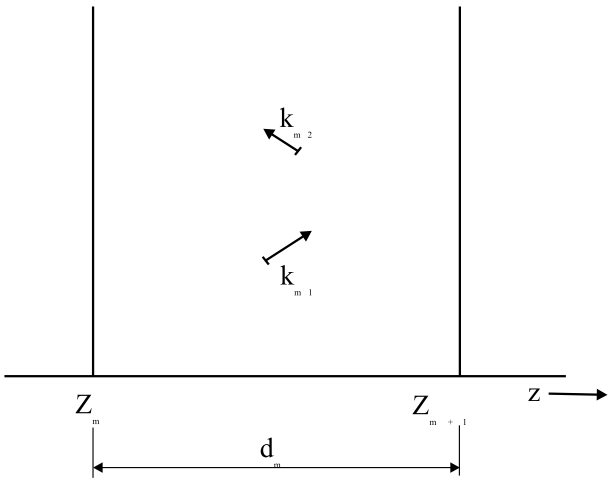

Figure 2.7: Nomenclature of waves inside the stack.

$$\begin{aligned} E_{m,2}\{z\} &= E_{m,2}\{z_m\} \exp\left\{ i(\vec{n} \circ \vec{k}_{m,2})(z - z_m) \right\} \\ &= E_{m,2}\{z_m\} \exp\left\{ -i(\vec{n} \circ \vec{k}_{m,1})(z - z_m) \right\} \quad . \end{aligned} \tag{2.14}$$

(2.13) and (2.14) can be summarized by a propagation matrix P

$$\begin{pmatrix} E_{m,1}\{z\} \\ E_{m,2}\{z\} \end{pmatrix} = P_m\{z - z_m\} \cdot \begin{pmatrix} E_{m,1}\{z_m\} \\ E_{m,2}\{z_m\} \end{pmatrix}$$

with

$$P_m\{z - z_m\} = \begin{pmatrix} \exp\left\{ i(\vec{n} \circ \vec{k}_m)(z - z_m) \right\} & 0 \\ 0 & \exp\left\{ -i(\vec{n} \circ \vec{k}_m)(z - z_m) \right\} \end{pmatrix} \quad . \tag{2.15}$$

Note that $E_{m,1}$ and $E_{m,2}$ are the amplitudes of the electric fields connected to the amplitude vectors by $E_m = \|\vec{E}_m\|$. For a proper description of the layer-boundaries we have to decompose the waves into transversal electric (TE) and transversal magnetic (TM) waves with respect to the boundaries. The conditions are $\vec{n} \circ \vec{E} = 0$ for a

pure TE wave and $\vec{n} \circ \vec{H} = 0$ for a pure TM-wave. With the decomposition $\vec{E} = (\vec{n} \circ \vec{E})\vec{n} + (\vec{n} \times \vec{E}) \times \vec{n}$ one can find the amplitude for TM- and TE-waves using $\vec{E}_{\mathrm{TM}} = (\vec{n} \circ \vec{E})\vec{n}$ and $\vec{E}_{\mathrm{TE}} = \vec{E} - \vec{E}_{\mathrm{TM}}$ with amplitudes $E_{\mathrm{TM}} = \|\vec{E}_{\mathrm{TM}}\| = \vec{n} \circ \vec{E}$ and $E_{\mathrm{TE}} = \|\vec{E}_{\mathrm{TM}}\| = \sqrt{\|\vec{E}\|^2 - (\vec{n} \circ \vec{E})^2}$.

The electric fields on both sides of a boundary are connected to each other by diffraction matrices D

$$D_m \cdot \begin{pmatrix} E_{m,1}\{z_{m+1}\} \\ E_{m,2}\{z_{m+1}\} \end{pmatrix} = D_{m+1} \cdot \begin{pmatrix} E_{m+1,1}\{z_{m+1}\} \\ E_{m+1,1}\{z_{m+1}\} \end{pmatrix} \quad .$$

Usually the boundary is described by the alternative form

$$\left.\begin{pmatrix} E_{m,1} \\ E_{m,2} \end{pmatrix}\right|_{z_{m+1}} = D_{m,m+1} \cdot \left.\begin{pmatrix} E_{m+1,1} \\ E_{m+1,2} \end{pmatrix}\right|_{z_{m+1}}$$

where D has the representation

$$D_{m,m+1} = D_m^{-1} \cdot D_{m+1} = \frac{1}{t_{m+1,m}} \begin{pmatrix} 1 & r_{m+1,m} \\ r_{m+1,m} & 1 \end{pmatrix} \tag{2.16}$$

with the well known reflection coefficients for TE-waves

$$r_{m+1,m} = \frac{\vec{n} \circ \left(\frac{\vec{k}_{m+1}}{\mu_{m+1}} - \frac{\vec{k}_m}{\mu_m} \right)}{\vec{n} \circ \left(\frac{\vec{k}_{m+1}}{\mu_{m+1}} - \frac{\vec{k}_m}{\mu_m} \right)} \quad . \tag{2.17}$$

and for TM-waves

$$r_{m+1,m} = \frac{\vec{n} \circ \left(\frac{\vec{k}_{m+1}}{\varepsilon_{m+1}} - \frac{\vec{k}_m}{\varepsilon_m} \right)}{\vec{n} \circ \left(\frac{\vec{k}_{m+1}}{\varepsilon_{m+1}} - \frac{\vec{k}_m}{\varepsilon_m} \right)} \tag{2.18}$$

and the transmission coefficients $t_{m+1,m} = 1 + r_{m+1,m}$.

Suppose there is an incident wave E_i from a medium number zero that generates a reflected wave with amplitude E_r. The stack may include N layers. The transmitted wave with amplitude E_t exits the stack into medium number $N+1$ as illustrated in figure 2.8.

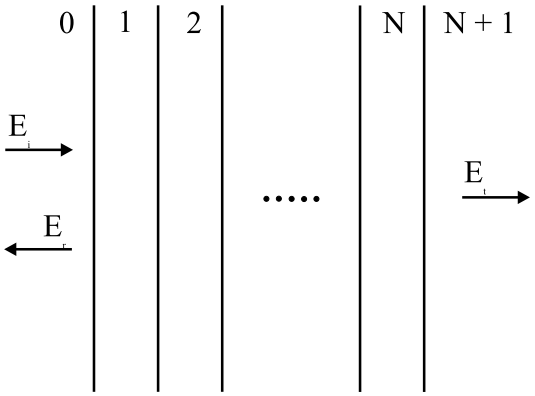

Figure 2.8: Stack of N layers with incident wave E_i that generates a reflected wave E_r and the transmitted wave E_t.

Using the propagation and diffraction matrices P and D the relations between the three wave-amplitudes can be written as

$$\begin{aligned}\begin{pmatrix} E_\mathrm{i} \\ E_\mathrm{r} \end{pmatrix} &= D_{0,1} \left(\prod_{m=1}^{N} P_m\{z_{m+1} - z_m\} \cdot D_{m,m+1} \right) \cdot \begin{pmatrix} E_\mathrm{t} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix} \begin{pmatrix} E_\mathrm{t} \\ 0 \end{pmatrix}\end{aligned} \tag{2.19}$$

resulting into $E_\mathrm{t} = \frac{1}{m_{1,1}} E_\mathrm{i}$ and $E_\mathrm{r} = \frac{m_{2,1}}{m_{1,1}} E_\mathrm{i}$. (2.19) is a compact description of a multilayer system. The matrices inside the parenthesis are called transfer matrices because they describe the transfer of waves along a distance $z - z_m$ in layer m and across a boundary between layers m and $m+1$.

As can be seen for the calculation of the matrices the normal component of $\vec{k}_m$ is needed for the propagation matrix P (2.15) and for the diffraction matrices D (2.16) via the reflection factors (2.17) and (2.18). From the reflection law and Snells law it is well known that for all wavenumber vectors the component parallel to the boundaries

are equal. Inside a homogeneous layer the components into normal direction, here z, are of the same magnitude but of opposite sign. Both can be written in the form

$$\vec{n} \circ (\vec{k}_{m,1} + \vec{k}_{m,2}) = 0 \tag{2.20}$$

$$\vec{n} \times (\vec{k}_{m,1} - \vec{k}_{\mathrm{i}}) = 0 \tag{2.21}$$

$$\vec{n} \times (\vec{k}_{m,2} - \vec{k}_{\mathrm{i}}) = 0 \quad , \tag{2.22}$$

remembering that $\vec{k}_{\mathrm{i}}$ is the wavenumber vector of the exciting wave.
With (2.21) and (2.22) it is possible to derive the normal component of $\vec{k}$ from the dispersion relation when $\vec{k}$ is decomposed as $\vec{k} = (\vec{n} \circ \vec{k})\vec{n} + (\vec{n} \times \vec{k}) \times \vec{n}$. The dispersion relation for plane waves in a medium with complex refractive index $\eta = \sqrt{\varepsilon\mu}$ is
$\eta^2 k_0^2 = \|\vec{k}\|^2 = (\vec{n} \circ \vec{k})^2 + \|\vec{n} \times \vec{k}\|^2$.
Replacing $\vec{n} \times \vec{k}$ by $\vec{n} \times \vec{k}_{\mathrm{i}}$ for each m $\|\vec{k}_m\|^2 = (\vec{n} \circ \vec{k}_m)^2 + \|\vec{n} \times \vec{k}_{\mathrm{i}}\|^2$ follows and the amplitude of the normal component results as $\vec{n} \circ \vec{k}_m = +\sqrt{\eta_m^2 k_0^2 - \|\vec{n} \times \vec{k}_{\mathrm{i}}\|^2}$ and using $\|\vec{n} \times \vec{k}_{\mathrm{i}}\| + (\vec{n} \circ \vec{k}_{\mathrm{i}})^2 = \eta_{\mathrm{i}}^2 k_0^2$ the normal component can be expressed only by the well known normal component of the incident wave and the refractive indices as

$$\vec{n} \circ \vec{k} = +\sqrt{(\eta^2 - \eta_{\mathrm{i}}^2)k_0 + (\vec{n} \circ \vec{k}_{\mathrm{i}})^2} \quad . \tag{2.23}$$

In several cases the waves are damped under propagation in z- direction. This can be attributed to a complex normal component of the wavenumber

$$\vec{n} \circ \vec{k} = \beta + i\alpha/2 \quad . \tag{2.24}$$

2.3 DFB- and DBR-Resonators

As we have seen in section 2.1 several modes are possible in a resonator only limited by the spectral width of the gain curve. In some cases it is desired to support only one

mode inside the gain region which can be done by a short resonator or by wavelength dependent reflectivities of the mirrors. This leads to the so called distributed Bragg Reflectors where the resonators are made of a certain number of quarter-wavelength layer pairs with high and low refractive index or in the case of a waveguide laser by introduction of a periodic perturbation, normally a grating, that has the same effect. Both techniques are illustrated in figure 2.9. The waveguide DBR is made of a grating in the vicinity of the waveguide.

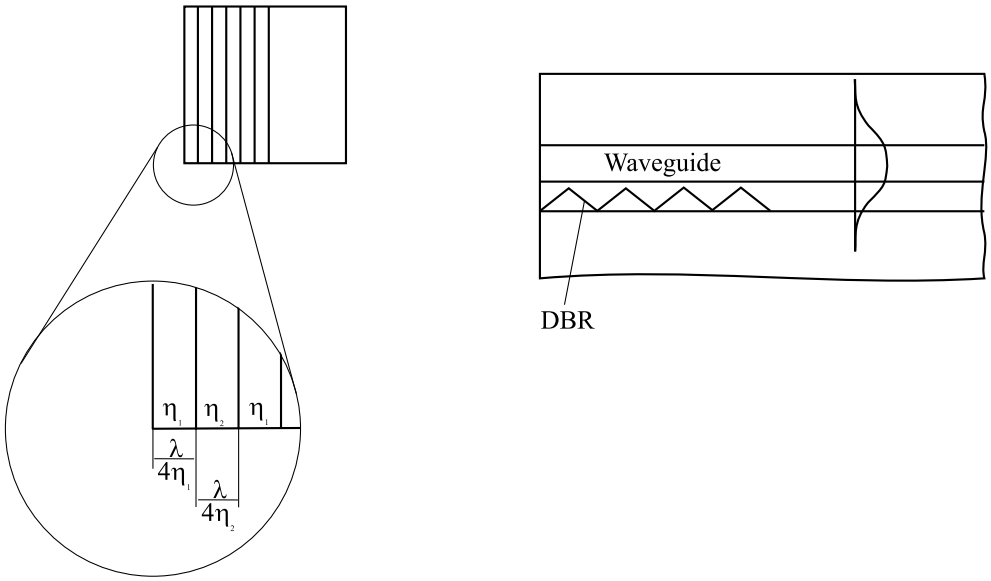

Figure 2.9: Distributed Bragg reflectors. Left: a stack of quarter-wavelength layers work as DBR. Right: The grating near the optical waveguide perturbs the waves resulting into high reflectivity at the resonance wavelength.

Due to the fact that the guided light expands into the regions around the waveguide the propagation is prohibited and at the resonance-wavelength of the DBR a high reflectivity is found. If one or both mirrors in the Fabry Resonator are exchanged against DBR's the resonator is called a DBR-Resonator.

In a waveguide resonator the grating can be extended over the whole length of the laser. In this case the resonator is called a Distributed Feedback (DFB)-Resonator.

The principle is depicted in figure 2.10.

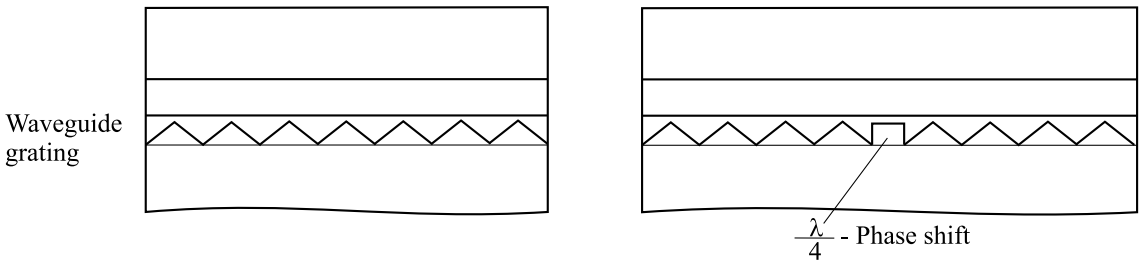

Figure 2.10: DFB-resonators. The grating extends over the whole laser resulting into two modes that exhibit the same reflectivity. The modes are placed symmetrically around the Bragg-frequency. The $\frac{\lambda}{4}-$phase-shift region in the right resonator breaks the degenarcy of the modes leading to emission just at the Bragg-wavelength.

Unfortunately a plain grating leads to two modes symmetrically around the Bragg-frequency of the system. When a $\frac{\lambda}{4}$ phase shift region is incorporated there are two identically DFB resonators that are coupled via the phase-shift region resulting into one mode just at the Bragg-wavelength where the reflectivity is highest. This leads to monomode operation and reduced threshold for a DFB-resonator with phase shift.

2.4 Optical Coatings

As mentioned in section 2 it may be useful in some cases to have an anti-reflection coating on a laser facet. On the ether hand sometimes it is desired to prevent a surface from damage without affecting its optical properties. Both can be attained by coating the surface with a single layer of certain thickness and refractive index as sketched in figure 2.11. For the calculation we can use the transfer-matrix method introduced in section 2.2.

First we have to calculate the normal component of each wavenumber-vector. Supposing the angle of the incident wave to be θ measured against the surface resulting into

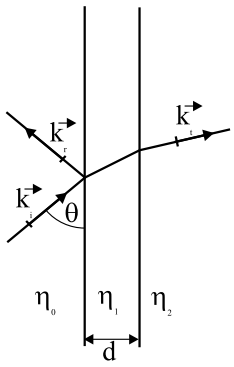

Figure 2.11: A surface of material with refractive index η_2 is coated with a thin layer of thickness d and refractive index η_1. The incident wave travels in a medium with index η_0.

$$\vec{n} \circ \vec{k}_{\mathrm{i}} = k_{\mathrm{i}} \sin\{\theta\} = k_0 \eta_0 \sin\{\theta\} \tag{2.25}$$

with the vacuum-wavenumber $k_0 = \frac{2\pi}{\lambda}$. Following (2.23)

$$\vec{n} \circ \vec{k}_1 = k_0 \sqrt{\eta_1^2 - \eta_0^2 + \eta_0^2 \sin^2\{\theta\}} = k_0 \sqrt{\eta_1^2 - \eta_0^2 \cos^2\{\theta\}} \tag{2.26}$$

$$\vec{n} \circ \vec{k}_2 = k_0 \sqrt{\eta_2^2 - \eta_0^2 \cos^2\{\theta\}} \tag{2.27}$$

are the desired values for D and P matrices. Using the definitions (2.15) and (2.16) we find

$$\begin{pmatrix} E_{\mathrm{i}} \\ E_{\mathrm{r}} \end{pmatrix} = \frac{1}{t_{2,1} t_{1,0}} \begin{pmatrix} 1 & r_{1,0} \\ r_{1,0} & 1 \end{pmatrix} \begin{pmatrix} \exp\left\{ i(\vec{n} \circ \vec{k}_1) d \right\} & 0 \\ 0 & \exp\left\{ -i(\vec{n} \circ \vec{k}_1) d \right\} \end{pmatrix} \cdot \begin{pmatrix} 1 & r_{2,1} \\ r_{2,2} & 1 \end{pmatrix} \begin{pmatrix} E_{\mathrm{t}} \\ 0 \end{pmatrix}$$

and after multiplication

$$\begin{pmatrix} E_{\mathrm{i}} \\ E_{\mathrm{r}} \end{pmatrix} = \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix} \begin{pmatrix} E_{\mathrm{t}} \\ 0 \end{pmatrix} \tag{2.28}$$

with

$$
\begin{aligned}
m_{1,1} &= \frac{1}{t_{2,1}t_{1,0}}\left(\exp\left\{i(\vec{n}\circ\vec{k}_1)d\right\} + r_{2,1}r_{1,0}\exp\left\{-i(\vec{n}\circ\vec{k}_1)d\right\}\right)\\
m_{1,2} &= \frac{1}{t_{2,1}t_{1,0}}\left(r_{2,1}\exp\left\{i(\vec{n}\circ\vec{k}_1)d\right\} + r_{1,0}\exp\left\{-i(\vec{n}\circ\vec{k}_1)d\right\}\right)\\
m_{2,1} &= \frac{1}{t_{2,1}t_{1,0}}\left(r_{1,0}\exp\left\{i(\vec{n}\circ\vec{k}_1)d\right\} + r_{2,1}\exp\left\{-i(\vec{n}\circ\vec{k}_1)d\right\}\right)\\
m_{2,2} &= \frac{1}{t_{2,1}t_{1,0}}\left(r_{2,1}r_{1,0}\exp\left\{i(\vec{n}\circ\vec{k}_1)d\right\} + \exp\left\{-i(\vec{n}\circ\vec{k}_1)d\right\}\right) \quad .
\end{aligned}
$$

The reflection coefficient is

$$
r = \frac{E_\mathrm{r}}{E_\mathrm{i}} = \frac{m_{2,1}}{m_{1,1}} = \frac{r_{1,0} + r_{2,1}\exp\left\{-i2(\vec{n}\circ\vec{k}_1)d\right\}}{1 + r_{2,1}r_{1,0}\exp\left\{-i2(\vec{n}\circ\vec{k}_1)d\right\}} \quad . \tag{2.29}
$$

For lossless media, i.e. $\mathrm{Im}\{\eta\} = 0$, two cases can be easily analyzed. Let us first suppose $(\vec{n}\circ\vec{k}_1)2d = m\cdot 2\pi$ giving $\exp\left\{-i(\vec{n}\circ\vec{k}_1)2d\right\} = 1$ we find

$$
r = \frac{r_{1,0} + r_{2,1}}{1 + r_{1,0}r_{2,1}} = r_{2,0}
$$

which is the same value as would be found without any coating. In this case the coating may act as surface protection layer. The thickness is

$$
d = \frac{m\pi}{\vec{n}\circ\vec{k}_1} = m\frac{\lambda}{2}\frac{1}{\sqrt{\eta_1^2 - \eta_0^2\cos^2\{\theta\}}}
$$

which simplifies for normal incidence $\theta = 90^\circ$ to $d = m\frac{\lambda}{2\eta_1}$. This is m times the half wavelength inside the layer. The second case is $(\vec{n}\circ\vec{k}_1)2d = (2m+1)\pi$ giving $r = \frac{r_{1,0}-r_{2,1}}{1-r_{2,1}r_{1,0}}$. Now it is possible to drive r to zero with $r_{1,0} = r_{2,1}$. For TE-waves

$$
\left(\frac{\vec{n}\circ\vec{k}_1}{\varepsilon_1}\right)^2 = \frac{\vec{n}\circ\vec{k}_\mathrm{i}}{\varepsilon_0}\cdot\frac{\vec{n}\circ\vec{k}_2}{\varepsilon_2} \tag{2.30}
$$

and for TM-waves

$$\left(\frac{\vec{n}\circ\vec{k}_1}{\mu_1}\right)^2 = \frac{\vec{n}\circ\vec{k}_{\mathrm{i}}}{\mu_0}\cdot\frac{\vec{n}\circ\vec{k}_2}{\mu_2} \tag{2.31}$$

satisfy the condition for $r = 0$. Taking again normal incidence $\theta = 90°$ (2.30) and (2.31) simplify to $\frac{\varepsilon_1}{\mu_1} = \sqrt{\frac{\varepsilon_0\varepsilon_2}{\mu_0\mu_2}}$ which in the case of non-magnetic media $\mu_1 = \mu_2 = \mu_0$ is

$$\eta_1 = \sqrt{\eta_0\eta_2} \quad . \tag{2.32}$$

The required thickness is

$$d = \left(m\frac{\lambda}{2} + \frac{\lambda}{4}\right)\cdot\frac{1}{\sqrt{\eta_1^2 - \eta_0^2\cos^2\{\theta\}}}$$

and for normal incidence at least $d = \frac{\lambda}{4\eta_1}$ which is a quarter wavelength inside the layer.

Such a quater wavelength layer is called anti-reflection coating because the suppression of reflections from the surface. The conditions (2.30) and (2.31) tell us which material we have to use for anti-reflection coatings. Often such materials are not available and in this case a multi-layer anti-reflection coating is applied. In that case normally three layers are deposited. The layer structure is depicted in figure 2.12.

Often the refractive indexes of the first and third layer are equal and lower than that of layer two. For given refractive indices there are three thicknesses to play around and minimize the reflectivity which is simulated in a computer using the transfer matrix model described above.

2.5 Mirrors

Mirrors for laser applications should exhibit high reflectivities near to 1. This can be obtained by distributed Bragg-reflectors made of a multi-layer stack of quarter-

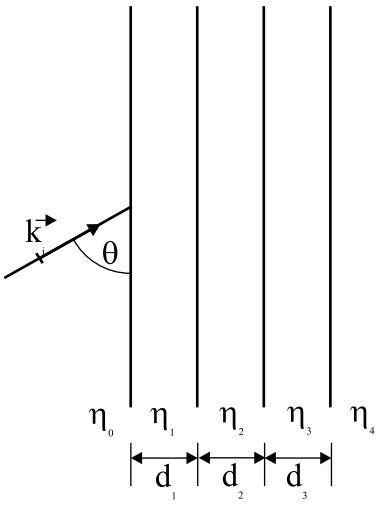

Figure 2.12: Multi-layer anti-reflection coating. Usually the refractive indices are chosen as $\eta_3 = \eta_1 < \eta_2$ and the thicknesses are optimized for minimal reflection.

wavelength layer-pairs. In figure 2.13 such a DBR-mirror is sketched.

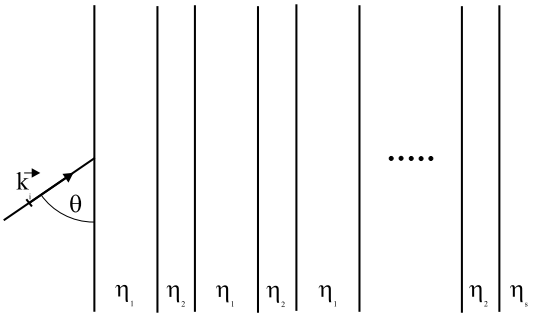

Figure 2.13: DBR-mirror of quarter-wavelength layers.

We will try to analyze the reflectivity of such a structure. Now for M layer pairs we find

$$\begin{pmatrix} E_\mathrm{i} \\ E_\mathrm{r} \end{pmatrix} = D_{0,1} \left(\prod_{m=1}^{M-1} P D_{2m-1,2m} \cdot P \cdot D_{2m,2m+1} \right) P \cdot D_{2M-1,2M} P \cdot D_{2M,\mathrm{s}} \begin{pmatrix} E_\mathrm{t} \\ 0 \end{pmatrix} \quad (2.33)$$

Remembering that the diffraction matrices $D_{2m-1,2m}$ and $D_{2m,2m+1} = D_{1,2}$ are inverse to each other for each m and noting that in the case of quarter-wavelength layers the propagation matrix is $P = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$ which is the same as a scalar multiplication with $\pm i.$, we can put them together into

$$D_{2m-1,2m+1} = P \cdot D_{2m-1,2m} \cdot P \cdot D_{2m,2m+1} = \frac{1}{t_{2,1}t_{1,2}} \begin{pmatrix} i & ir_{2,1} \\ -ir_{2,1} & -i \end{pmatrix} \begin{pmatrix} i & ir_{1,2} \\ -ir_{1,2} & -i \end{pmatrix}$$
$$= D_{1,3} = \frac{-1}{1-(r_{1,2})^2} \begin{pmatrix} 1 & -r_{1,2} \\ r_{1,2} & -1 \end{pmatrix} \begin{pmatrix} 1 & r_{1,2} \\ -r_{1,2} & -1 \end{pmatrix} \tag{2.34}$$

regarding the facts $r_{1,2} = -r_{2,1}$ and $t = 1 + r$. The product is $D_{1,3}$ raised to power $M-1$. For this reason we write $D_{1,3}$ first in the form $D_{1,3} = [w] \cdot [\lambda] \cdot [w]^{-1}$ where $[\lambda]$ is the diagonal matrix of eigenvalues of $D_{1,3}$ and $[w]$ is the eigenvector-matrix giving $D_{1,3}^{M-1} = [w] \cdot [\lambda^{M-1}] \cdot [w]^{-1}$. With

$$D_{1,3} = \frac{-1}{1-r_{1,2}^2} \begin{pmatrix} 1+r_{1,2}^2 & 2r_{1,2} \\ 2r_{1,2} & 1+r_{1,2}^2 \end{pmatrix}$$

the eigenvalues are $\lambda_{1,2} = -\frac{(1\pm r_{1,2})^2}{1-r_{1,2}^2} = -\frac{(1\pm r_{1,2})^2}{(1+r_{1,2})(1-r_{1,2})}$
and therefore $\lambda_1 = \frac{1}{\lambda_2} = -\frac{1+r_{1,2}}{1-r_{1,2}}$ and the eigenvector-matrix results into

$$[w] = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

with the inverse

$$[w]^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad .$$

Now the product can be replaced by

$$D_{1,3}^{M-1} = \frac{1}{2} \begin{pmatrix} \lambda_1^{M-1} + \lambda_2^{M-1} & \lambda_1^{M-1} - \lambda_2^{M-1} \\ \lambda_1^{M-1} - \lambda_2^{M-1} & \lambda_1^{M-1} + \lambda_2^{M-1} \end{pmatrix} \cdot \left(\frac{-1}{1-r_{1,2}^2} \right)^{M-1}$$

summarizing (2.33) into the usual representation (2.28) we find

$$
\begin{aligned}
m_{1,1} &= \frac{1}{t'}\left[\lambda_1^{M-1}(1+r_{1,0})(1-r_{1,2}r_{\mathrm{s},2}) + \lambda_2^{M-1}(1-r_{1,0})(r_{1,2}+r_{\mathrm{s},2})\right] \\
m_{1,2} &= \frac{1}{t'}\left[\lambda_1^{M-1}(1+r_{1,0})(r_{1,2}-r_{\mathrm{s},2}) + \lambda_2^{M-1}(1-r_{1,0})(1-r_{1,2}r_{\mathrm{s},2})\right] \\
m_{2,1} &= \frac{1}{t'}\left[\lambda_1^{M-1}(1+r_{1,0})(1-r_{1,2}r_{\mathrm{s},2}) - \lambda_2^{M-1}(1-r_{1,0})(r_{1,2}+r_{\mathrm{s},2})\right] \\
m_{2,2} &= \frac{1}{t'}\left[\lambda_1^{M-1}(1+r_{1,0})(r_{1,2}-r_{\mathrm{s},2}) + \lambda_2^{M-1}(1-r_{1,0})(1-r_{1,2}r_{\mathrm{s},2})\right]
\end{aligned}
$$

$$
t' = 2 \cdot t_{1,0} \cdot t_{2,1} \cdot t_{\mathrm{s},2} \quad .
$$

The reflectivity of the whole structure is $r = \frac{m_{2,1}}{m_{1,1}}$ which gives

$$
r = \frac{\lambda_1^{2(M-1)} \frac{1+r_{1,0}}{1-r_{1,0}} \cdot \frac{1-r_{1,2}r_{\mathrm{s},2}}{r_{1,2}+r_{\mathrm{s},2}} - 1}{\lambda_1^{2(M-1)} \frac{1+r_{1,0}}{1-r_{1,0}} \cdot \frac{1-r_{1,2}r_{\mathrm{s},2}}{r_{1,2}+r_{\mathrm{s},2}} + 1} \quad . \tag{2.35}
$$

A closer look at the factors gives for TE-waves $\lambda_1 = \frac{\vec{n}\circ\vec{k}_1}{\vec{n}\circ\vec{k}_2} \cdot \frac{\varepsilon_2}{\varepsilon_1}$ and for TM-waves $\lambda_1 = \frac{\vec{n}\circ\vec{k}_1}{\vec{n}\circ\vec{k}_2} \cdot \frac{\mu_2}{\mu_1}$. The second factor is

$$
\frac{1+r_{1,0}}{1-r_{1,0}} = \frac{\vec{n}\circ\vec{k}_1}{\vec{n}\circ\vec{k}_{\mathrm{i}}} \cdot \begin{cases} \frac{\varepsilon_{\mathrm{i}}}{\varepsilon_1} & \text{for TE-waves} \\ \frac{\mu_{\mathrm{i}}}{\mu_1} & \text{for TM-waves} \end{cases}
$$

and the third factor may be written as

$$
\frac{1-r_{1,2}r_{\mathrm{s},2}}{r_{1,2}+r_{\mathrm{s},2}} = \begin{cases} \dfrac{\frac{\vec{n}\circ\vec{k}_2}{\varepsilon_2} \cdot \left(\frac{\vec{n}\circ\vec{k}_1}{\varepsilon_1} + \frac{\vec{n}\circ\vec{k}_{\mathrm{s}}}{\varepsilon_{\mathrm{s}}}\right)}{\frac{\vec{n}\circ\vec{k}_1}{\varepsilon_1} \cdot \frac{\vec{n}\circ\vec{k}_{\mathrm{s}}}{\varepsilon_{\mathrm{s}}} - \frac{(\vec{n}\circ\vec{k}_2)^2}{\varepsilon_2^2}} & \text{for TE-waves} \\ \dfrac{\frac{\vec{n}\circ\vec{k}_2}{\mu_2} \cdot \left(\frac{\vec{n}\circ\vec{k}_1}{\mu_1} + \frac{\vec{n}\circ\vec{k}_{\mathrm{s}}}{\varepsilon_{\mathrm{s}}}\right)}{\frac{\vec{n}\circ\vec{k}_1}{\varepsilon_1} \cdot \frac{\vec{n}\circ\vec{k}_{\mathrm{s}}}{\varepsilon_{\mathrm{s}}} - \frac{(\vec{n}\circ\vec{k}_2)^2}{\varepsilon_2^2}} & \text{for TM-waves} \end{cases} \quad .
$$

which cannot be further simplified. For big numbers of layer pairs M it can easily be seen $\lim_{M\to\infty} r \to \pm 1$ regardless $|\lambda_1| < 1$ or $|\lambda_1| > 1$ and TE- or TM-waves. In non-magnetic media and for normal incidence (2.35) is

$$r\{\theta = 90°\} = \frac{\left(\frac{\eta_1}{\eta_2}\right)^{2(M+1)} \cdot \frac{\eta_i}{\eta_1} \cdot \frac{\eta_2(\eta_1+\eta_2)}{\eta_2^2-\eta_1\eta_s} - 1}{\left(\frac{\eta_1}{\eta_2}\right)^{2(M+1)} \cdot \frac{\eta_i}{\eta_1} \cdot \frac{\eta_2(\eta_1+\eta_2)}{\eta_2^2-\eta_1\eta_s} + 1} \quad . \tag{2.36}$$

As can be seen from (2.36) with bigger index contrast $\Delta\eta = \eta_1 - \eta_2$ fewer pairs are needed to reach a given value of r. Moreover the first layer seen from incidence should exhibit a bigger refractive index than the second of the pairs. A complete analysis of such a structure must be done by (2.33).

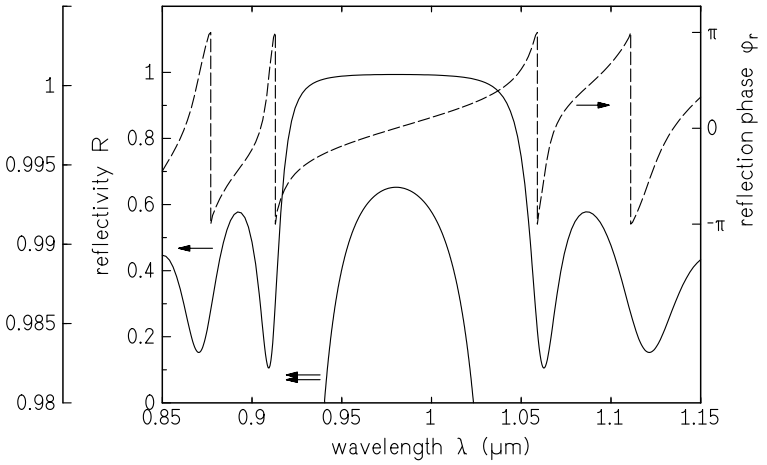

Figure 2.14: Reflectivity- and phase spectra for a 30 pair DBR on GaAs-substrate ($\eta = 3.5$) at normal incidence from air. The pairs are made of AlGaAs ($\eta_1 = 3.3$, $\eta_2 = 3.4$) and have a thickness of 68.2nm and 66.2nm which correspondents to a resonance wavelength of 900nm.

In figure 2.14 the resulting reflectivity r and phase $\arg\{r\}$ is depicted for different wavelength of the wave assuming normal incidence from air. The refractive index of the substrate is $r_s = 3.5$, the 30 pairs are made of material with $\eta_1 = 3.3$ and $\eta_2 = 3.4$

2.6 Photonic Bandgaps

Assume a material that incorporates areas of different refractive index in a lattice-like structure as sketched in figure 2.15. The wave is diffracted at every pillar and a DBR-like reflectivity-spectrum is found.

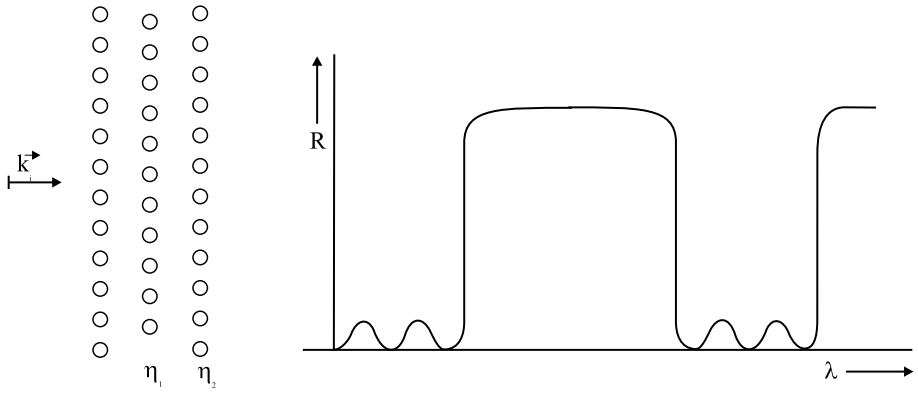

Figure 2.15: Lattice like perturbation of refractive index. The incident wave can only travel through such a structure when it has a wavelength that lies inside a photonic band. Otherwise it is completely reflected as sketched in the reflectivity spectrum.

Only waves inside a transmission-band ($r = 0$) can travel through the structure, other wavelength are completely reflected. This is an analogy to the electronic band-model with bandgaps where the electrons are not able to change their state.
With such structure waveguide can be made that have big dimensions and support only one mode can't be obtained with rigid waveguide-boundaries. Two examples are shown in figure 2.16.

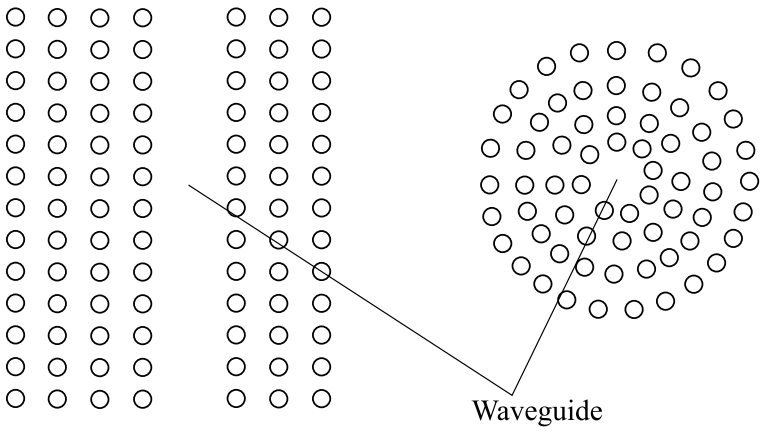

Figure 2.16: Photonic bandgap waveguides. In both cases the waveguide is build by a defect in the perturbation lattice. Such waveguides are monomode although they have dimensions which would allow many modes to exist in a conventional waveguide.

Chapter 3

Optical Waveguiding

Modern communication systems make use of the high bandwidth obtainable with light as information carrier. The light has to be guided from source to receiver by means of optical waveguides. They are mainly made of dielectric material which exhibits low losses. Connections are ranging from chip-to-chip over board-to-board to house-to house in far distance communication. For chip-to-chip connections new attempts are using optical fibers that are embedded inside the printed circuit boards sketched in figure 3.1.

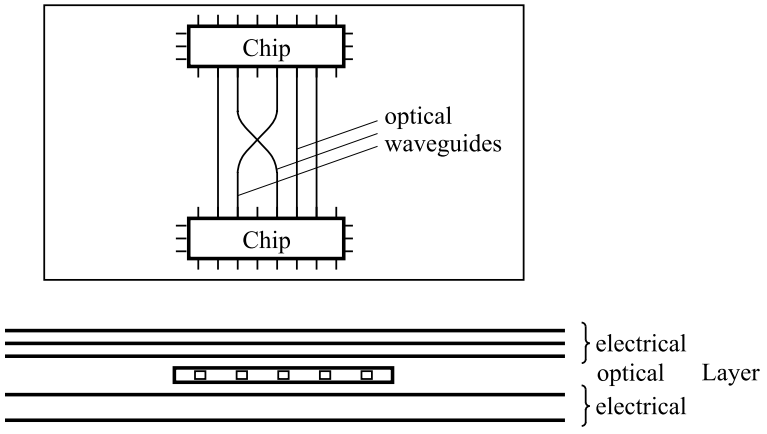

Figure 3.1: Chip to Chip communication by optical waveguides that are embedded in a multi-layer printed circuit board.

Board-to-board communication can be done by free space coupling or via an optical backplane as shown in figure 3.2.

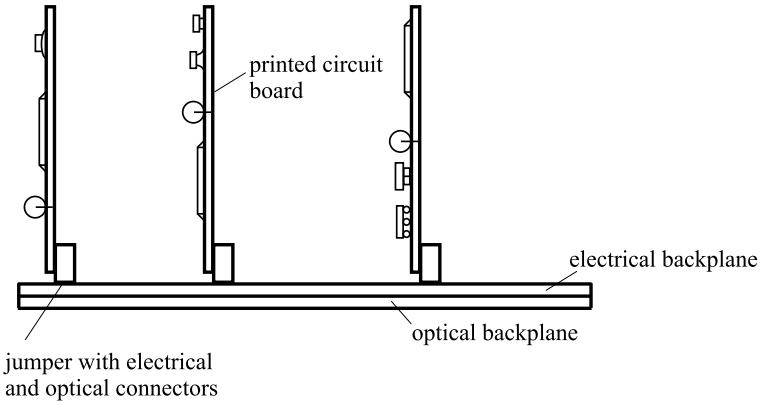

Figure 3.2: Optical connection between printed circuit boards by an optical backplane that couples the board optically.

Longer distances are usually connected via optical fibers made of glass or plastic. An often discussed connection is the intra chip connection, i.e. for clock distribution in a processor chip. There planar optical waveguides made of dielectric layers are under evaluation.

3.1 Planar and Cylindrical optical Waveguides

In the following section we will discuss both planar and cylindrical waveguides. The names are some what confusing and indicate the cross-section of the waveguides to have parallel layers or be circular respectively. Waveguides with parallel layers are planar in the bird eyes view giving the name. On the other hand waveguides with circular cross section look like very long cylinders form which the name origins. It is not necessary for cylindrical waveguides to have circular cross sections. Elliptical and even rectangular waveguides are also cylindrical but are not summarized under this name. We don't lock at that types of waveguides.

3.1.1 Wave Equation

The wave equation for waveguides is derived from the general wave equation assuming a time harmonic electrical or magnetic field into $z-$direction

$$\begin{aligned} \vec{E} &= \vec{\hat{E}}\{x,y\} \cdot \exp\{i(k_z\, z - \omega t)\} \\ \vec{H} &= \vec{\hat{H}}\{x,y\} \cdot \exp\{i(k_z\, z - \omega t)\} \end{aligned} \tag{3.1}$$

where $E\{x,y\}$ and $H\{x,y\}$ can also be noted as $E\{\varrho,\varphi\}$ and $H\{\varrho,\varphi\}$ in cylindrical coordinates. It is usual to write the wavenumber k_z in the form

$$k_z = \beta - i\frac{\alpha}{2} \tag{3.2}$$

with the propagation constant β and intensity damping factor α. Damping occurs when there are losses in the waveguide due to scattering or material losses. The first one have to be considered separately the latter can be incorporated assuming a conductive waveguide material with conductivity σ. Substituting (3.1) into the Maxwell equations assuming a homogeneous region we find

$$\begin{aligned} \frac{\partial^2}{\partial x^2}E_z + \frac{\partial^2}{\partial y^2}E_z + k_\varrho^2\, E_z = 0 \\ \frac{\partial^2}{\partial x^2}H_z + \frac{\partial^2}{\partial y^2}H_z + k_\varrho^2\, H_z = 0 \end{aligned} \tag{3.3}$$

where $E_z = \vec{e}_z \circ \vec{\hat{E}}$ and $H_z = \vec{e}_z \circ \vec{\hat{H}}$. The other components can be calculated from the solutions of (3.3) by

$$E_x = \frac{1}{k_\varrho^2}\left(ik_z\frac{\partial}{\partial x}E_z + i\omega\mu\mu_0\frac{\partial}{\partial y}H_z\right)$$

$$
\begin{aligned}
E_{\mathrm{y}} &= \frac{1}{k_\varrho^2}\left(ik_{\mathrm{z}}\frac{\partial}{\partial y}E_{\mathrm{z}} + i\omega\mu\mu_0\frac{\partial}{\partial x}H_{\mathrm{z}}\right)\\
H_{\mathrm{x}} &= \frac{1}{k_\varrho^2}\left((\theta - i\omega\varepsilon\varepsilon_0)\frac{\partial}{\partial y}E_{\mathrm{z}} + ik_{\mathrm{z}}\frac{\partial}{\partial x}H_{\mathrm{z}}\right)\\
H_{\mathrm{y}} &= \frac{1}{k_\varrho^2}\left((-\theta + i\omega\varepsilon\varepsilon_0)\frac{\partial}{\partial x}E_{\mathrm{z}} + ik_{\mathrm{z}}\frac{\partial}{\partial y}H_{\mathrm{z}}\right)
\end{aligned}
\tag{3.4}
$$

with

$$
\begin{aligned}
k_\varrho^2 &= \omega^2\varepsilon_0\mu_0\varepsilon\mu - k_{\mathrm{z}}^2 + i\omega\sigma\mu_0\mu\\
&= \eta^2 k_0^2 - k_{\mathrm{z}}^2
\end{aligned}
\tag{3.5}
$$

where $k_0^2 = \omega^2\varepsilon_0\mu_0 = \left(\frac{2\pi}{\lambda}\right)^2$, $\eta^2 = \mu\left(\varepsilon + i\frac{\sigma}{\omega\varepsilon_0}\right) = \mu\tilde{\varepsilon}$ and $\tilde{\varepsilon} = \varepsilon + i\frac{\sigma}{\omega\varepsilon_0}$.
In the case when $E_{\mathrm{z}} = 0$ we have a pure transversal electric (TE) wave in the other case $H_{\mathrm{z}} = 0$ the wave is a so called transversal magnetic (TM) wave. In both cases $k_\varrho \neq 0$ is required for finite fields from (3.4). In the case $k_\varrho = 0$ we have to fulfill $E_{\mathrm{z}} \equiv 0$ and $H_{\mathrm{z}} \equiv 0$ simultaneously to prevent infinite fields resulting form (3.4). This are transversal electromagnetic (TEM) waves. Now from the Maxwell equations follows $-k_{\mathrm{z}}E_{\mathrm{y}} = \omega\mu\mu_0 H_{\mathrm{x}}$ and $k_{\mathrm{z}}E_{\mathrm{x}} = \omega\mu\mu_0 H_{\mathrm{y}}$. Both fields turn out to be source free $\nabla \circ \vec{E}_{\mathrm{TEM}} = 0$ and $\nabla \circ \vec{H}_{\mathrm{TEM}} = 0$ and can be calculated by solving ether $\Delta\vec{H}_{\mathrm{TEM}} = 0$ or $\Delta\vec{E}_{\mathrm{TEM}} = 0$ obeying the relation between E and H given above. For all three types of waves

$$
\vec{E} \circ \vec{H} = 0 \tag{3.6}
$$

holds which tells us that the fields $\vec{E}$ and $\vec{H}$ are orthogonal. Currently used optical waveguides support mainly TE and TM modes and therefore we only discuss this types further. Pure TE or TM waves that obey (3.3) can be written in an alternative form with magnetic and electric vectorpotentials. For TE-waves $E_{\mathrm{z}} = 0$ holds and we may assume an electrical vectorpotential $\vec{F}$ such that

$$\varepsilon_0 \, \tilde{\varepsilon} \, \vec{E}_{\mathrm{TE}} = \vec{D}_{\mathrm{TE}} = \nabla \times \vec{F}_{\mathrm{TE}} \tag{3.7}$$

holds. To assure $E_{\mathrm{z}} = 0$ we have to take

$$\vec{F}_{\mathrm{TE}} = F_{\mathrm{TE}} \cdot \exp\left\{i(k_{\mathrm{z}} - \omega t)\right\} \vec{e}_{\mathrm{z}} \tag{3.8}$$

and with the same discussion we introduce a magnetic vectorpotential for TM-waves

$$\vec{A}_{\mathrm{TM}} = A_{\mathrm{TM}} \cdot \exp\left\{i(k_{\mathrm{z}} - \omega t)\right\} \vec{e}_{\mathrm{z}} \tag{3.9}$$

$$\mu\mu_0 \vec{H}_{\mathrm{TM}} = \vec{B}_{\mathrm{TM}} = \nabla \times \vec{A}_{\mathrm{TM}} \quad . \tag{3.10}$$

From (3.7) and (3.10) follows that only source-free fields $\nabla \circ \vec{E} = 0$ and $\nabla \circ \vec{H} = 0$ can be described with $\vec{F}_{\mathrm{TE}}$ and $\vec{A}_{\mathrm{TM}}$. Substituting $\vec{E}_{\mathrm{TE}}$ or $\vec{H}_{\mathrm{TM}}$ in the Maxwell equations we find
$\vec{H}_{\mathrm{TE}} = -i\frac{\omega}{k^2}\left(k^2 F_{\mathrm{TE}} \vec{e}_{\mathrm{z}} + ik_{\mathrm{z}} \nabla F_{\mathrm{TE}}\right)$ and $\vec{E}_{\mathrm{TM}} = i\frac{\omega}{k^2}\left(k^2 A_{\mathrm{TM}} \vec{e}_{\mathrm{z}} + ik_{\mathrm{z}} \nabla A_{\mathrm{TM}}\right)$ and therefore

$$\begin{aligned} \vec{H}_{\mathrm{TE}} \circ \vec{e}_{\mathrm{z}} &= -i\omega \frac{k_\varrho^2}{k^2} F_{\mathrm{TE}} \\ \vec{E}_{\mathrm{TM}} \circ \vec{e}_{\mathrm{z}} &= i\omega \frac{k_\varrho^2}{k^2} A_{\mathrm{TM}} \quad . \end{aligned}$$

Substituting this into (3.3) two similar scalar wave equations

$$\Delta A_{\mathrm{TM}} + k_\varrho^2 A_{\mathrm{TM}} = 0 \tag{3.11}$$

$$\Delta F_{\mathrm{TE}} + k_\varrho^2 F_{\mathrm{TE}} = 0 \tag{3.12}$$

result for A and F. It has to be stated that $\vec{A}$ and $\vec{F}$ in this case are Lorentz-calibrated. As can be seen in (3.8) and (3.9) the vector potentials are waves that travel in z-direction. In general every wave can be expanded into plane waves by

$$\psi = \psi_0 \cdot \exp\left\{i(\vec{k} \circ r\vec{e}_{\mathrm{r}} - \omega t)\right\} = \psi_0 \exp\left\{i(\Phi - \omega t)\right\}$$

with wavenumbervector

$$\vec{k} = \nabla\Phi \quad .$$

The dispersion relation $\|\vec{k}_{\|}^2 = k^2 = \eta^2 k_0^2$ has to be fulfilled for every contributing plane wave. We write $\vec{k} = \vec{\beta} + i\vec{a}$ and find $|\vec{\beta}|^2 - |\vec{a}|^2 + i2\vec{\beta} \circ \vec{a} = \eta^2 k_0^2$. In the case of lossless media ($\mathrm{Im}\{\eta\} = 0$) two possible cases for the required condition $\vec{a} \circ \vec{\beta} = 0$ are resulting.

$\vec{a} = 0$: The wave travels like a plane wave in direction of $\vec{\beta}$ without any losses. This is the trivial case.

$\vec{a} \neq 0$: Now $\vec{a} \circ \vec{\beta} = 0$ has to be met. This means that the wave travels into direction of $\vec{\beta}$ but exhibits losses transverse to $\vec{\beta}$. Such waves travel again into $\vec{\beta}$-direction but exhibit no energy loss in that direction resulting into losses energy transport in $\vec{\beta}$-direction. Such waves are called transverse-damped waves. A typical example occurs at total reflection where the transmitted wave travels along the interface and is evanescent in direction normal to the interface.

In general we can note the condition for lossless energy transport

$$\begin{aligned} \nabla\Phi &= \vec{\beta} + i\vec{a} \\ \vec{a} \circ \vec{\beta} &= 0 \quad . \end{aligned} \tag{3.13}$$

As a last point we have to consider waves at an interface between waveguide layers. For charge- and current-free interfaces the boundary conditions are

$$\vec{n} \circ (\varepsilon_0\, \tilde{\varepsilon}\, \vec{E}) = \mathrm{const} \tag{3.14}$$

$$\vec{n} \circ (\mu\, \mu_0\, \vec{H}) = \mathrm{const} \tag{3.15}$$

$$\vec{n} \times \vec{E} = \mathrm{const} \tag{3.16}$$

$$\vec{n} \times \vec{H} = \mathrm{const} \quad . \tag{3.17}$$

The interfaces are perpendicular to the $z-$direction $\vec{n} \circ \vec{e}_z = 0$ and therefore in the case of TE-waves (3.15) and for TM-waves (3.14) is already fulfilled.
For simplicity all media concerned in the next section are assumed to be nonmagnetic, that is $\mu = 1$.
In (3.11) and (3.12) we have assumed the media to be homogeneous, that is $\nabla\tilde{\varepsilon} = 0$ and $\nabla\mu = 0$. In real waveguides the media may be inhomogeneous, at least perpendicular to the z direction. With Lorentz calibrated vector potentials we find for inhomogeneous media

$$\Delta\vec{A} + k^2\vec{A} + \frac{1}{\mu}(\nabla\mu) \times (\nabla \times \vec{A}) = 0 \tag{3.18}$$

$$\Delta\vec{F} + k^2\vec{F} + \frac{1}{\tilde{\varepsilon}}(\nabla\tilde{\varepsilon}) \times (\nabla \times \vec{F}) = 0 \quad . \tag{3.19}$$

We take again $\vec{A}_{\mathrm{TM}} = A_{\mathrm{TM}} \cdot \vec{e}_z$ and $\vec{F}_{\mathrm{TE}} = F_{\mathrm{TE}} \cdot \vec{e}_z$ for TM- and TE-waves respectively and assume $\tilde{\varepsilon}$ and μ to be $z-$independent $\frac{\mathrm{d}}{\mathrm{d}z}\,\tilde{\varepsilon} = 0$ and $\frac{\mathrm{d}}{\mathrm{d}z}\,\mu = 0$. Now

$$\Delta A_{\mathrm{TM}} + k_\varrho^2 A_{\mathrm{TM}} + \frac{1}{\mu}(\nabla_\varrho\,\mu) \circ (\nabla_\varrho\,A_{\mathrm{TM}}) = 0 \tag{3.20}$$

and

$$\Delta F_{\mathrm{TE}} + k_\varrho^2 F_{\mathrm{TE}} + \frac{1}{\tilde{\varepsilon}}(\nabla_\varrho\,\tilde{\varepsilon}) \circ (\nabla_\varrho\,F_{\mathrm{TE}}) = 0 \tag{3.21}$$

have to be solved. For non magnetic media $\mu = 1$ (3.20) changes back to (3.11). For slowly varying dielectric constants $\tilde{\varepsilon}$

$$|(\nabla\tilde{\varepsilon}) \circ (\nabla F_{\mathrm{TE}})| \ll k_\varrho^2\,\tilde{\varepsilon}\,F_{\mathrm{TE}} \tag{3.22}$$

we can use (3.12) as a good approximation for (3.21) which simplifies the analysis enormously.

3.1.2 Step-Index Planar Waveguides

In figure (3.3) the cross section for a planar waveguide is shown. Due to the fact that the structure extends in $y-$direction to infinity we expect no $y-$dependence in the fields and have therefore $k_\varrho = k_\mathrm{x}$.

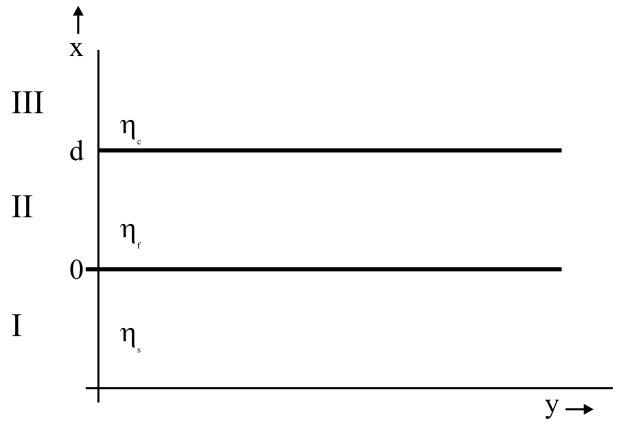

Figure 3.3: Planar optical waveguide. We assume region II to be the guiding film with refractive index η_f on a substrate with η_s and coated by a cover with η_c

For both TE and TM waves we can attempt F_TE and A_TM in the three regions as

$$\begin{aligned}
\psi_\mathrm{I} &= \psi_{\mathrm{I},1} \cdot \exp\{-ik_{\mathrm{x}_1}\, x\} \cdot \exp\{i(k_\mathrm{z}\, z - \omega t)\} \\
\psi_\mathrm{II} &= (\psi_{\mathrm{II},1} \cdot \exp\{ik_{\mathrm{x}_2}\, x\} + \psi_{\mathrm{II},2} \cdot \exp\{-ik_{\mathrm{x}_2}\, x\}) \exp\{i(k_\mathrm{z}\, z - \omega t)\} \qquad (3.23) \\
\psi_\mathrm{III} &= \psi_{\mathrm{III},1} \cdot \exp\{ik_{\mathrm{x}_3}(x-d)\} \exp\{i(k_\mathrm{z}\, z - \omega t)\} \quad .
\end{aligned}$$

For TE-waves we substitute ψ by F and have to fulfill the boundary equations (3.14), (3.16) and (3.17) at $x = 0$ and $x = d$. With $\vec{n} = \vec{e}_\mathrm{x}$ we find that (3.14) is fulfilled and from (3.17) and (3.16) $\vec{E} \circ \vec{e}_\mathrm{z}$ and $\vec{H} \circ \vec{e}_\mathrm{z}$ have to be steady at the boundaries which gives

$$\left(\frac{k_{\mathrm{x}_3}}{k_3}\right)^2 F_\mathrm{III}\Big|_{x=d} = \left(\frac{k_{\mathrm{x}_2}}{k_2}\right)^2 F_\mathrm{II}\Big|_{x=d}$$

$$\left(\frac{k_{\mathrm{x}_2}}{k_2}\right)^2 F_{\mathrm{II}}\Big|_{x=0} = \left(\frac{k_{\mathrm{x}_1}}{k_1}\right)^2 F_{\mathrm{I}}\Big|_{x=0} \tag{3.24}$$

$$\begin{aligned} \frac{1}{\tilde{\varepsilon}_3}\,\frac{\mathrm{d}}{\mathrm{d}x}F_{\mathrm{III}}\Big|_{x=d} &= \frac{1}{\tilde{\varepsilon}_2}\,\frac{\mathrm{d}}{\mathrm{d}x}F_{\mathrm{II}}\Big|_{x=d} \\ \frac{1}{\tilde{\varepsilon}_2}\,\frac{\mathrm{d}}{\mathrm{d}x}F_{\mathrm{II}}\Big|_{x=0} &= \frac{1}{\tilde{\varepsilon}_1}\,\frac{\mathrm{d}}{\mathrm{d}x}F_{\mathrm{I}}\Big|_{x=0} \end{aligned} \tag{3.25}$$

respectively, where the abbrveations

$$\begin{aligned} k_{\mathrm{x}_1}^2 &= \eta_1^2\, k_0^2 - k_{\mathrm{z}}^2 \\ k_{\mathrm{x}_2}^2 &= \eta_2^2\, k_0^2 - k_{\mathrm{z}}^2 \\ k_{\mathrm{x}_3}^2 &= \eta_3^2\, k_0^2 - k_{\mathrm{z}}^2 \end{aligned} \tag{3.26}$$

are used. The four equations can be written in matrix form

$$\begin{pmatrix} 0 & \left(\frac{k_{\mathrm{x}_2}}{k_2}\right)^2 \exp\{ik_{\mathrm{x}_2}\, d\} & \left(\frac{k_{\mathrm{x}_2}}{k_2}\right)^2 \exp\{-ik_{\mathrm{x}_2}\, d\} & -\left(\frac{k_{\mathrm{x}_3}}{k_3}\right)^2 \\ -\left(\frac{k_{\mathrm{x}_1}}{k_1}\right)^2 & \left(\frac{k_{\mathrm{x}_2}}{k_2}\right)^2 & \left(\frac{k_{\mathrm{x}_2}}{k_2}\right)^2 & 0 \\ 0 & \frac{k_{\mathrm{x}_2}}{\tilde{\varepsilon}_2} \cdot \exp\{ik_{\mathrm{x}_2}\, d\} & -\frac{k_{\mathrm{x}_2}}{\tilde{\varepsilon}_2} \cdot \exp\{-ik_{\mathrm{x}_2}\, d\} & -\frac{k_{\mathrm{x}_3}}{\tilde{\varepsilon}_3} \\ \frac{k_{\mathrm{x}_1}}{\tilde{\varepsilon}_1} & \frac{k_{\mathrm{x}_2}}{\tilde{\varepsilon}_2} & -\frac{k_{\mathrm{x}_2}}{\tilde{\varepsilon}_2} & 0 \end{pmatrix} \begin{pmatrix} F_{\mathrm{I},1} \\ F_{\mathrm{II},1} \\ F_{\mathrm{II},2} \\ F_{\mathrm{III},1} \end{pmatrix} = 0 \tag{3.27}$$

This is an eigenvalue problem. Solutions are only possible when the determinant of the matrix equals zero. We can vary k_{z} and therefore k_{x_1}, k_{x_2} and k_{x_3} are varied. When the determinant is zero we have a possible solution and k_{z} is the eigenvalue of that

solution. With one amplitude given, i.e. $F_{\mathrm{II},1}$, all other amplitudes follow from (3.27). From the determinant of (3.27) the well known resonance equation

$$r_{1,2} \cdot r_{2,3} \cdot \exp\{i\, 2k_{\mathrm{x}_2} d\} = 1 \tag{3.28}$$

follows directly. The resonance equation can be split into amplitude condition $|r_{1,2}| \cdot |r_{2,3}| \cdot \exp\{-2\,\mathrm{Im}\{k_{\mathrm{x}_2}\}\, d\} = 1$ and phase condition $\arg\{r_{1,2}\} + \arg\{r_{2,3}\} + 2\mathrm{Re}\{k_{\mathrm{x}_2}\}\, d = m \cdot 2\pi$. For lossless waveguiding in lossless media the amplitude condition requires $|r_{1,2}| \cdot |r_{2,3}| = 1$ due to $\mathrm{Im}\{k_{\mathrm{x}_2}\} = 0$. This can be obtained when total reflection occurs at the interfaces and the waves in both regions I and III are transversal damped. With the transformations

$$\begin{aligned} V &= k_0\, d \sqrt{\eta_2^2 - \eta_1^2}, \\ B &= \frac{\left(\frac{k_z}{k_0}\right)^2 - \eta_1^2}{\eta_2^2 - \eta_1^2} \\ a_{\mathrm{TE}} &= \frac{\eta_1^2 - \eta_3^2}{\eta_2^2 - \eta_1^2} \end{aligned}$$

(3.28) can be written in the form

$$V = \frac{1}{\sqrt{1-B}} \left(m\pi + \arctan\left\{ \sqrt{\frac{B}{1-B}} \right\} + \arctan\left\{ \sqrt{\frac{B + a_{\mathrm{TE}}}{1-B}} \right\} \right)$$

where V can be calculated as a function of B as depicted in figure 3.4

At this point it can be stated that (3.28) follows also for TM-waves. The only difference for the $B - V-$diagram is the asymmetry-parameter which has to be calculated from

$$a_{\mathrm{TM}} = \left(\frac{\eta_2}{\eta_1}\right)^4 \cdot a_{\mathrm{TE}} \quad .$$

In the definition of B the ratio $\frac{k_z}{k_0}$ occurs which is called the effective index

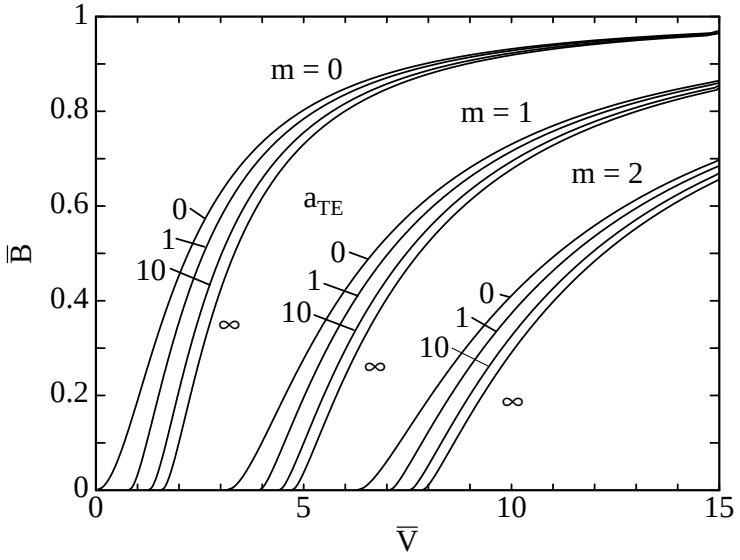

Figure 3.4: $B - V-$diagram for TE-waves in lossless media. The starting points are separated by π on the $V-$axis and begin at $V = \arctan\left\{\sqrt{a_{\mathrm{TE}}}\right\}$.

$$\eta_{\mathrm{eff}} = \frac{k_{\mathrm{z}}}{k_0} \quad . \tag{3.29}$$

This is the index of a virtual homogeneous material a plane wave would travel in at the same speed (phase velocity).

The phase condition can be alternately written in the form

$$2 \int_0^d k_{\varrho \mathrm{t}} \, \mathrm{d}\varrho - \Phi_{\mathrm{s}} - \Phi_{\mathrm{c}} = m \cdot 2\pi \tag{3.30}$$

with phase changes at reflection from substrate Φ_{s} and cover Φ_{c} and transversal wavenumber $k_{\varrho \mathrm{n}} = \vec{e}_{\mathrm{t}} \circ \vec{k}_{\varrho}$. This representation of the phase condition can be interpreted as a wave traveling in transverse direction with $k_{\varrho \mathrm{t}}$ accumulating phase on the way. The optical path length is effectively somewhat longer than d due to the phase shifts occurring at reflection. With constant $k_{\varrho \mathrm{t}}$ the effective waveguide width

can be estimated from (3.30) to be $d_{\text{eff}} = d - \frac{\Phi_s+\Phi_c}{2k_{\varrho t}}$ such that $2\, d_{\text{eff}} \cdot k_{\varrho n} = m \cdot 2\pi$ is satisfied. Remembering that the phase shift at cover and substrate follow $\Phi_s < 0$ and $\Phi_c < 0$ the statement of a bigger effective waveguide width than the geometrical one is understood easily.

3.1.3 Graded-Index Planar Waveguides

In a graded index waveguide the refractive index changes transverse to $z-$direction. In practice this is done by doping of layers or diffusion of a dopand. A good approximation for the index profile $\eta\{x\}$ is

$$\eta^2\{x\} = \begin{cases} \eta_f^2 - (\eta_f^2 - \eta_s^2) \cdot \left(\frac{x-x_0}{d}\right)^2 & \text{for } 0 \le x \le d \\ \eta_s^2 & \text{for } x < 0 \\ \eta_c^2 & \text{for } x > d \end{cases} \quad . \tag{3.31}$$

In the case of a symmetric waveguide $\eta_c = \eta_s$ the index-maximum is found in the middle of the waveguide $x = w_0 = \frac{d}{2}$. If $\Delta\eta = \eta_f - \eta_s$ is small compared to η_f we make no big failure, when we extend the parabolic profile over the interfaces with $\eta^2 = \eta_f - 2\,\eta_f \cdot \Delta\eta \cdot \left(\frac{x}{d} - \frac{1}{2}\right)^2$ for $\eta_f \gg \Delta\eta = \eta_f - \eta_s$. Assuming further that (3.22) is fulfilled for small $\Delta\eta$ we have to solve (3.11) and (3.12) for TM and TE-waves respectively. For simplicity of the following calculations we take $\eta_s \simeq \eta_c$ and make a coordinate transformation on the $x-$axis such that the refractive index profile looks as figure 3.5 shows.

With the new coordinates the refractive index is described by

$$\eta^2 = \eta_f^2 - \left(\eta_f^2 - \eta_s^2\right) \cdot \left(\frac{x}{a}\right)^2 \quad . \tag{3.32}$$

Assuming that (3.22) is met for sufficiently small $\Delta\eta$ (3.11) and (3.12) can be used and have the same form

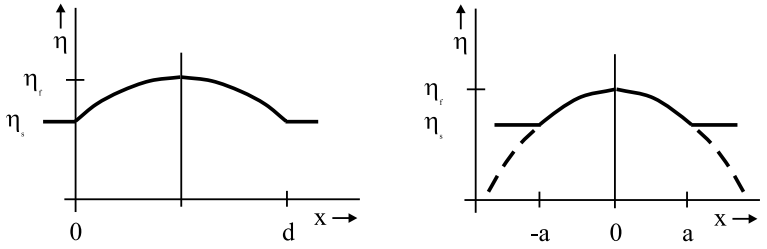

Figure 3.5: Model of refractive index profile of a symmetric graded index planar waveguide. The left sketch shows the model with coordinates as used in section 3.1.2. In the right graph the transformed coordinate system is used.

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi + \left(\eta_\mathrm{f}^2 k_0^2 - k_\mathrm{z} - \left(\eta_\mathrm{f}^2 - \eta_\mathrm{s}^2\right) k_0^2 \left(\frac{x}{a}\right)^2\right)\psi = 0 \quad . \tag{3.33}$$

The overall solution is a superposition of Hermite-Gaussian modes ψ_m

$$\psi = \sum_{m=0}^{\infty} C_m \cdot \psi_m \cdot \exp\{ik_{\mathrm{z}m}z - \omega t\} \tag{3.34}$$

with $k_\mathrm{z} = k_{\mathrm{z}m}$, amplitude C_m and mode profile

$$\psi_m = \mathrm{H}_m\{\frac{x}{p}\} \cdot \exp\left\{-\left(\frac{x}{2\,p}\right)^2\right\} \tag{3.35}$$

for a mode of integer order m. The parameters in (3.34) and (3.35) are

$$k_{\mathrm{z}m}^2 = k^2\,\eta_\mathrm{f}^2 - (2m+1)\frac{k}{a}\sqrt{\eta_\mathrm{f}^2 - \eta_\mathrm{s}^2} \tag{3.36}$$

and

$$\frac{1}{p^2} = \frac{2k}{a}\sqrt{\eta_\mathrm{f}^2 - \eta_\mathrm{s}^2} \quad . \tag{3.37}$$

The Hermite polynomials H_m in (3.35) follow

$$\mathrm{H}_0\{x\} = 1 \tag{3.38}$$

$$\mathrm{H}_1\{x\} = x \tag{3.39}$$

$$\mathrm{H}_{m+1}\{x\} = x \cdot \mathrm{H}_m\{x\} - \mathrm{H}_{m-1}\{x\} \tag{3.40}$$

$$\frac{\partial^2}{\partial x^2}\ \mathrm{H}_m - x\frac{\partial}{\partial x}\ \mathrm{H}_m + m\ \mathrm{H}_m = 0 \tag{3.41}$$

$$\frac{1}{\sqrt{\pi\, m!}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{x^2}{2}\right\} \cdot \mathrm{H}_m\{x\} \cdot \mathrm{H}_n\{x\}\ \mathrm{d}x = \delta_{m,n} \tag{3.42}$$

$$\mathrm{H}_m\{x\} = (-1)^m \cdot \exp\left\{\frac{x^2}{2}\right\} \frac{\mathrm{d}^m}{\mathrm{d}x^m} \cdot \exp\left\{-\frac{x^2}{2}\right\} \quad . \tag{3.43}$$

We have to note at this point that there exists an other widely used representation for the Hermite polynomials

$$\mathrm{H\,e}_m\{\tilde{x}\} = (-1)^m \exp\{\tilde{x}^2\} \cdot \frac{\mathrm{d}^m}{\mathrm{d}x^m}\ \exp\{-\tilde{x}^2\} \tag{3.44}$$

giving

$$\mathrm{H\,e}_0\{\tilde{x}\} = 1 \tag{3.45}$$

$$\mathrm{H\,e}_1\{\tilde{x}\} = 2\,\tilde{x} \quad . \tag{3.46}$$

They follow $\mathrm{H\,e}''_m - 2\,\tilde{x}\ \mathrm{H\,e}'_m + 2\,m\ \mathrm{H\,e}_m = 0$

We can easily transform our solution to this set by $x = \sqrt{2}\,\tilde{x}$ giving

$$\psi = \sum_{m=0}^{\infty} \psi_m \cdot \mathrm{H}\,\mathrm{e}_m \left\{ \sqrt{2}\frac{\tilde{x}}{w} \right\} \cdot \exp\left\{ -\left(\frac{\tilde{x}}{\tilde{w}}\right)^2 \right\} \tag{3.47}$$

with $\tilde{w} = 2p$. The modes m in (3.35) are orthogonal as can be easily see by application of (3.42) giving

$$\frac{1}{\pi\, m!\, n!} \int_{-\infty}^{\infty} \psi_m \cdot \psi_n \ \mathrm{d}\left(\frac{x}{p}\right) = \delta_{m,n} \quad . \tag{3.48}$$

3.1.4 Optical fibers

For cylindrical waveguides we can use vectorpotentials as well as for planar waveguides. As we have seen in section 3.1.2 we can use a matrix representation for the boundary condition to solve the problem. Since the fields E and H can be divided into a superposition of TE and TM portions $\vec{E} = \vec{E}_{\mathrm{TE}} + \vec{E}_{\mathrm{TM}}$, $\vec{E}_{\mathrm{TE}} \circ \vec{E}_{\mathrm{TM}} = 0$ and $\vec{H} = \vec{H}_{\mathrm{TE}} + \vec{H}_{\mathrm{TM}}$, $\vec{H}_{\mathrm{TE}} \circ \vec{H}_{\mathrm{TM}} = 0$ we may try to solve the boundary equations in one step.
First of all we assume for all following cylindrical waveguides a circular cross-section as depicted in figure 3.6.

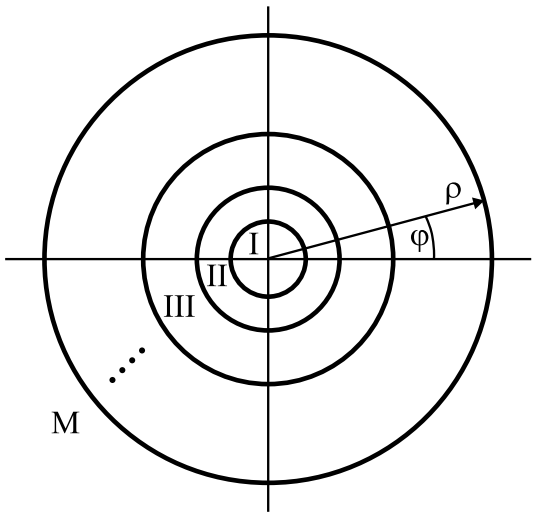

Figure 3.6: General cross-section of cylindrical waveguides with circular apertures

In every part of the cross-section we try to take a potential of the form

$$\psi = \sum_m R_m\{k_\varrho \cdot \varrho\} \cdot \exp\{i\, m\varphi\} \tag{3.49}$$

where ϱ and φ denote the radius from axis and φ is the azimuth. The radial functions R_m have to be regular in every region. This implies for region I which includes $\varrho = 0$ that we have to take Bessel-functions

$$R_{\mathrm{I},m} = C_{\mathrm{I},m} \cdot \mathrm{J}_m\{k_{\varrho\mathrm{I}} \cdot \varrho\} \quad . \tag{3.50}$$

The last region M extends to infinity and we take there Hankel-function

$$R_{\mathrm{M},m} = C_{\mathrm{M},m} \cdot \mathrm{H}_m^{(2)}\{k_{\varrho\mathrm{M}} \cdot \varrho\} \quad . \tag{3.51}$$

For regions between I and M we take ether

$$R_{m,j} = C_{1,m,j} \cdot \mathrm{J}_m\{k_{\varrho_j} \cdot \varrho\} + C_{2,m,j} \cdot \mathrm{N}_m\{k_{\varrho_j} \cdot \varrho\} \tag{3.52}$$

or

$$R_{m,j} = \widetilde{C}_{1,m,j} \cdot \mathrm{H}_m^{(1)}\{k_{\varrho_j} \cdot \varrho\} + \widetilde{C}_{2,m,j} \cdot \mathrm{H}_m^{(2)}\{k_{\varrho_j} \cdot \varrho\} \quad . \tag{3.53}$$

If the fiber incorporates a graded-index core we will try to take a similar model as in section 3.1.3 for the graded index planar waveguide. Waveguides with graded index show lower dispersion at the same radius a where the cladding-refractive-index is reached when many guided modes are incorporated.

3.1.5 Step-Index Optical Fibers

The step-index optical fiber can be modeled as sketched in figure 3.7. We assume the cladding-diameter b to be that large that the evanescent fields do not reach that interface. In this case b can be assumed to be infinite.

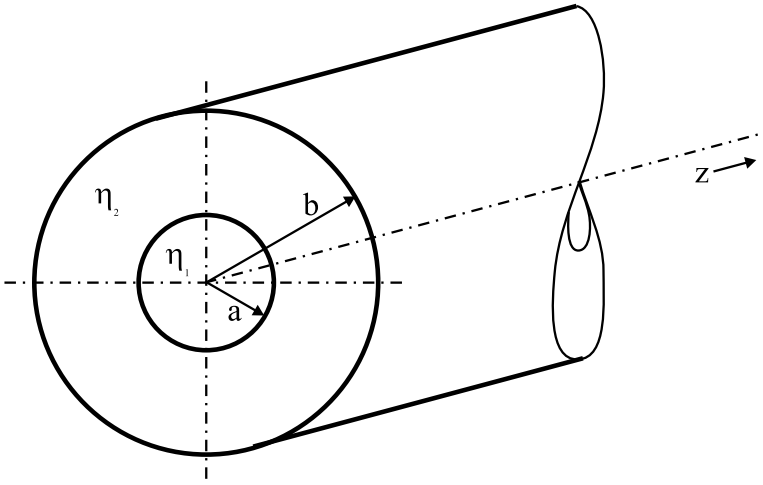

Figure 3.7: Optical fiber with core radius a. The radius of the cladding is $b \gg a$ such that it can be taken to extend to infinity.

We use again $\vec{A}_{\mathrm{TM}} = A_{\mathrm{TM}} \cdot \vec{e}_{\mathrm{z}}$ and $\vec{F}_{\mathrm{TE}} = F_{\mathrm{TE}} \cdot \vec{e}_{\mathrm{z}}$ for both regions and set

$$\begin{aligned}
A_{\mathrm{TM,I}} &= \sum_m A_m \cdot \mathrm{J}_m\{k_{\varrho 1} \cdot \varrho\} \cdot \exp\{i\, m\varphi\} \exp\{i(k_{\mathrm{z}m} z - \omega t)\} \\
F_{\mathrm{TE,I}} &= \sum_m B_m \cdot \mathrm{J}_m\{k_{\varrho 1} \cdot \varrho\} \cdot \exp\{i\, m\varphi\} \exp\{i(k_{\mathrm{z}m} z - \omega t)\} \\
A_{\mathrm{TM,II}} &= \sum_m C_m \cdot \mathrm{K}_m\{k_{\varrho 2} \cdot \varrho\} \cdot \exp\{i\, m\varphi\} \exp\{i(k_{\mathrm{z}m} z - \omega t)\} \\
F_{\mathrm{TE,II}} &= \sum_m D_m \cdot \mathrm{K}_m\{k_{\varrho 2} \cdot \varrho\} \cdot \exp\{i\, m\varphi\} \exp\{i(k_{\mathrm{z}m} z - \omega t)\} \quad . \qquad (3.54)
\end{aligned}$$

In the following discussion we regard only one azimuthal order m and take for simplicity $k_{\mathrm{z}m} = k_{\mathrm{z}}$ as well as $A = A_m, B = B_m, C = C_m$ and $D = D_m$. From (3.54) follows the total field in both regions I and II for each azimuthal order m

$$\begin{aligned}
\vec{E} = \Bigg[& \left(i\, C_{\mathrm{H}} \frac{m}{\varrho\, \tilde{\varepsilon}\, \varepsilon_0} R - C_{\mathrm{E}}\, \omega\, \frac{k_{\mathrm{z}} k_\varrho}{k^2} R' \right) \cdot \vec{e}_\varrho \\
& - \left(C_{\mathrm{H}} \frac{k_\varrho}{\tilde{\varepsilon}\, \varepsilon_0} R' + C_{\mathrm{E}}\, m \frac{\omega k_{\mathrm{z}}}{k^2 \varrho} R \right) \cdot \vec{e}_\varphi - i\, C_{\mathrm{E}}\, \omega \frac{k_\varrho^2}{k^2} R \cdot \vec{e}_{\mathrm{z}} \Bigg] \cdot \exp\{i(\, m\varphi + k_{\mathrm{z}}\, z - \omega t)\} \\
\vec{H} = \Bigg[& \left(i\, C_{\mathrm{E}} \frac{m}{\mu\, \mu_0\, \varrho} R + C_{\mathrm{H}}\, \omega\, \frac{k_{\mathrm{z}} k_\varrho}{k^2} R' \right) \cdot \vec{e}_\varrho
\end{aligned}$$

$$+\left(i\,C_{\mathrm{H}}\,m\,\frac{\omega k_{\mathrm{z}}}{k^2\varrho}R - C_{\mathrm{E}}\frac{k_\varrho}{\mu\mu_0}R'\right)\cdot\vec{e}_\varphi - i\,C_{\mathrm{H}}\,\omega\frac{k_\varrho^2}{k^2}R\cdot\vec{e}_{\mathrm{z}}\Big]\cdot\exp\{i(\,m\varphi + k_{\mathrm{z}}\,z - \omega t)\}$$

with $R'\{x\} = \frac{\partial}{\partial x}\,R\{x\}$, $R\{k_\varrho\,\varrho\} \in \{\mathrm{J}_m\{k_\varrho\,\varrho\}, \mathrm{K}_m\{k_\varrho\,\varrho\}\}$, $C_{\mathrm{H}} \in \{B, D\}$, $C_{\mathrm{E}} \in \{A, C\}$ depending on the region I or II.

Knowing that different azimuthal orders m are not possible in adjacent layers due to the boundary conditions at $\varrho = a$ which follows from (3.16) and (3.17) we find

$$\begin{aligned}\left(\vec{E}_{\mathrm{I}} - \vec{E}_{\mathrm{II}}\right)\circ\vec{e}_{\mathrm{z}} &= 0\\ \left(\vec{H}_{\mathrm{I}} - \vec{H}_{\mathrm{II}}\right)\circ\vec{e}_{\mathrm{z}} &= 0\\ \left(\vec{E}_{\mathrm{I}} - \vec{E}_{\mathrm{II}}\right)\circ\vec{e}_{\varphi} &= 0\\ \left(\vec{H}_{\mathrm{I}} - \vec{H}_{\mathrm{II}}\right)\circ\vec{e}_{\varphi} &= 0 \quad .\end{aligned} \tag{3.55}$$

If we try to take pure TE or TM waves there is no solution for higher orders m. When we take the superposition of both types solutions are possible. Applying (3.28) to (3.27) we find the condition

$$0 = [M]\cdot(A, B, C, D)^T \tag{3.56}$$

with matrix

$$M = \begin{pmatrix} \frac{k_{\varrho 1}^2}{k_1^2}\,\mathrm{J}_m\{k_{\varrho 1}a\} & 0 & \frac{k_{\varrho 2}^2}{k_2^2}\,\mathrm{K}_m\{-i\,k_{\varrho 2}a\} & 0\\ 0 & \frac{k_{\varrho 1}^2}{k_1^2}\,\mathrm{J}_m\{k_{\varrho 1}a\} & 0 & \frac{k_{\varrho 2}^2}{k_2^2}\,\mathrm{K}_m\{-i\,k_{\varrho 2}a\}\\ \frac{k_{\varrho 1}^2}{\mu_1\,\mu_0}\,\mathrm{J}'_m\{k_{\varrho 1}a\} & i\,n\frac{\omega k_{\mathrm{z}}}{k_1^2 a}\,\mathrm{J}_m\{k_{\varrho 1}a\} & i\frac{k_{\varrho 2}^2}{\mu_2\,\mu_0}\,\mathrm{K}'_m\{-i\,k_{\varrho 2}a\} & -i\,n\frac{\omega k_{\mathrm{z}}}{k_2^2 a}\,\mathrm{K}_m\{-i\,k_{\varrho 2}a\}\\ i\,n\frac{\omega k_{\mathrm{z}}}{k_1^2 a}\,\mathrm{J}_m\{k_{\varrho 1}a\} & \frac{k_{\varrho 1}}{\tilde{\varepsilon}_1\,\varepsilon_0}\,\mathrm{J}'_m\{k_{\varrho 1}a\} & -i\,n\frac{\omega k_{\mathrm{z}}}{k_2^2 a}\,\mathrm{K}_m\{-i\,k_{\varrho 2}a\} & i\frac{k_{\varrho 2}}{\tilde{\varepsilon}_2\,\varepsilon_0}\,\mathrm{K}'_m\{-i\,k_{\varrho 2}a\} \end{pmatrix} \tag{3.57}$$

Like for the planar waveguide we find nontrivial solutions for the coefficients A, B, C, D in (3.56) when $\mathrm{Det}\{[M]\} = 0$.

For guided modes $k_{\varrho 2}$ is imaginary. It is common to substitute

$$u = k_{\varrho 1}\, a$$
$$w = -i\, k_{\varrho 2} a$$

giving the fiber parameter

$$V^2 = u^2 + w^2 = k^2 a^2 \cdot \left(\eta_1^2 - \eta_2^2\right)$$

and the phase parameter

$$B = \frac{w^2}{u^2 + v^2} = \frac{w^2}{V^2} = \frac{k_z^2 - k^2 \eta_2^2}{\eta_1^2 - \eta_2^2} \quad .$$

For the lowest order mode $m = 0$ no coupling between $\{A, C\}$ and $\{B, D\}$ exist. In this case we have pure TE- and TM-waves. We then have to solve

$$[M_{\mathrm{TM}}] \begin{pmatrix} A \\ C \end{pmatrix} = 0 \tag{3.58}$$
$$[M_{\mathrm{TE}}] \begin{pmatrix} B \\ D \end{pmatrix} = 0 \tag{3.59}$$

with

$$[M_{\mathrm{TE}}] = \begin{pmatrix} \frac{u^2}{(k_1 a)^2}\, \mathrm{J}_0\{u\} & \frac{-w^2}{(k_2 a)^2}\, \mathrm{K}_0\{w\} \\ \frac{u}{a\tilde{\varepsilon_1}\, \varepsilon_0}\, \mathrm{J}_0'\{u\} & -\frac{w}{a\tilde{\varepsilon_2}\, \varepsilon_0}\, \mathrm{K}_0'\{w\} \end{pmatrix} \tag{3.60}$$
$$[M_{\mathrm{TM}}] = \begin{pmatrix} \frac{u^2}{(k_1 a)^2}\, \mathrm{J}_0\{u\} & \frac{-w^2}{(k_2 a)^2}\, \mathrm{K}_0\{w\} \\ \frac{u}{a\mu_1\, \mu_0}\, \mathrm{J}_0'\{u\} & -\frac{w}{a\mu_2\, \mu_0}\, \mathrm{K}_0'\{w\} \end{pmatrix} \tag{3.61}$$

Satisfaction of (3.57) requires $\mathrm{Det}\{[M_{\mathrm{TE}}]\} = 0$ and $\mathrm{Det}\{[M_{\mathrm{TM}}]\} = 0$ for non vanishing fields (i.e. $A, B, C, D \neq 0$). For TE- and TM-waves the conditions

$$\frac{u^2 w}{\mu_1}\,\mathrm{J}_0\{u\}\frac{\mathrm{d}}{\mathrm{d}w}\,\mathrm{K}_0\{w\} = \frac{w^2 u}{\mu_2}\,\mathrm{K}_0\{w\}\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{J}_0\{u\} \tag{3.62}$$

and

$$\frac{u^2 w}{\tilde{\varepsilon}_1}\,\mathrm{J}_0\{u\}\frac{\mathrm{d}}{\mathrm{d}w}\,\mathrm{K}_0\{w\} = \frac{w^2 u}{\tilde{\varepsilon}_2}\,\mathrm{K}_0\{w\}\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{J}_0\{u\} \tag{3.63}$$

have to be met. Note that for cylinder-functions $Y_m \,\epsilon\{\mathrm{J}_m, \mathrm{K}_m\}$ the first derivative can be written as

$$\frac{\mathrm{d}}{\mathrm{d}x}\,Y_m\{x\} = Y' = -Y_{m+1}\{x\} + \frac{m}{x}\,Y_m\{x\} \tag{3.64}$$

and follow the recursion

$$Y_{m+1}\{x\} = \frac{2m}{x}\,Y_m - Y_{m-1} \quad . \tag{3.65}$$

With substitution of the derivatives in (3.62) and (3.63) by (3.64) the eigenvalue-equations for $u \cdot w \neq 0$ are

$$\frac{u\,\mathrm{J}_0\{u\}}{\mathrm{J}_1\{u\}} = \frac{w\,\mathrm{K}_0\{w\}}{\mathrm{K}_1\{w\}} \cdot \begin{cases} \frac{\mu_1}{\mu_2} \text{ for TE-waves} \\ \frac{\varepsilon_1}{\varepsilon_2} \text{ for TM-waves} \end{cases} . \tag{3.66}$$

For higher order modes $m \neq 0$ the matrix $[M]$ can be rewritten with substitutions u and w and $k_\mathrm{z} = \frac{1}{a}\sqrt{(k_1 a)^2 - u^2} = \frac{1}{a}\sqrt{(k_2 a)^2 - w^2}$. It then has the form

$$[M] = \begin{pmatrix} \left(\frac{u}{k_1 a}\right)^2 \mathrm{J}_m & 0 & -\left(\frac{w}{k_2 a}\right)^2 \mathrm{K}_m & 0 \\ 0 & \left(\frac{u}{k_1 a}\right)^2 \mathrm{J}_m & 0 & -\left(\frac{w}{k_2 a}\right)^2 \mathrm{K}_m \\ \frac{u}{a\mu_1\,\mu_0}\,\mathrm{J}'_m & i\,m\frac{\omega\sqrt{(k_1 a)^2-u^2}}{(k_1 a)^2}\,\mathrm{J}_m & -\frac{w}{a\mu_2\,\mu_0}\,\mathrm{K}'_m & -i\,m\frac{\omega\sqrt{(k_2 a)^2-u^2}}{(k_2 a)^2}\,\mathrm{K}_m \\ i\,m\frac{\omega\sqrt{(k_1 a)^2-u^2}}{(k_1 a)^2}\,\mathrm{J}_m & \frac{u}{a\tilde{\varepsilon}_1\,\varepsilon_0}\,\mathrm{J}'_m & i\,m\frac{w\sqrt{(k_2 a)^2-w^2}}{(k_2 a)^2}\,\mathrm{K}_m & -\frac{w}{a\tilde{\varepsilon}_2\,\varepsilon_0}\,\mathrm{K}'_m \end{pmatrix} \tag{3.67}$$

and from $\mathrm{Det}\{[M]\} = 0$ follows the exact solution for this approach.
We find the eigenvalue condition

$$\begin{aligned} 0 = m^2 &\left(\frac{u}{k_1 a}\sqrt{\frac{1}{w^2}+\left(\frac{1}{k_2 a}\right)^2} + \frac{w}{k_2 a}\sqrt{\frac{1}{u^2}+\left(\frac{1}{k_1 a}\right)^2}\right)^2 \cdot \left(\frac{u}{k_1 a}\cdot\frac{w}{k_2 a}\right)^2 \\ &+ \left(\frac{\mathrm{K}'_m}{\mathrm{K}_m}\frac{u}{k_1 a}\sqrt{\frac{\tilde{\varepsilon}_1}{\mu_1}} - \frac{\mathrm{J}'_m}{\mathrm{J}_m}\frac{w}{k_2 a}\sqrt{\frac{\tilde{\varepsilon}_2}{\mu_2}}\right)\left(\frac{\mathrm{K}'_m}{\mathrm{K}_m}\frac{u}{k_1 a}\sqrt{\frac{\mu_1}{\tilde{\varepsilon}_1}} - \frac{\mathrm{J}'_m}{\mathrm{J}_m}\frac{w}{k_2 a}\sqrt{\frac{\mu_1}{\tilde{\varepsilon}_2}}\right)\left(\frac{u}{k_1 a}\frac{w}{k_2 a}\right)^2 \end{aligned} \tag{3.68}$$

The solutions for $m = 0$ are found in the second summand. For TE-waves the difference in the first parentheses pair is zero and for TM-waves the same applys to the second pair. For higher order modes the analytical solution of (3.68) is very complicated.
Normally $\mathrm{Det}\{[M]\} = 0$ is solved numerical. The variables u and w are replaced by B and V with $w = B \cdot V^2$ and $u = V\sqrt{1-B}$. The way to solve the eigenvalue equation is to give a B and search for a corresponding V.
When the matrix-elements are calculated it is found that the elements $m_{3,2}$, $m_{3,4}$, $m_{4,1}$ and $m_{4,3}$ have small values compared to the other ones even for big numbers of the refractive index η provided the refractive index contrast $\eta_1^2 - \eta_2^2$ is small. This is normally found in fibers made of quartz glas. If we take $\varepsilon_1 \simeq \varepsilon_2$ and $\mu_1 \simeq \mu_2$ this

elements can be set to zero and we find pure TE- and TM-waves again as approximation for the true waves. Using (3.64) we find

$$\left(\tilde{\varepsilon}_2\frac{\mathrm{K}_{m+1}}{w\,\mathrm{K}_m}-\tilde{\varepsilon}_1\frac{\mathrm{J}_{m+1}}{u\,\mathrm{J}_m}\right)u^2w^2=m\,\tilde{\varepsilon}_1\,\tilde{\varepsilon}_2\,a^2\cdot\left[k_0^2(\mu_1-\mu_2)+k_\mathrm{z}^2\left(\frac{1}{\tilde{\varepsilon}_1}-\frac{1}{\tilde{\varepsilon}_2}\right)\right] \quad (3.69)$$

as eigenvalue equation for TE-waves where we can set the right side to zero when we apply $\tilde{\varepsilon}_1 \simeq \tilde{\varepsilon}_2$ and $\mu_1 \simeq \mu_2$ as above. A similar equation is found for the TM-waves. For both TE- and TM-waves we have the eigenvalue condition

$$\frac{u\,\mathrm{J}_m}{\mathrm{J}_{m+1}}=\frac{w\,\mathrm{K}_m}{\mathrm{K}_{m+1}} \quad (3.70)$$

In (3.70) the right side is slowly varying with w whereas the left expression oscillates and takes several times the same value. This gives for each azimuthal order m a set of many possible solutions.

All this approximate solutions are linearly polarized and therefore called $LP_{m,n}$-modes. Weakly guided modes extend widely into the cladding having a small $w \to 0$. In this case we find $\mathrm{J}_m\{u_{m,n}\}=0$ for the $LP_{m,n}$-waves. On the other hand for strongly guided modes the fields are nearly only in the core with $w \gg 1$. Now u for the $LP_{m,n}$-waves is determined by $\mathrm{J}_{m+1}\{u_{m,n}\}=0$. The guided modes start as weakly guided ones with $k_\mathrm{z} \simeq \eta_1 k_0$ and for high frequencies they travel nearly only in the core with $k_\mathrm{z} \simeq \eta_2 k_0$. For a $LP_{m,n}$-wave with azimuthal order m therefore the values of u are in the range

$$\begin{aligned} 0 \le u \le u_{m,1} &\text{ for } n=1 \\ u_{m+1,n-1} \le u \le u_{m,n} &\text{ for } n>1 \end{aligned} \quad (3.71)$$

were $\mathrm{J}_m\{u_{m,n}\}=0$ and n denotes the $n-$th zero of a Bessel-function. The first zero of J_0 is $u_{0,1}=2.405$, the first zero of J_1 is $u_{1,1}=3.83$. The lowest order mode starts at $u=0$ which is a trivial solution for (3.68). This gives for $LP_{0,1}$:

$$0 \leq u \leq 2.405$$

and for $LP_{m,1}$:

$$0 \leq u \leq u_{m,n}$$

Exact solutions of (3.68) are small deviations from the LP-modes. They are superpositions of for example of a strong TE-mode and a small TM-mode which means $A \ll B$ and $C \ll D$. TE-waves are also called $H-$waves because they have only a $H-$component in $z-$direction. TM-waves are also called $E-$waves. The exact solution are called $EH-$modes when they are mainly TM ($A \gg B$) and HE when they are mainly TE ($A \ll B$). It can be shown that the approximative $LP_{m,n}-$waves and the exact $HE_{m+1,n}$ and $EH_{m-1,n}$ waves have nearly the same k_z and therefore the same curves in the $u - w-$diagram as shown in figure 3.8.
For $m = 0$ we have seen this already under the assumption that there is no big difference in the refractive indices. In a given structure with fixed wavelength $V = \sqrt{u^2 + w^2}$ is constant. Intersections of V and $u - w-$curves are the solution of (3.68).

3.1.6 Graded-Index Optical Fibers

As pointed out before graded index waveguides exhibit less dispersion than a comparable multi-mode step-index waveguide. We first want to consider an index with quadratic radius dependence and then look for the optimal profile with respect to dispersion. In both cases we take the variation of index to be very small like we assumed in section 3.1.3 and we start form (3.11) or (3.12) which in cylindrical coordinates have the form

$$\frac{\mathrm{d}^2}{\mathrm{d}\varrho^2}\psi + \frac{1}{\varrho} \cdot \frac{\mathrm{d}}{\mathrm{d}\varrho}\psi + \frac{1}{\varrho^2} \cdot \frac{\mathrm{d}^2}{\mathrm{d}\varphi^2}\psi + k_\varrho^2 \psi = 0 \tag{3.72}$$

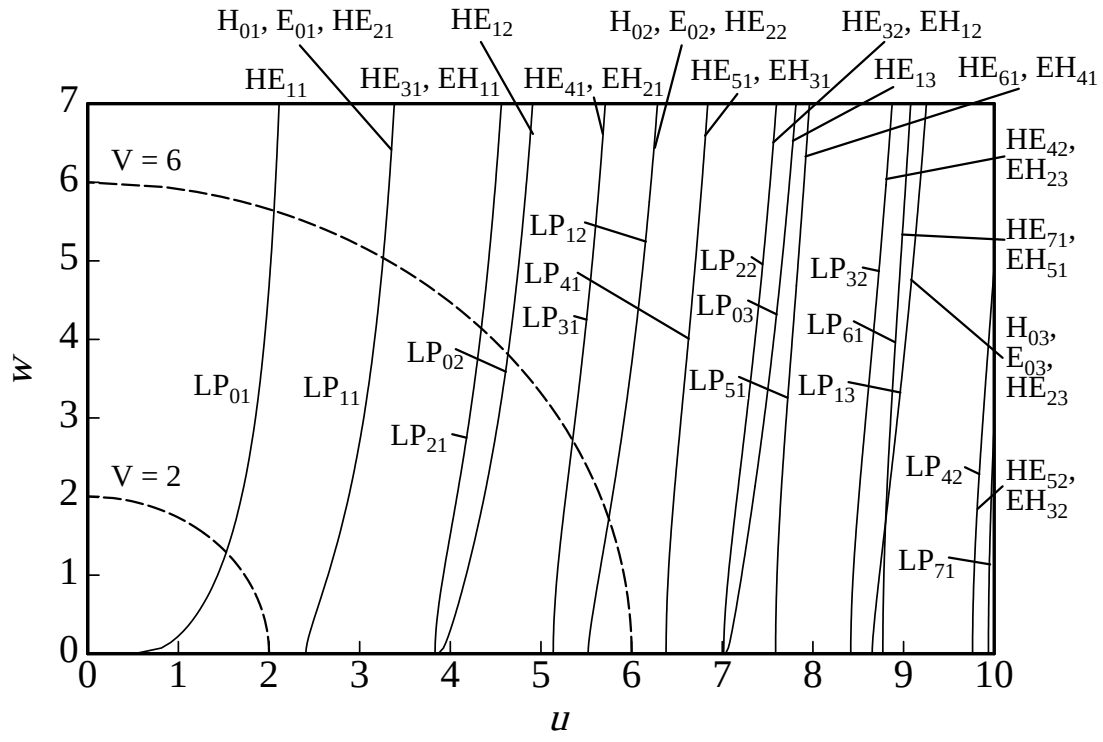

Figure 3.8: $u - w$ curves for the approximate LP modes. It was assumed that $\tilde{\varepsilon}_1 \simeq \tilde{\varepsilon}_2$ and $\mu_1 \simeq \mu_2$. Intersections between the $u - w$ curves and V give the solution for a given structure and wavelength.

and with

$$\eta^2 = \eta_1^2 - \left(\eta_1^2 - \eta_2^2\right) \cdot \left(\frac{\varrho}{a}\right)^2 \tag{3.73}$$

k_ϱ^2 can be written as

$$k_\varrho^2 = k_0^2\eta_1^2 - k_z^2 - k_0^2\left(\eta_1^2 - \eta_2^2\right)\left(\frac{\varrho}{a}\right)^2 \quad . \tag{3.74}$$

From section 3.1.3 we know that parabolic index profiles support gaussian waves. We try

$$\psi_m = R\left\{\tfrac{\varrho}{\varrho_1}\right\} \cdot \exp\{i\, m\varphi\} \cdot \exp\left\{-\left(\frac{\varrho}{w}\right)^2\right\} \tag{3.75}$$

and find for R

$$R\{\tfrac{\varrho}{\varrho_1}\} = \left(\frac{\sqrt{2}\,\varrho}{w}\right)^m \cdot \mathrm{L}_p^{(m)}\left\{2\left(\frac{\varrho}{w}\right)^2\right\} \tag{3.76}$$

with

$$\frac{1}{w^2} = \frac{k_0}{2a} \cdot \sqrt{\eta_1^2 - \eta_2^2} \tag{3.77}$$

and

$$k_{zm,p}^2 = k_0^2\eta_1^2 - (2p + m + 1)\frac{2k_0}{a}\sqrt{\eta_1^2 - \eta_2^2} \quad . \tag{3.78}$$

The function $\mathrm{L}_p^{(m)}$ denotes Laguerre polynomials that follow

$$\begin{aligned}
\mathrm{L}_0^{(m)}\{x\} &= 1 \\
\mathrm{L}_1^{(m)}\{x\} &= 1 + m - x \qquad (3.79)\\
\mathrm{L}_{p+1}^{(m)}\{x\} &= \frac{1}{p+1}\left[(1 + m + 2p - x) \cdot \mathrm{L}_p^{(m)}\{x\} - (p + m) \cdot \mathrm{L}_p^{(m)}\{x\}\right] \qquad (3.80)
\end{aligned}$$

where 3.80 is valid for $p \geq 1$. They are orthogonal in the form

$$\frac{1}{\Gamma\{1+m\}\binom{p+m}{p}} \cdot \int_0^\infty \mathrm{L}_p^{(m)}\{x\} \cdot \mathrm{L}_{p'}^{(m)}\{x\} \cdot x^m \cdot \exp\{-x\}\ \mathrm{d}x \quad . \tag{3.81}$$

Both vector potentials therefore result to

$$A, F = \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} C_{m,p} \cdot \psi_{m,p} \cdot \exp\{i\, k_{\mathrm{z}m,p} z - \omega t\} \tag{3.82}$$

with

$$\psi_{m,p} = \mathrm{L}_p^{(m)} \left\{ 2 \left(\frac{\varrho}{w} \right)^2 \right\} \cdot \left(\frac{\sqrt{2}\,\varrho}{w} \right)^m \cdot \exp\left\{ - \left(\frac{\varrho}{w} \right)^2 \right\} \cdot \exp\{i\, m\varphi\} \quad . \tag{3.83}$$

and amplitudes $C_{m,p}$.

3.1.7 Dispersion in Fibers

We want to estimate the dispersion in a multi mode graded index fiber. From general mode theory we know that signal transport velocity in a non dispersive medium ($\frac{\mathrm{d}}{\mathrm{d}\omega}\,\varepsilon = 0$, $\frac{\mathrm{d}}{\mathrm{d}\omega}\,\mu = 0$) is the group velocity c_{gr}. The inverse is signal traveling time per unit length

$$\tau = \frac{1}{c_{\mathrm{gr}}} = \frac{\mathrm{d}}{\mathrm{d}\omega}\, k_{\mathrm{z}} \quad . \tag{3.84}$$

For graded index fibers we find in (3.78) an expression for k_{z}. Taking the derivative of (3.78) results into

$$\tau = \frac{k_0 \eta_1}{k_{\mathrm{z}}} \left(\frac{\eta_1}{c_0} + k_0 \cdot \frac{\mathrm{d}}{\mathrm{d}\omega}\, \eta_1 \right) + \frac{\eta_1^2 k_0^2 - k_{\mathrm{z}}^2}{2 k_{\mathrm{z}}} \left(\frac{1}{\omega} - \frac{\eta_1 \frac{\mathrm{d}}{\mathrm{d}\omega}\, \eta_1 - \eta_2 \frac{\mathrm{d}}{\mathrm{d}\omega}\, \eta_2}{\eta_1^2 - \eta_2^2} \right)$$

assuming that signal transport in a low dispersive medium does not deviate too much from the group velocity. Using $\frac{\mathrm{d}}{\mathrm{d}\omega}\lambda = -\frac{\lambda}{\omega}$ and the group index $\eta' = \eta - \lambda \cdot \frac{\mathrm{d}}{\mathrm{d}\lambda}\eta$ we can simplify

$$\tau = \frac{1}{c_0}\left(\frac{\eta_1 k_0}{k_\mathrm{z}}\eta_1' + \frac{\eta_1^2 k_0^2 - k_\mathrm{z}^2}{2k_0 k_\mathrm{z}} \cdot \frac{\eta_1\eta_1' - \eta_2\eta_2'}{\eta_1^2 - \eta_2^2}\right) \quad . \tag{3.85}$$

High order modes are traveling near $k_\mathrm{z} = \eta_2 \cdot k_0$ in the cladding and low order modes near $k_\mathrm{z} = \eta_1 \cdot k_0$ are concentrated to the core. The time difference for signal transport is

$$\Delta\tau = \frac{1}{c_0}\left(\eta_1'\frac{\eta_1 - \eta_2}{\eta_2} + \frac{\eta_1\eta_1' - \eta_2\eta_2'}{\eta\eta_2}\right) = \frac{3\eta_1\eta_1' - \eta_2(\eta_1' + \eta_2')}{2\eta_2 c_0} \quad .$$

For comparison we want to calculate the same value for a step index fiber. Here we have (3.70) as implicit equation for k_z in EH and HE modes. At $z = 0$ $LP_{n,m}$- modes are exited which are a superposition of $HE_{n-1,m}$ and $EH_{n+1,m}$ modes. The amplitudes $C_{\mathrm{H}_{n-1,m}}$ and $C_{\mathrm{E}_{n+1,m}}$ are correlated and the signals are carried by two modes. With intensive calculations it can be shown that the lowest order modes in the multimode fiber have $\tau = \frac{\eta_1'}{c_0}$ and the highest order modes transport signals with $\tau = \frac{\eta_2'}{c_0}$ giving

$$\Delta\tau = \frac{1}{c_0}\left(\eta_1' - \eta_2'\right)$$

which for quartz fibers is bigger than $\Delta\tau$ for graded index quartz fibers as calculated before.

Going back to (3.85) the question arises if there is an optimum for k_z. In the case of a parabolic profile the optimum lies out of the possible range for guided modes. But if we assume

$$\eta^2 = \eta_1^2 - (\eta_1^2 - \eta_2^2) \cdot \left(\frac{\varrho}{a}\right)^g \tag{3.86}$$

we can find an optimized g where the dispersion of the fiber can be further minimized. This special form of index profile is called power law profile. We cannot find an analytical solution for the vector potential but for the calculation of dispersion it is not necessary to know the exact fields. It is sufficient to have a good approximation of k_z which is directly dependent on k_ρ due to the dispersion relation (3.5). In several cases $k_r ho$ and therefore k_z is known as an implicit expression $I\{k_\mathrm{z}, \omega\} = 0$. With Taylor expansion $I\{k_\mathrm{z}, \omega\} \simeq \frac{\partial}{\partial\omega} I \,\mathrm{d}\omega + \frac{\partial}{\partial k_\mathrm{z}} I \,\mathrm{d}k_\mathrm{z}$ the disperion is calculated by

$$\tau = \frac{\mathrm{d}}{\mathrm{d}\omega} k_\mathrm{z} \simeq \frac{\frac{\partial}{\partial k_\mathrm{z}} I}{\frac{\partial}{\partial\omega} I} \quad .$$

For the power law profile we try to find I analog to (3.30) with

$$\int_{a_1}^{a_2} k_{\varrho m} \,\mathrm{d}\varrho = m\pi + \frac{1}{2}(\Phi\{a_1\} + \Phi\{a_2\})$$

where a_1 and a_2 are the inner and outer reversing points $k_{\varrho m}\{a_1\} = 0$. In radial modes there is no inner turning point and we have to set $a_1 = 0$ and $\Phi\{a_1\} = 0$, in helix modes both points a_1 and a_2 exists and the phase change is $\Phi = -\frac{\pi}{2}$ because we there find the transition form transverse propagation to evanescent declining. Due to (3.5)

$$\int_{a_1}^{a_2} k_{\varrho m}\{\varrho\} \,\mathrm{d}\varrho = (m + \frac{1}{2})\pi \tag{3.87}$$

with azimuthal dependence $\exp\{i\, m\varphi\}$ and therefore

$$k_{\varrho m} = k_\varrho^2 - \left(\frac{m}{\varrho}\right)^2 \tag{3.88}$$

is the implicit equation for k_z. Differentiation of (3.5) and (3.87) yields the dispersion

$$\tau = \frac{1}{c_0} \cdot \frac{k_0}{k_\mathrm{z}} \cdot \frac{\int_{a_1}^{a_2} \frac{\eta\cdot\eta'}{k_{\varrho m}} \,\mathrm{d}\varrho}{\int_{a_1}^{a_2} \frac{1}{k_{\varrho m}} \,\mathrm{d}\varrho} \quad . \tag{3.89}$$

Low order modes are traveling with $k_z \simeq \eta_1 \cdot k_0$ giving $k_{\varrho m} \simeq 0$ and $a_1 \simeq a_2$ and therefore $\tau \simeq \frac{\eta_1'}{c_0}$. We want to choose g such that τ for high order modes equals that for low oder modes. There is no analytical solution for $k_z < \eta_1 \cdot k_0$.
The way to find an appropriate expression for k_ρ in high order modes follows a wired way by counting the number of guided modes. This number is assumed to be constant with a small change in light frequency resulting into the required value for τ. For more information refer to appendix B. The optimal profile exponent g follows to be

$$g_{\text{opt}} = 2 \cdot \frac{\eta_2}{\eta_1} \cdot \frac{(\eta_1' - \eta_2')}{(\eta_1 - \eta_2)} \quad .$$

With g_{opt} the highest order modes travel as fast as the lowest order modes and the dispersion for optimal power law profile results to

$$\Delta\tau \simeq -\frac{1}{8}\left(\frac{\eta_1 - \eta_2}{\eta_1}\right)^2 \frac{\eta_1'}{c_0} \quad .$$

τ gives only the time a signal travels per unit length. Under modulation we have to consider not only a single frequency for the light but a spectrum of modulation width $\Delta\omega$ around the carrier ω. Assuming small bandwidth $\Delta\omega$ compared to ω we can state

$$\tau \simeq \tau\Big|_\omega + \frac{\mathrm{d}}{\mathrm{d}\omega}\tau\Big|_\omega \cdot \Delta\omega = \tau\Big|_\omega + \frac{\mathrm{d}}{\mathrm{d}\lambda}\tau \cdot \Delta\lambda \tag{3.90}$$

where $\frac{\mathrm{d}}{\mathrm{d}\omega}\tau$ gives the dispersion in the fiber. Due to the small diameter required for a single mode fiber, i.e. $2a < \lambda\frac{2.405}{\pi} \cdot \frac{1}{\sqrt{\eta_1^2 - \eta_2^2}}$ for step index fibers, the fabrication of a graded index profile is very difficult and usually only step index fibers are used for single mode transmission. Taking $\frac{\mathrm{d}}{\mathrm{d}\lambda}(\eta_1' - \eta_2') \simeq 0$ in weakly guiding fibers and using V and B we have

$$\tau' = \frac{\mathrm{d}}{\mathrm{d}\lambda}\tau \simeq \frac{1}{c_0}\left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\eta_1' - \frac{\eta_1' - \eta_2'}{\lambda} \cdot V \cdot \frac{\mathrm{d}^2}{\mathrm{d}V^2}(VB)\right) \quad . \tag{3.91}$$

The first term is called material dispersion, the second originates form wave guiding and is called waveguide dispersion. In silica fibers $\frac{\mathrm{d}}{\mathrm{d}\lambda}\eta_1 > 0$ for $\lambda > 1.27\mu$m and $\frac{\mathrm{d}^2}{\mathrm{d}V^2}VB > 0$ for $V < B$. It is therefore possible to compensate the total chromatic dispersion τ' for wavelength $\lambda > 1.27\mu$m. In standard silica single mode fibers the dispersion minimum is located at 1.3μm. With enhanced profiles the maximum in wave guide dispersion can be shifted to the attenuation minimum at 1.55μm. Such fibers are called dispersion shifted fibers. When the refractive index profile has the form of a W as sketched in figure 3.9 the total chromatic dispersion τ' exhibits two zeros and is very small in the spectral region between the zeros. This leads to the name dispersion flattened fiber.

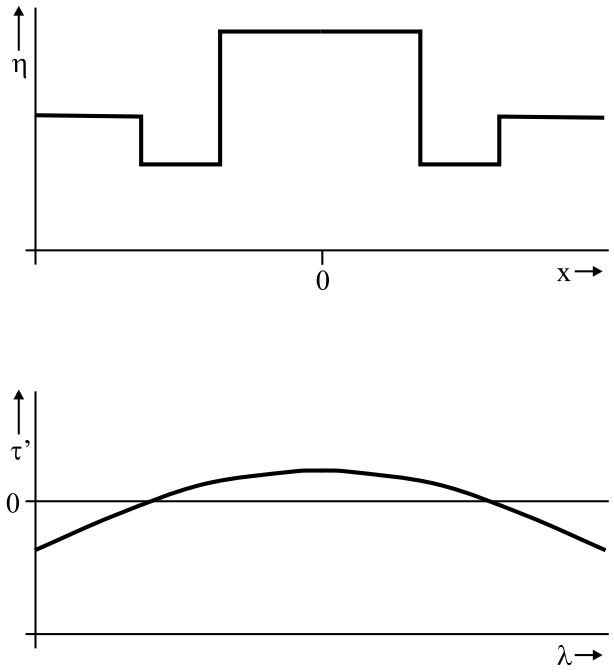

Figure 3.9: Refractive index of a dispersion flattened fiber. Because the index profile looks like a W it is also called W- profile fiber. The total chromatic dispersion exhibits two zeros and is very small between that zeros.

3.1.8 Optical Fiber Losses

There are two basic mechanisms of losses in fibers: scattering and absorption. Absorption is a material characteristic as described in section 1.2.3. In most fibers intentionally and non intentionally impurities are incorporated which build states that can optically be exited and recombine nonradiative thus leading to attenuation. In standard silica fibers water (OH)molecules are the severest problem leading to high attenuation 1.2μm and 1.4μm wavelength. On the long wavelength side resonances of the fiber material limit the usable spectral range leading to minimal losses in standard silica fibers at 1.55μm. With heavy metal fluorid fibers the material resonances are at bigger wavelength and therefore such fibers exhibit an absorption minimum near 2μm and higher.

The most popular scattering mechanism in fibers is Rayleigh scattering which takes place when a wave incidents a small volume of different refractive index. Such variations are technological problems coming from nonhomogeneous glass melts when the fiber is drawn leading to attenuation proportional to λ^{-4} and thus limiting the useful spectral range on the short wavelength side.

A second scattering mechanism is bending of a fiber as schematically sketched in figure 3.10.

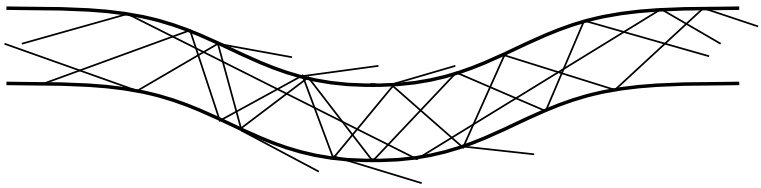

Figure 3.10: Bend in an optical fiber. At the bend the optical ray changes to angle of incidence which can be explained as a transition to another propagation constant or even to another modal order. Modes with lower propagation constants or higher orders exhibit higher losses because the extend more into the cladding.

When the wave enters the bend the angle of incidence is higher and therefore the

propagation constant is lowered. The wave extends more into the cladding and in some cases is not any longer guided by total internal reflection leading to losses. In multimode fibers also the modal order can change at a bend which is referred to as mode mixing. Higher order modes are likely to be transferred to non guided waves (so called radiative waves) because they are near their on-off value of V. This circumstance leads to the fact that the overall attenuation at bends in a multimode fiber is higher than that for singlemode fibers. In real optical fiber cables there are thousands of microscopic bends. In singlemode fibers the attenuation is proportional to the cablelength.

A simple model for waveguide bend attenuation is given later in section 3.5 and shows that bend angles should be less than 1° to avoid excessive attenuation. The attenuation behaviour of multimode fibers can only be described by the bending spectrum and statistical coupling of modes. The result is that for the first kilometer the length dependence is not linear and after that length the attenuation growth linear cable length due to a somewhat steady state mode mixing.

Scattering also occurs at joints between two fibers (splices, plugs). From a wave point of view every scattering center can be taken as a reflecting point. With an Optical-Time-Domain-Reflectometer (OTDR) this effect can be used to evaluate a fiber connection. The basis of this instrument is that a narrow pulse is launched into one end of the fiber and then the amount of reemerging light from this same end is monitored as a function of time after the initial launch. As the pulse travels down the fiber it will be attenuated and partially reflected. Taking the velocity of light inside the fiber as $\frac{c}{\eta_1}$the reemerging light at a certain time t after launching at t_0 is closely related to the distance $L = \frac{c}{2\eta_1(t-t_0)}$ where the light is reflected with $R\{L\}$. With constant attenuation the portion

$$R_\mathrm{i}\{t\} = R\{L\} \cdot \exp\{-2\alpha L\} \tag{3.92}$$

is found at the input and from the graph $R_\mathrm{i}\{t\}$ one sees at which conditions the connection is working.

3.2 Effective-Index Method

As we have seen before only in special geometries a closed solution for the wave-equation can be found. If we take a separation-trial by products assuming propagation in $z-$direction for example in cartesian coordinates such a trial is

$$\vec{A}(\vec{r}) = X\{k_\mathrm{x}\, x\} \cdot Y\{k_\mathrm{y}\, y\} \cdot \exp\{i(k_\mathrm{z}\, z - \omega t)\} \cdot \vec{e}_\mathrm{z} \tag{3.93}$$

for pure TM-waves. The dispersion relation now in every layer is

$$k_\mathrm{x}^2 + k_\mathrm{y}^2 + k_\mathrm{z}^2 = k^2 \quad . \tag{3.94}$$

Let us take the waveguide in figure 3.11 as an example.

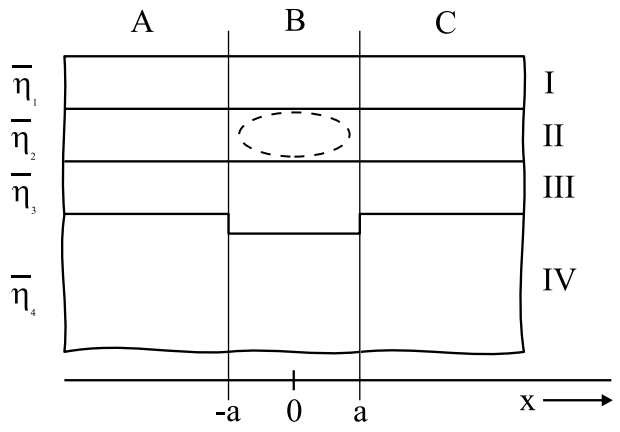

Figure 3.11: Ridge guided waveguide. Light is guided in layer II above the ridge. The refractive indices follow $\eta_2 > \eta_1$ and $\eta_2 > \eta_3 > \eta_4$.

If the spatial dependence in $x-$direction is small compared to the $y-$direction we can set as a first attempt $k_\mathrm{x} = 0$ and $X = 1$ finding the dispersion relation

$$k_\mathrm{y}^2 + k_{\mathrm{z},0}^2 = k^2 \quad . \tag{3.95}$$

We may solve the wave equation for the three regions A, B, and C separately assuming they are homogeneous in $x-$direction. In every region from boundary condition we find a $k_{\mathrm{z},0}$ as solution which is equivalent to an effective index $\eta_{\mathrm{eff}} = \frac{k_{\mathrm{z},0}}{k_0}$. It is the refractive index of a homogeneous material where the light would travel with the same velocity as in the structure. Due to the differences in the three regions A,B and C the effective refractive indices are different. With respect to the $x-$direction the structure in figure 3.11 now can be viewed as a structure of three homogeneous layers with indices $\eta_{\mathrm{eff,A}}$, $\eta_{\mathrm{eff,B}}$ and $\eta_{\mathrm{eff,C}}$ without any change in $y-$direction. Rewriting (3.94) we find with (3.95)

$$k_{\mathrm{x}}^2 + k_{\mathrm{z}}^2 = k^2 - k_{\mathrm{y}}^2 = k_{\mathrm{z},0}^2 = \eta_{\mathrm{eff}}^2 \cdot k_0^2 \quad . \tag{3.96}$$

If in the first step a TM-wave with respect to the $y-z-$ plane has been considered the $H-$field is oriented into $x-$direction. As a consequence in the second step we have to take a TE-wave because there is a strong $H-$field perpendicular to the interfaces of the three layer structure.

The general method is to split a problem into two independent ones neglecting one of them in a first calculation step. The resulting propagation constants are attributed to homogeneous layers of effective refractive index η_{eff}. In the second step the neglected problem is solved independently replacing all structures from the first step by homogeneous layers with different effective indices. The usage of effective indices to describe the wave propagation in a complicated structure is the origin of the method name.

3.3 Beam Propagation Method

In the last section a method for the eigenmode calculation in planar step-index waveguides has been introduced. If the refractive index is varying continuously the structure can be split into a lot of parallel layers with small thickness such that the index in each layer is nearly constant as sketched in figure 3.12.

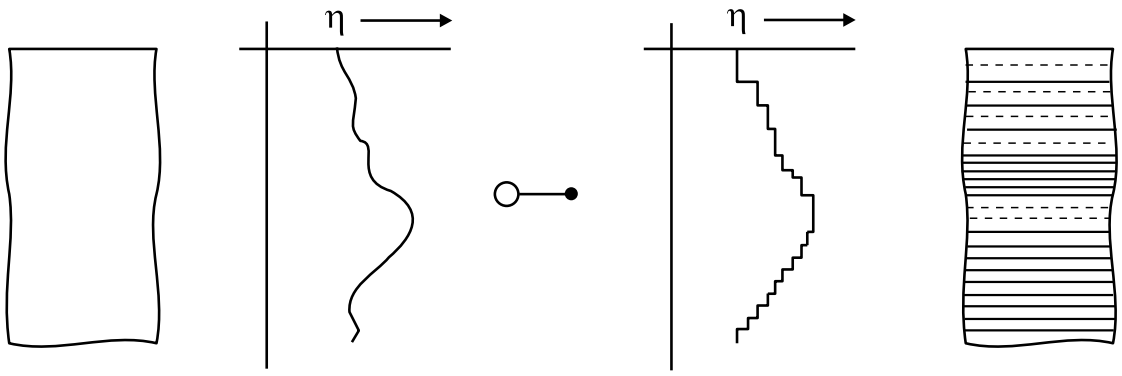

Figure 3.12: A waveguide with continuous varying refractive index is split into many homogeneous layers with constant index. For numerical reasons more layers than necessary are introduced (dashed lines) increasing the effort to solve the problem.

This method leads to a big number of layers and therefore high numerical effort to solve the problem. Moreover the thicknesses of the layers are varying which leads to additional effort in the programming of a computer. For this reason additional layers are introduced such that a grid with constant period is build increasing the numerical effort again. For waveguide splits or junctions as sketched in figure 3.13 the cross section of the waveguides is $z-$dependent. An additional grid has to be implemented in $z-$direction and the modes in each section have to be calculated increasing the effort again.

Often it is not necessary to know the propagation coefficient k_z. Only the field distribution is desired for a given input. In this case the beam propagation method is widely used. This method goes back to the knowledge that each field distribution can be expanded into plane waves by Fourier transformation. Both the plane waves as

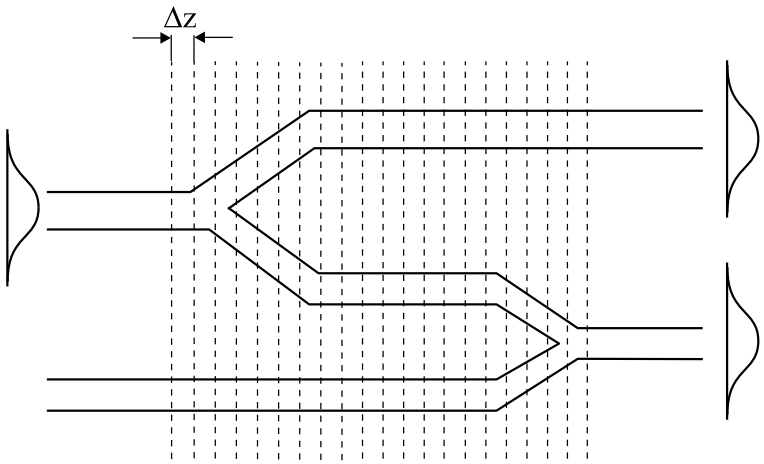

Figure 3.13: Waveguide splits, joints and bends in an integrated optical cross connect. For the calculation of light propagation the structure is split into sections. Each section contains a nearly $z-$independent cross section.

well as the real distribution are propagating in $z-$direction over distance $\Delta z = z - z_0$ with $\exp\{i\, k_\mathrm{z} \cdot \Delta z\}$. The trick is to split the refractive index into a constant and a cross-section dependent portion

$$\eta\{x, y\} = \eta_0 + \Delta\eta\{x, y\} \quad . \tag{3.97}$$

If $\Delta\eta \ll \eta_0$ the propagation coefficient k_z results from the dispersion relation to

$$\begin{aligned} k_\mathrm{z} &= \sqrt{\eta^2 k_0^2 - k_\mathrm{x}^2 - k_\mathrm{y}^2} = \sqrt{\eta_0^2 k_0^2 - k_\mathrm{x}^2 - k_\mathrm{y}^2 + 2\Delta\eta\, k_0 \left(\eta_0 + \frac{\Delta\eta}{2}\right) k_0} \\ &\simeq \sqrt{\eta_0^2 k_0^2 - k_\mathrm{x}^2 - k_\mathrm{y}^2} \cdot \left(1 + \Delta\eta\, k_0 \cdot \frac{\left(\eta_0 + \frac{\Delta\eta}{2}\right) k_0}{\eta_0^2 k_0^2 - k_\mathrm{x}^2 - k_\mathrm{y}^2}\right) \\ &\simeq k_{\mathrm{z}0} + \Delta\eta\, k_0 = k_{\mathrm{z}0} + \Delta\eta\, \frac{2\pi}{\lambda} \end{aligned} \tag{3.98}$$

with

$$k_{\mathrm{z}0} = \sqrt{\eta_0^2 k_0^2 - k_\mathrm{x}^2 - k_\mathrm{y}^2} \quad . \tag{3.99}$$

With (3.98) the propagation can be split into two factors $\exp\{i\,k_z \cdot \Delta z\} \simeq \exp\{i\,k_{z0}\Delta z\} \cdot \exp\left\{i\,2\pi\Delta\eta \cdot \frac{\Delta z}{\lambda}\right\}$.

Now the field propagation is calculated recursively in four steps

$$\tilde{E}_{\mathrm{in}}\{m\} = \mathrm{F}\{E_{\mathrm{in}}\{m\}\} \tag{3.100}$$

$$\tilde{E}_{\mathrm{k}} = \tilde{E}_{\mathrm{in}}\{m\} \cdot \exp\{i\,k_{z0}\Delta z\} \tag{3.101}$$

$$E_{\mathrm{k}} = \mathrm{F}^{-1}\{\tilde{E}_{\mathrm{k}}\} \tag{3.102}$$

$$E_{\mathrm{out}}\{m\} = E_{\mathrm{k}} \cdot \exp\left\{i\,2\pi\Delta\eta\{x,y\} \cdot \frac{\Delta z}{\lambda}\right\} \tag{3.103}$$

where for each section m is set to $E_{\mathrm{in}}\{m\} = E_{\mathrm{out}}\{m-1\}$. The Fourier transformation and its inverse follow

$$\mathrm{F}\{E\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\{x,y\} \cdot \exp\{-i\,k_{\mathrm{x}}x\} \cdot \exp\{-i\,k_{\mathrm{y}}y\}\ \mathrm{d}x\ \mathrm{d}y \tag{3.104}$$

$$\mathrm{F}^{-1}\{\tilde{E}\} = \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{E}\{k_{\mathrm{x}},k_{\mathrm{y}}\} \cdot \exp\{i\,k_{\mathrm{x}}x\} \cdot \exp\{i\,k_{\mathrm{y}}y\}\ \mathrm{d}k_{\mathrm{x}}\ \mathrm{d}k_{\mathrm{y}} \quad . \tag{3.105}$$

Often this method is described as operations in Fourier- and real space. With (3.100) one changes from real space to the Fourier space. In the Fourier space the plane waves are propagating with $\exp\{i\,k_{z0} \cdot \Delta z\}$ (3.101). After changing back to the real space (3.102) the missing cross section dependent propagation (3.103) follows. This roundtrip is performed for each section Δz. The assumptions made in (3.98) are only valid for small k_{x} and k_{y} compared to $k = \eta_0 k_0$. This means that in (3.104) only for small values of k_{x} and k_{y} remarkable values for $\tilde{E}\{k_{\mathrm{x}},k_{\mathrm{y}}\}$ should result which is true for paraxial beams.

Only they are considered and their propagation is split into a constant part in Fourier space and a cross-section depending part in real space as schematically sketched in figure 3.14. Due to the fact that propagation of paraxial beams is performed the method is named **B**eam **P**ropagation **M**ethod or short BPM.

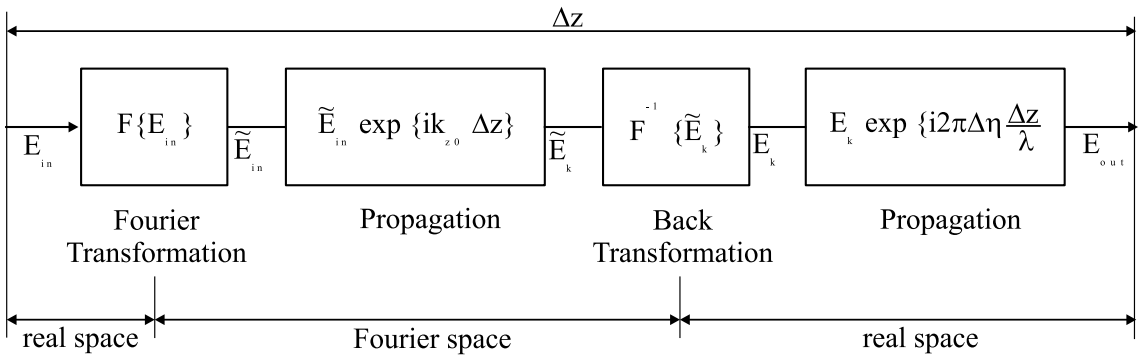

Figure 3.14: Schematical description of the beam propagation method.

3.4 Properties of Optical Modes

In an optical system waveguides are connected by splices and connectors or via splitters, Moreover such a system can contain couplers and filters. For the description of this elements it is useful to employ optical modes. The general properties of guided optical modes provide insight into the function of that elements. Both discontinuities and perturbations can be treated with optical modes.

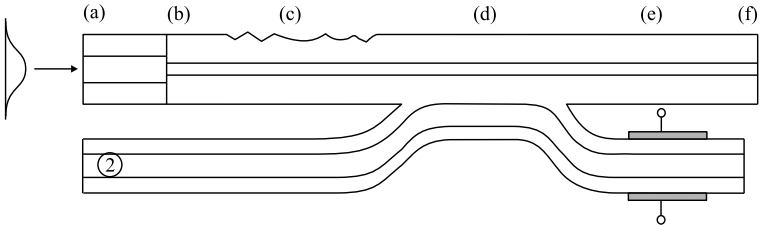

Figure 3.15: Perturbations of a waveguide. The light propagation behaviour can be described by the modes of non perturbed waveguides.

The perturbation can be accounted by an additional polarization which interacts with the optical modes. As an example we may use the situation sketched in figure 3.15 where the exitation of optical modes from incident light at position (a) and the butt coupling of waveguides at position (b) leads to mode exitation respectively so called mode coupling at that discontinuities. More complicated is the perturbation by a

surface corrugation as is sketched at position (c) or the coupling between optical modes in different waveguides (d). At least modes can be influenced by external fields which lead to optical anisotropy (e). The radiation into free space (f) can be described as a mode coupling to paraxial waves like the butt coupling (b). Paraxial waves can be regarded as the modes of free space under certain assumptions. The properties of guided modes i.e. orthogonality and reciprocity are key elements to evaluate the system.

Modes of order m [1] have the propagation constant k_{zm}. In the following chapters we use the field representations

$$\begin{aligned} \vec{E} &= \vec{E}_m\{x,y\} \cdot \exp\{i(k_{\mathrm{z}m}\, z - \omega t)\} \\ \vec{H} &= \vec{H}_m\{x,y\} \cdot \exp\{i(k_{\mathrm{z}m}\, z - \omega t)\} \end{aligned} \tag{3.106}$$

with expansions into longitudinal and transversal components

$$\begin{aligned} \vec{E}_m &= (\vec{E}_{\mathrm{t}m} + E_{\mathrm{z}m} \cdot \vec{e}_{\mathrm{z}}) \\ \vec{H}_m &= (\vec{H}_{\mathrm{t}m} + H_{\mathrm{z}m} \cdot \vec{e}_{\mathrm{z}}) \quad . \end{aligned}$$

The expansions are related by (see appendix C)

$$\begin{aligned} \vec{E}_{\mathrm{t}m} &= \vec{E}^*_{tm} \\ \vec{H}_{\mathrm{t}m} &= \vec{H}^*_{\mathrm{t}m} \\ E_{\mathrm{z}m} &= -E^*_{\mathrm{z}m} \\ H_{\mathrm{z}m} &= -H^*_{\mathrm{z}m} \quad . \end{aligned} \tag{3.107}$$

[1] Normally there are two numbers m, n that are assigned to a mode. For simplicity in this section we put them together into one modenumber m which can be regarded as the number of an element in a series.

This tells us that in traveling modes transversal electric and magnetic fields $\vec{E}_\mathrm{t}$ and $\vec{H}_\mathrm{t}$ are in phase but exhibit a $\frac{\pi}{2}$-phase-shift for both longitudinal fields E_z and H_z.
Time averaged energy transport can be calculated for a mode of order m with the complex pointing-vector $\vec{S}_m = \vec{E}_m \times \vec{H}_m^* = \vec{E}_{\mathrm{t}m} \times \vec{H}_{\mathrm{t}m}^* + E_{\mathrm{z}m}(\vec{e}_\mathrm{z} \times \vec{H}_{\mathrm{t}m}^* + H_{\mathrm{z}m}(\vec{e}_\mathrm{z} \times \vec{E}_{\mathrm{t}m}^*$ giving the radiated power

$$P_m = \frac{1}{2}\mathrm{Re}\left\{\iiint_V \nabla \circ \vec{S}_m \,\mathrm{d}^3 r\right\} = \frac{1}{2}\mathrm{Re}\left\{\oiint_{S_V} \vec{S}_m \circ \,\mathrm{d}^2\vec{S}\right\} \quad . \tag{3.108}$$

As long as guided modes are regarded the fields vanish transversely and energy transport folows directly from the z- component of $\vec{S}_m$ resulting with integration over the transversal plane S_t into

$$P_m = \frac{1}{2}\Re\iint_{S_\mathrm{t}} \vec{S}_m \circ \vec{e}_\mathrm{z} \,\mathrm{d}^2 r = \Re\langle S_{\mathrm{z}m}\rangle \quad . \tag{3.109}$$

In non dispersive media P_m is directly related to the energy densities of electric and magnetic fields

$$w_m = \frac{1}{2}\left(\varepsilon\varepsilon_0 \vec{E}_m \circ \vec{E}_m^* + \mu\mu_0 \vec{H}_m \circ \vec{H}_m^*\right)$$

by the group velocity

$$c_{\mathrm{gr}m} = \frac{\mathrm{d}}{\mathrm{d}\omega}\mathrm{Re}\left\{k_{\mathrm{z}m}\right\} = \frac{\mathrm{d}}{\mathrm{d}\omega}\beta_m$$

giving

$$P_m = c_{\mathrm{gr}m} \iint_{S_\mathrm{t}} w_m \;\mathrm{d}^2 r \tag{3.110}$$

implying the photon picture with photons of energy $\hbar\omega$ and density $N_{\mathrm{phot}m} = \frac{1}{\hbar\omega}\, w_m$ which gives

$$P_m = \hbar\omega \cdot c_{\mathrm{gr}m} \iint\limits_{S_\mathrm{t}} N_{\mathrm{phot}m}\, \mathrm{d}^2 r \tag{3.111}$$

where the integral is a measure of photons that flow through the area S_t transporting energy with the group-velocity.

3.4.1 Reciprocity

For the evaluation of perturbations in a given structure it would be nice to describe the effects by the known fields (modes) of the unperturbed structure. As a first step the relation between two arbitrary choosen field distributions (modes) that satisfy the Maxwell equations is given by (see appendix C.1)

$$\nabla \circ (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2^* \times \vec{H}_1) = 0 \quad . \tag{3.112}$$

This means that energy transport is the same for both combinations $\vec{E}$ and $\vec{H}$. As a model a setup with two antennas can be taken. When they are acting as sources with the same input power they produce the field distributions 1 and 2 respectively. Equation (3.112) says that the energy transport from 1 to 2 is the same as from 2 to 1. In a perturbed structure the field can be expressed as the field of the unperturbed structure together with a polarization that accounts for the field change caused by the perturbation. Assuming the perturbed field as $\vec{D}_1 = \varepsilon_0 \varepsilon \vec{E}_1 + \vec{P}_1$ (3.112) is modified to

$$\nabla \circ (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2^* \times \vec{H}_1) = i\omega(\vec{P}_1 \circ \vec{E}_2^*) \quad . \tag{3.113}$$

Now the energy transport from 1 to 2 does not equal the way back due to the perturbation. Equation (3.113) is the key to calculate a field distribution $\vec{E}_2, \vec{H}_2$ that carrys

just the same energy as is generated by the perturbation. Instead of (3.113) it is more convenient to use its integral representation

$$\oiint_{s_\mathrm{V}} (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2^* \times \vec{H}_1) \circ \mathrm{d}^2\vec{S} = i\omega \iiint_V (\vec{P}_1 \circ \vec{E}_2^*)\, \mathrm{d}^3 r \quad . \tag{3.114}$$

3.4.2 Orthogonality, Normalization

It is not very convenient to calculate with fields $\vec{E}$ and $\vec{H}$ in (3.114). The modal behaviour is well known for the structure and therefore it would be nice when the fields could be expanded into a series of mode fields. For this purpose the modes must be orthogonal and build a complete set. Furthermore the amplitudes should be normalized for the ease of calculation. Orthogonality of two different modes m and p that are not degenerate ($k_{zm} \neq k_{zp}$) is ahown in appendix C.2 as

$$\iint_{S_\mathrm{t}} (\vec{E}_{\mathrm{t}p} \times \vec{H}_{\mathrm{t}m}^*)\, \mathrm{d}^2 r = 0$$

where S_t is the whole transverse plane. The energy transport of mode m follows directly from (3.109) with (3.107) to

$$\langle S_{zm} \rangle = \iint_{S_\mathrm{t}} (\vec{E}_{\mathrm{t}m} \times \vec{H}_{\mathrm{t}m}^*) \circ \vec{e}_\mathrm{z}\, \mathrm{d}^2 r \quad . \tag{3.115}$$

Normalization is carried out by

$$\begin{aligned} \vec{\tilde{E}}_m &= \frac{\vec{E}_m}{\sqrt{\langle S_{\mathrm{z},m} \rangle}} \\ \vec{\tilde{H}}_m &= \frac{\vec{H}_m}{\sqrt{\langle S_{\mathrm{z},m} \rangle}} \quad . \end{aligned} \tag{3.116}$$

With the normalized mode fields the orthogonal relation

$$\iint_{S_\mathrm{t}} (\vec{\bar{E}}_{\mathrm{t}m} \times \vec{\bar{H}}^*_{\mathrm{t}p}) \circ \vec{e}_\mathrm{z}\, \mathrm{d}^2 r = \delta_{m,p} \quad . \tag{3.117}$$

follows and transverse fields are easily expanded into a series of the normalized mode-fields given above.

3.4.3 Field Expansions by Modes

With normalization of the orthogonal mode fields we can try to expand any given field $\vec{E}$ or $\vec{H}$ into a series with $m \geq 0$ regarding $\vec{E}_{\mathrm{t}m} = \vec{E}_{\mathrm{t},-m}$ and $\vec{H}_{\mathrm{t}m} = -\vec{H}_{\mathrm{t},-m}$

$$\begin{aligned} \vec{E}_\mathrm{t} &= \sum_{m=1}^{\infty} (a_m + a_{-m}) \vec{\bar{E}}_{\mathrm{t}m} \\ \vec{H}_\mathrm{t} &= \sum_{m=1}^{\infty} (a_m - a_{-m}) \vec{\bar{H}}_{\mathrm{t}m} \quad . \end{aligned} \tag{3.118}$$

The z- components of $\vec{E}$ and $\vec{H}$ follow from the Maxwell equations (see C.2) and the mode amplitudes a_m can be calculated using (3.115), (3.116) and (3.117)

$$a_{\pm m} = \frac{1}{2} \iint_{S_\mathrm{t}} (\vec{E} \times \vec{\bar{H}}^*_{\mathrm{t}m} \pm \vec{\bar{E}}^*_{\mathrm{t}m} \times \vec{H}) \circ \vec{e}_\mathrm{z}\ \mathrm{d}^2 r \quad . \tag{3.119}$$

This expansion holds only for nondegenerate modes. In the case of degenerate modes it can be shown that they are as well orthogonal but not to the simple relation (3.115) and therefore the extraction of amplitudes is somewhat more complicated. Energy transport into z- direction follows for guided modes from

$$\langle S_\mathrm{z} \rangle = \iint_{S_\mathrm{t}} (\vec{E}_\mathrm{t} \times \vec{H}^*_\mathrm{t}) \circ \vec{e}_\mathrm{z}\, \mathrm{d}^2 r$$

$$= \iint\limits_{S_\mathrm{t}} \sum_{m=1}^{\infty}\sum_{p=1}^{\infty}((a_m + a_{-m})\vec{\underline{E}}_\mathrm{t} \times (a_p^* - a_{-p^*})\vec{\underline{H}}_\mathrm{t}^{\,*}) \circ \vec{e}_\mathrm{z}\,\mathrm{d}^2r$$

$$= \sum_{m=1}^{\infty}\sum_{p=1}^{\infty}(a_m + a_{-m})(a_p^* - a_{-p}^*)\delta_{m,p} = \sum_{m=1}^{\infty} |a_m|^2 - |a_{-m}|^2$$

which is the difference between forward and backward transported energy.

As discussed in appendix C.2 it has to be noted that (3.118) is only applicable to arbitrary transversal fields when the modes are a full set. This is fulfilled in the most practical cases and must be shown for each structure under consideration.

For the calculation of a_m in (3.119) $\vec{E}_\mathrm{t} = E_\mathrm{t}\vec{e}_\mathrm{t}$ and $\vec{H}_\mathrm{t} = H_\mathrm{t}\vec{e}_\mathrm{t}$ can be set. The propagation folows $\vec{k} = k_\mathrm{z}\vec{e}_\mathrm{z} + k_\ell\vec{e}_\ell$ where

$vece_\ell$ is defined as

$vece_\ell = \vec{e}_\mathrm{z} \times \vec{e}_\mathrm{t}$ and k can be taken as $k \simeq k_\mathrm{z} \gg k_\ell$. In the most practical waveguides the modes are nearly TE or TM polarized. In this case the electric and magnetic fields are related by

$$\vec{H} \simeq \frac{1}{kZ} \cdot (\vec{k} \times \vec{E}) = \frac{1}{kZ}\left((k_\ell E_\mathrm{z} - k_\mathrm{z}E_\ell)\vec{e}_\mathrm{t} + k_\mathrm{z}E_\mathrm{t}\vec{e}_\ell - k_\ell E_\mathrm{t}\vec{e}_\mathrm{z}\right)$$

with $Z = \sqrt{\frac{\mu\mu_0}{\varepsilon\varepsilon_0}}$. Equation (3.119) simplifies with $(\vec{E}_{tm}^* \times \vec{H}) \circ \vec{e}_\mathrm{z} = \frac{k_\mathrm{z}}{kZ}E_{tm}^* E_\mathrm{t} = \frac{k_\mathrm{z}}{kZ}\vec{E} \circ \vec{E}_{tm}^*$ to

$$a_m \simeq \frac{k_{zm}}{kZ} \iint\limits_{S_\mathrm{t}} \vec{E} \circ \vec{\underline{E}}_{tm}^{\,*}\ \mathrm{d}^2r \simeq \frac{1}{Z} \iint\limits_{S_\mathrm{t}} \vec{E} \circ \vec{\underline{E}}_{tm}^{\,*}\ \mathrm{d}^2r \quad . \tag{3.120}$$

The z- component of the space averaged complex pointing vector follows with $\vec{E}_{tm} = \vec{e}_\mathrm{z} \times \vec{E}_m \times \vec{e}_\mathrm{z}$ from

$$\langle S_{zm}\rangle \simeq \frac{1}{Z} \iint\limits_{S_\mathrm{t}} \left|\vec{E}_m \times \vec{e}_\mathrm{z}\right|^2\ \mathrm{d}^2r \tag{3.121}$$

Equation (3.120) holds for planar and circular waveguides and it must be shown for other types. Nevertheless (3.120) is extremely useful and widely used in literature instead of (3.119).

A typical application of (3.120) is the calculation of coupling efficiency from a light source into a fiber. We assume the amplitude of the incident electrical field to be $\vec{E}_\mathrm{i}$. The coupling into a mode m is calculated by the amount of power coupled into mode m which is proportional to $|a_m|^2$ relative to the input power $\langle S_{zi} \rangle$ giving the coupling efficiency

$$\eta_m = \frac{\iint\limits_{S_\mathrm{t}} \left(\vec{E}_\mathrm{i} \circ \vec{\bar{E}}^*_{\mathrm{t}m} \right)^2 \, \mathrm{d}^2 r}{\iint\limits_{S_\mathrm{t}} |\vec{E}_\mathrm{i} \times \vec{e}_\mathrm{z}|^2 \, \mathrm{d}^2 r} \quad . \tag{3.122}$$

3.5 Waveguide Resonators

We want to evaluate the amplitude dependence of a mode m along the waveguide. Starting from pure transverse fields E_1 and H_1 the exitation through a mode m with fields E_2 and H_2 can be found going back to (3.112) and taking $\nabla = \nabla_\mathrm{t} + \frac{\partial}{\partial z} \vec{e}_\mathrm{z}$. We expand the transverse fields $\vec{E}_1$ and $\vec{H}_1$ as in (3.118) to $\vec{E}_1 = \sum_m (a_m + a_{-m}) \vec{\bar{E}}_{\mathrm{t},m}$ $\vec{H}_1 = \sum_m (a_m - a_{-m}) \vec{\bar{H}}_{\mathrm{t},m}$. Integration over the whole transverse plane gives with Stokes law

$$\iint\limits_{S_\mathrm{t}} \nabla_\mathrm{t} \circ \vec{b} \, \mathrm{d}^2 r = \oint\limits_{C_{S_\mathrm{t}}} \vec{b} \circ \mathrm{d}\vec{\ell} \tag{3.123}$$

for (3.112)

$$\iint\limits_{S_\mathrm{t}} \frac{\partial}{\partial z} \left(\vec{E}_1 \times \vec{H}_2^* + \vec{E}_2 \times \vec{H}_1^* \right) \circ \vec{e}_\mathrm{z} \, \mathrm{d}^2 r = 0 \tag{3.124}$$

because the integral over the transverse part vanishes due to vanishing fields $\vec{E}_1$ and $\vec{H}_1$ at infinite distances to the $z-$axis. Assuming further $\vec{E}_2 =$

$\left(\vec{E}_{\mathrm{t},m} + E_{\mathrm{z},m} \cdot \vec{e}_{\mathrm{z}}\right) \exp\{i(k_{\mathrm{z},m}\, z - \omega t)\}$ and $\vec{H}_2 = \left(\vec{H}_{\mathrm{t},m} + H_{\mathrm{z},m} \cdot \vec{e}_{\mathrm{z}}\right) \exp\{i(k_{\mathrm{z},m}\, z - \omega t)\}$ as in (3.106) condition (3.124) can only be met with a z- dependent a_m

$$\frac{1}{a_m} \frac{\mathrm{d}}{\mathrm{d}z}\, a_m - i\, k_{\mathrm{z},m} = 0 \tag{3.125}$$

$$\frac{1}{a_{-m}} \frac{\mathrm{d}}{\mathrm{d}z}\, a_{-m} + i\, k_{\mathrm{z},-m} = 0 \tag{3.126}$$

where the solution is

$$a_{\pm m}\{z\} = a_{\pm m,0} \cdot \exp\{\pm i\, k_{\mathrm{z},m}\, z\} \quad . \tag{3.127}$$

This means that for a given mode of order m the propagation along the $z-$axis is like that of a plane wave in free space with refractive index $\eta_{\mathrm{eff},m} = \frac{k_{\mathrm{z},m}}{k_0}$. With this result waveguide resonators can be treated in the same way as the equivalent resonators in free space.

Chapter 4

Optical Mode Coupling and Radiation

4.1 Coupling of Optical Modes

The optical modes of a given waveguide are independent from each other which means that they can have arbitrary amplitudes. When the waveguide is contains discontinuities or is perturbed the modes can be coupled. We take figure 4.1 as an example as we did already in section 3.4. The simplest discontinuity is the end of a waveguide (a), (b), (f). More complicated are bends or corrugations of surface and ingots (c). While the end of a waveguide can be handled by simply adding the entire modes to fulfill boundary conditions the other ones must be treated by coupled mode theory. The mode exitation at (a) has already been discussed in section 3.4.3. Butt coupling (b) will be evaluated in 4.1.1. The influence of a corrugation is described in section 4.1.2 and disturbance by closely neighboring waveguides is subject of section 4.1.4. In both cases a differential equation system for the z- dependence of the mode amplitudes results. The general solution is given in section 4.1.3. Field induced optical anisotropy (e) is discussed in section 4.1.5. Radiation into free space is subject of section 4.2. Here we look for a solution of the wave equation under assumption of paraxial wave

propagation. As a result we find Hermite-Gauss and Laguerre-Gauss waves as modes of free space. Mode exitation can then be dealed as for butt coupling (b).

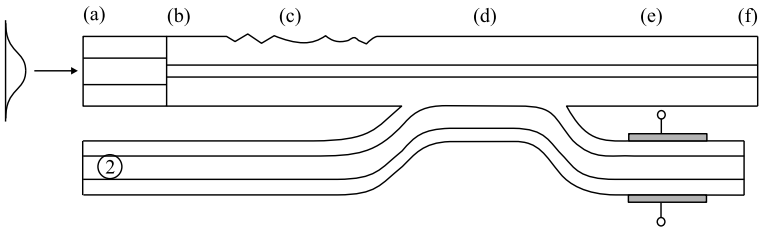

Figure 4.1: Perturbations of a waveguide. The light propagation behaviour can be described by the modes of non perturbed waveguides.

4.1.1 Joint Mode-Coupling

We assume that two waveguides are joint at a certain point z as sketched in figure 4.2.

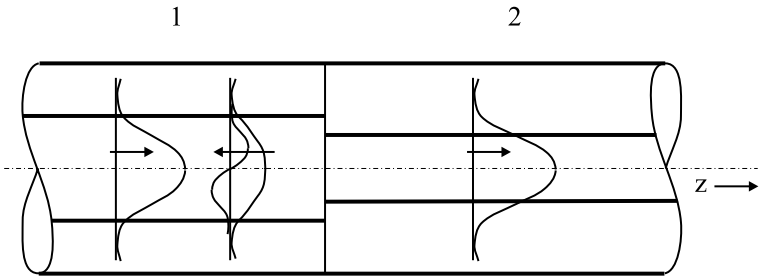

Figure 4.2: Joint of two waveguides. The incident mode in the left waveguide generates modes in both waveguides that travel away from the joint.

A mode of order m is incident from the left waveguide. Boundary conditions say that at charge- and current- free interfaces the transverse electrical and magnetical fields and the normal magnetic induction and electric displacement must be steady. This leads for plane waves to the well known Fresnell-refraction-laws. But in the case of modes we have additionally to meet the lateral field distributions. As we have seen before we

can expand any field distribution at a given point z into the modes of the waveguide by usage of (3.119) or (3.120) if we handle weakly guided modes. Imagine the lowest order mode of the left-hand waveguide to be incident to the joint as depicted in figure 4.2. At the joint it is reflected partly due to index changes. The reflection follows mainly the Fresnell-laws of reflection. The ground state mode in the right waveguide is exited. If both waveguides have the same modes and are well aligned this is the whole story. But in general the fibers may have an offset in the axes or are tilted or even worse have different modes. All this cases must be treated by assuming the whole entity of modes in both waveguides to be exited at the joint and traveling away from it.
With incident transverse field amplitude $\vec{E}_{\mathrm{t},i}$ we have to set

$$\vec{E}_{\mathrm{t},i} + \sum_m b_m \cdot \vec{\vec{E}}_{\mathrm{t},m} = \sum_n a_n \cdot \vec{\vec{E}}_{\mathrm{t},n} \tag{4.1}$$

where m denotes the modes of the incident waveguide and n the modes of the other one. We make use of (3.120) and find

$$a_n = \frac{2k_{\mathrm{z},n}}{kZ} \iint\limits_{S_\mathrm{t}} \left[\left(\vec{E}_{\mathrm{t},i} + \sum_m b_m \cdot \vec{\vec{E}}_{\mathrm{t},m} \right) \circ \vec{\vec{E}}_{\mathrm{t},n} \right] \, \mathrm{d}^2 r \quad . \tag{4.2}$$

For $m = i$ we have Fressnell reflection $\vec{E}_{\mathrm{t},m} = r_i \cdot \vec{E}_{\mathrm{t},i}$. Equation (4.2) has to be solved for every n and we can formally note

$$\vec{a}_n = [M_{n,m}] \cdot \vec{b}_m + \vec{g}_{i,n} \tag{4.3}$$

where the elements of $\vec{a}_n$ and $\vec{b}_m$ are the amplitudes of exited modes. $[M_{n,m}]$ is a coupling matrix with elements

$$M_{n,m} = \frac{2k_{\mathrm{z},n}}{kZ} \iint\limits_{S_\mathrm{t}} (\vec{\vec{E}}_{\mathrm{t},m} \circ \vec{\vec{E}}_{\mathrm{t},n}) \, \mathrm{d}^2 r \tag{4.4}$$

which are well known from quantum theory as transition matrix elements between two states n and m. The elements of $\vec{g}_{i,n}$ are exitation values

$$g_{i,n} = \frac{2k_{\mathrm{z},n}}{kZ} \iint\limits_{S_\mathrm{t}} (\vec{E}_{\mathrm{t},i} \circ \vec{\tilde{E}}_{\mathrm{t},n}) \ \mathrm{d}^2 r \quad . \tag{4.5}$$

Unfortunately in (4.1) both $\vec{a}_n$ and $\vec{b}_m$ are unknown and there is no unique solution. But we have to remember that on charge free interfaces the normal dielectric displacement must be steady which gives an additional equation to an unique solution of (4.3). This additional equation is a matrix equation like (4.3) with the same amplitude vectors $\vec{a}_n$ and $\vec{b}_m$ but another matrix and another source vector making the solution somewhat complicated but solvable.

A way to omit the second matrix equation is the recursive calculation of an approximate solution for (4.3). In the first step $\vec{b}_m$ is set to zero and $\vec{a}_n$ is directly calculated from $g_{i,n}$. Now $g_{i,n}$ is set to zero and a first approximation for $\vec{b}_m$ is calculated with the inverse matrix $[M_{n,m}]$ by

$$\vec{b}_m = [M_{n,m}]^{-1} \cdot \vec{a}_n \quad . \tag{4.6}$$

Now $\vec{b}_m$ is normalized to meet $b_m \cdot \vec{\tilde{E}}_{\mathrm{t},m} = r_i \cdot \vec{E}_{\mathrm{z},i}$ for $m = i$ which is done by

$$\vec{b}'_m = \frac{r_i}{b_i} \cdot \vec{b}_m \quad . \tag{4.7}$$

Now $\vec{b}'_m$ form (4.7) is taken as an approximation for $\vec{b}_m$ in (4.3) and a new $\vec{a}_n$ is calculated. Steps (4.6) and (4.7) follow and so on until the changes in $\vec{a}_n$ and $\vec{b}_m$ are below a preset limit at which the solution is assumed to be steady state. This is computer work and can be done very quick. The biggest afford is to calculate the matrix $[M_{n,m}]$.

The problem is to figure out which is the best approximation for r_i, especially when the second waveguide supports modes of other dimensions. In this case the way out is to

look at $\vec{g}_{i,n}$ and find out, which elements is the biggest. That element is considered to represent the mode which mainly contributes to the energy transport. The associated wavevector $\vec{k}_{i,n}$ can be taken to calculate the value of r_i using (2.18).

4.1.2 Mode Coupling at Perturbations

In this section we evaluate the influence of a perturbation on a waveguide. We assume an incident field $\vec{E}_1$ that is perturbed by a corrugation of the waveguide surface which can be a grating for example. The following calculations are provided under the assumption that the interaction between corrugation and field is small, which is found when only the evanescent field interacts with the corrugation. In this case we speak of weak coupling. For strong coupling the method presented here is not applicable. At the perturbation new modes are generated traveling back and forth. The effect that an incident wave mode generates new modes at a perturbation is called mode coupling. Figure 4.3 shows schematically the procedure how to calculate the new modes.

For weak mode coupling the incident field is first expanded into a series of modes (4.9). Each mode generates a perturbation polarization (4.14) which in turn generates new modes. The generated modes are calculated by expansion of the polarization into modes of the non perturbed waveguide (4.11, 4.13). When the perturbation is z-dependent as we find for a grating the mode coupling must in principle be calculatet for each positoin z. The calculation can be significantly simplified a Fourier- transform of the perturbation polarization (4.17, 4.20) is introduced. With this step boundary conditions can be fulfilled when the Fourier transformed polarization is expanded into the fields of the modes. As a result the exitation of mode m by incident mode n follows as a differential equation. With some mathematics called coupled mode theory the exitation strength can be calculated.

For evaluation of the influence of perturbations we use equation (C.12) as starting point. The incident field E_1 produces a polarization P_1 at the perturbation which in turn excites the field E_2 as output. Considering (C.16) we can simplify to

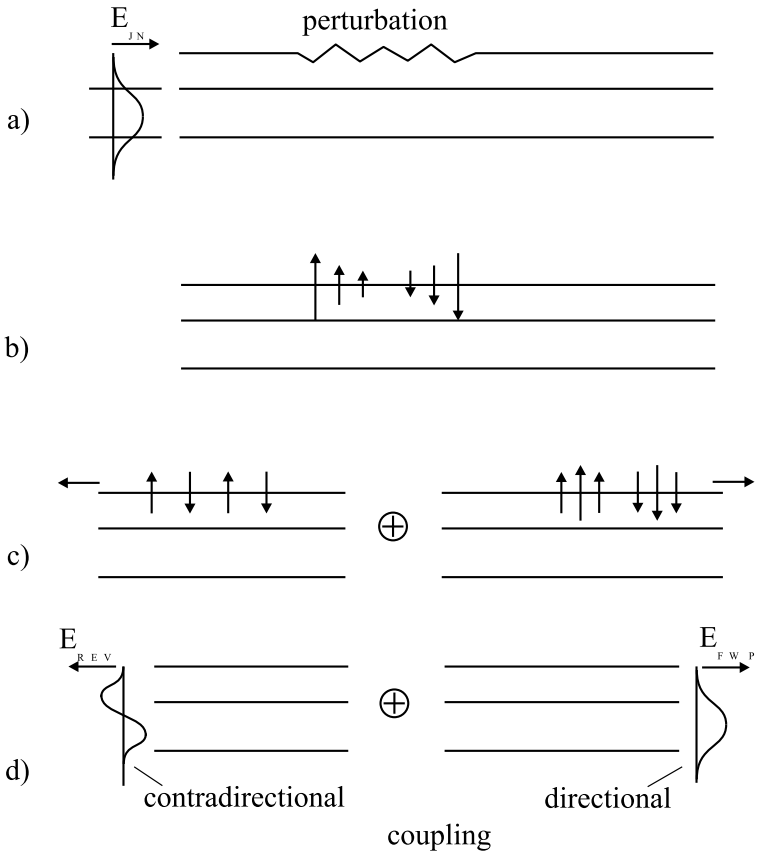

Figure 4.3: An incident wave E_{IN} produces the perturbation polarization P. Via Fourier transformation and mode expansion P itself generates forward and backward traveling waves in the waveguide.

$$\iint\limits_{S_\mathrm{t}} \frac{\partial}{\partial z}\left(\vec{E}_1 \times \vec{H}_2^* + \vec{E}_2^* \times \vec{H}_1\right) \circ \vec{e}_\mathrm{z}\,\mathrm{d}^2r = i\,\omega \iint\limits_{S_\mathrm{t}} \vec{P}_1 \circ \vec{E}_2^*\,\mathrm{d}^2r \quad . \tag{4.8}$$

Only the transverse components of the arbitrary fields $\vec{E}_1$ and $\vec{H}_1$ contribute to the left integral and we may expand

$$\begin{aligned} \vec{E}_{\mathrm{t}1} &= \sum_{n=1}^{\infty}(a_n + a_{-n})\vec{\bar{E}}_{\mathrm{t}n} \\ \vec{H}_{\mathrm{t}1} &= \sum_{n=1}^{\infty}(a_n - a_{-n})\vec{\bar{H}}_{\mathrm{t}n} \end{aligned} \tag{4.9}$$

as we did in section 3.4. We want to know how strong the exitation of mode m is and set

$$\begin{aligned} \vec{E}_2 &= \vec{\bar{E}}_m \cdot \exp\{i(k_{\mathrm{z},m}\,z - \omega t)\} \\ \vec{H}_2 &= \vec{\bar{H}}_m \cdot \exp\{i(k_{\mathrm{z},m}\,z - \omega t)\} \quad . \end{aligned} \tag{4.10}$$

For guided waves we find due to orthogonality only contributions for $n = m$ in (4.8) and therefore

$$\begin{aligned} \frac{\partial}{\partial z}a_m - i\,k_{\mathrm{z},m}a_m &= i\,\omega \iint\limits_{S_\mathrm{t}} \vec{P}_1 \circ \vec{\bar{E}}_m^*\,\mathrm{d}^2r \\ \frac{\partial}{\partial z}a_{-m} + i\,k_{\mathrm{z},m}a_{-m} &= -i\,\omega \iint\limits_{S_\mathrm{t}} \vec{P}_1 \circ \vec{\bar{E}}_{-m}^*\,\mathrm{d}^2r \quad . \end{aligned} \tag{4.11}$$

The amplitudes of E_m may vary not only harmonic as we found in (3.127) especially when the perturbation is arbitrary z- dependent. With variation of the constant $a_{\pm m,0} = A_m\{z\}$ we rewrite (3.127)

$$a_{\pm m}\{z\} = A_{\pm m}\{z\} \cdot \exp\{\pm i\,k_{\mathrm{z},m}\,z\} \quad . \tag{4.12}$$

Now we can further simplify 4.11

$$\frac{\partial}{\partial z} A_{\pm m} = \pm i\,\omega \exp\{\mp i\, k_{\mathrm{z},m}\, z\} \cdot \iint\limits_{S_\mathrm{t}} \vec{P}_1 \circ \vec{\vec{E}}^{\,*}_{\pm m}\, \mathrm{d}^2 r \quad . \tag{4.13}$$

Equation (4.11) tells us that new modes m are generated at a perturbation by cross-correlation of the perturbation $\vec{P}_1$ and the electric field of a mode m. From (4.13) we learn that the exitation of a mode m at the perturbation is only efficient when the perturbation has a Fourier component k_z conjugated to k_{zm} of the mode. This case is called synchronization or phase matching. As we will see in section 3.6.3 for asynchronous coupling nearly no energy exchange occurs.

What we have to find is an expression for the integral in (4.13). The perturbation polarization can be attributed to a change in the dielectric constant $\Delta\varepsilon$ to

$$\vec{P}_1 = \varepsilon_0 \cdot \Delta\varepsilon \cdot \vec{E}_1 \tag{4.14}$$

where $\Delta\varepsilon$ can be a scalar, a tensor or a tensor that depends on the electric field present. In the first case we have a so called scalar perturbation the second and third case is called anisotropic perturbation where the third one is not only anisotropic but may be also nonlinear as it is observed with the Pockels effect or the Kerr effect.

In this section we discuss only scalar perturbations. We expand the polarization into $\vec{P}_1 = \vec{P}_\mathrm{t} + P_\mathrm{z} \cdot \vec{e}_\mathrm{z}$ and use (4.14) with a scalar $\Delta\varepsilon$. For the transverse polarization we can use a field expansion

$$\vec{P}_\mathrm{t} = \varepsilon_0 \cdot \Delta\varepsilon \cdot \vec{E}_{\mathrm{t}n} = \varepsilon_0 \cdot \Delta\varepsilon \cdot \sum_n (a_n + a_{-n}) \vec{\vec{E}}_{\mathrm{t}n} \quad . \tag{4.15}$$

The z- component can be connected to the magnetic field using 4.14 and C.2 by

$$\nabla_\mathrm{t} \times \vec{H}_\mathrm{t} = -i\,\omega(\varepsilon + \Delta\varepsilon)\varepsilon_0 E_\mathrm{z} \cdot \vec{e}_\mathrm{z} = -i\,\omega \left(\frac{\varepsilon + \Delta\varepsilon}{\Delta\varepsilon}\right) P_\mathrm{z} \cdot \vec{e}_\mathrm{z}$$

and with field expansion we find

$$\begin{aligned} P_{\mathrm{z}} &= \frac{i}{\omega} \cdot \left(\frac{\Delta\varepsilon}{\varepsilon + \Delta\varepsilon} \right) (\nabla_{\mathrm{t}} \times \vec{H}_{\mathrm{t}n}) \circ \vec{e}_{\mathrm{z}} \\ &= \frac{i}{\omega} \cdot \left(\frac{\Delta\varepsilon}{\varepsilon + \Delta\varepsilon} \right) \sum_{n=1}^{\infty} \left[\nabla_{\mathrm{t}} \times (a_n - a_{-n}) \vec{\bar{H}}_{\mathrm{t}n} \right] \circ \vec{e}_{\mathrm{z}} \quad . \end{aligned}$$

Finally results with $(\nabla_{\mathrm{t}} \times \vec{H}_{\mathrm{t}n}) \circ \vec{e}_{\mathrm{z}} = -i\,\omega\varepsilon\varepsilon_0 \cdot \bar{E}_{\mathrm{z}n}$

$$P_{\mathrm{z}} = \Delta\varepsilon \cdot \varepsilon_0 \cdot \left(\frac{\varepsilon}{\varepsilon + \Delta\varepsilon} \right) \sum_{n=1}^{\infty} (a_n - a_{-n}) \bar{E}_{\mathrm{z}n} \quad . \tag{4.16}$$

Taking (4.15) and (4.16) for the coupling-integral in (4.13) we find coupling-coefficients

$$\begin{aligned} K_{\mathrm{t}m,n} &= \omega \iint \Delta\varepsilon \cdot \varepsilon_0 (\vec{\bar{E}}_{\mathrm{t}n} \circ \vec{\bar{E}}^{\,*}_{\mathrm{t}m})\, \mathrm{d}^2 r \\ K_{\mathrm{z}m,n} &= \omega \iint \Delta\varepsilon\varepsilon_0 \left(\frac{\varepsilon}{\varepsilon + \Delta\varepsilon} \right) \cdot \bar{E}_{\mathrm{z}n} \cdot \bar{E}^{*}_{\mathrm{z}m}\, \mathrm{d}^2 r \end{aligned} \tag{4.17}$$

and going back to (4.13) a forward traveling wave is exited with (4.12)

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}z} A_m = {} & i \sum_{n=1}^{\infty} A_n (K_{\mathrm{t}m,n} + K_{\mathrm{z}m,n}) \cdot \exp\{i(k_{\mathrm{z}n} - k_{\mathrm{z}m})z\} \\ & + i \sum_{n=1}^{\infty} A_{-n} (K_{\mathrm{t}m,n} - K_{\mathrm{z}m,n}) \cdot \exp\{-i(k_{\mathrm{z}n} + k_{\mathrm{z}m})z\} \quad . \end{aligned} \tag{4.18}$$

Backward traveling wave exitation follows with $\vec{\bar{E}}_{-m} = \vec{\bar{E}}_m$ from C.5

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}z} A_{-m} = {} & -i \sum_{n=1}^{\infty} A_n (K_{\mathrm{t}m,n} - K_{\mathrm{z}m,n}) \cdot \exp\{i(k_{\mathrm{z}n} + k_{\mathrm{z}m})z\} \\ & - i \sum_{n=1}^{\infty} A_{-n} (K_{\mathrm{t}m,n} + K_{\mathrm{z}m,n}) \cdot \exp\{-i(k_{\mathrm{z}n} - k_{\mathrm{z}m})z\} \quad . \end{aligned} \tag{4.19}$$

As we pointed out before effective coupling can only occur for modes that fulfill synchronization. This means that $K_{\mathrm{t}m,n}$ and $K_{\mathrm{z}m,n}$ must vary in $z-$direction ether as fast as $\exp\{\pm i(k_{\mathrm{z}n}+k_{\mathrm{z}m})z\}$ or $\exp\{\pm i(k_{\mathrm{z}n}-k_{\mathrm{z}m})z\}$.

Codirectional coupling

The second $z-$dependence in (4.19) and the first in (4.18) originate from waves that travel in the same direction forward or backward as the incident wave. They are so called codirectional waves. Even for high order modes the $z-$dependence can be slowly varying with z. Optical mode coupling is therefore very likely. Due to z- dependent $\Delta\varepsilon$ both $K_{\mathrm{t}m,n}$ and $K_{\mathrm{z}m,n}$ are z- dependent. When $K_{\mathrm{t}m,n}$ and $K_{\mathrm{z}m,n}$ are expanded into a Fourier-series as

$$K^{(\pm)}_{m,n} = K_{\mathrm{t}m,n} \pm K_{\mathrm{z}m,n} = \sum_{\ell=-\infty}^{\infty} \kappa^{(\pm)}_{\ell,m,n} \cdot \exp\{i\,\ell\,\gamma_{m,n} z\}^{(\pm)} \tag{4.20}$$

(4.18) and (4.19) are relatively simple to solve.

The frequencies $\gamma_{m,n}$ follow directly from the geometry of the perturbation. If for example a grating of infinite length and period Λ is analyzed $\gamma_{m,n} = 2\pi/\Lambda$ follows for each pair (m,n). On the other hand a uniform perturbation of length L leads to $\gamma_{m,n} = 2\pi/L$. Other configurations are also possible up to ones with a continuous spectrum of $\gamma_{m,n}$. In that case the l- sum has to be replaced by an integral. For lossless media $\Delta\varepsilon$ is real which gives $K_{m,n} = K^*_{m,n}$ and therefore $\kappa_{\ell,m,n} = \kappa^*_{-\ell,m,n}$ and

$$K^{(+)}_{m,n} = \sum_{\ell=1}^{\infty} \kappa^{(+)}_{\ell,m,n} \exp\{i\,\ell\,\gamma^{(+)}_{m,n} z\} + \kappa^{(+)*}_{\ell,m,n} \exp\{-i\,\ell\,\gamma^{(+)}_{m,n} z\} \quad . \tag{4.21}$$

For codirectional coupling we have to take only the components with $k_{\mathrm{z}m} - k_{\mathrm{z}n}$ in (4.18) and (4.19) and find

$$\frac{\mathrm{d}}{\mathrm{d}z}A_m = i\sum_{n=1}^{\infty}\sum_{\ell=1}^{\infty} A_n \left(\kappa^{(+)}_{\ell,m,n}\exp\{i\,(k_{\mathrm{z}n} - k_{\mathrm{z}m} + \ell\,\gamma^{(+)}_{m,n})z\}\right.$$
$$\left. +\ \ \kappa^{(+)*}_{\ell,m,n}\exp\{i\,(k_{\mathrm{z}n} - k_{\mathrm{z}m} - \ell\,\gamma^{(+)}_{m,n})z\}\right) \tag{4.22}$$

where the first term is a fast varying one that can be neglected which results into

$$\frac{\partial}{\partial z}A_m \simeq i\sum_{n=1}^{\infty}\sum_{\ell=1}^{\infty} A_n\kappa^{(+)*}_{\ell,m,n}\exp\{i\,(k_{\mathrm{z}n} - k_{\mathrm{z}m} - \ell\,\gamma^{(+)}_{m,n})z\} \tag{4.23}$$

and with the same discussion

$$\frac{\partial}{\partial z}A_{-m} \simeq -i\sum_{n=1}^{\infty}\sum_{\ell=1}^{\infty} A_{-n}\kappa^{(+)*}_{\ell,m,n}\exp\{-i\,(k_{\mathrm{z}n} - k_{\mathrm{z}m} - \ell\,\gamma^{(+)}_{m,n})z\} \quad . \tag{4.24}$$

Coupling between forward or backward traveling modes follows the same mathematical equation and we can concentrate on (4.23). Two modes m and n are coupled via

$$\begin{aligned}\frac{\partial}{\partial z}A_m &= i\,\kappa^{(+)*}_{\ell,m,n}\exp\{i\,2\delta^{(+)}_{l,m,n}z\}\cdot A_n\\ \frac{\partial}{\partial z}A_n &= i\,\kappa^{(+)}_{\ell,m,n}\exp\{-i\,2\delta^{(+)}_{l,m,n}z\}\cdot A_m\end{aligned} \tag{4.25}$$

with coupling frequency

$$2\delta^{(+)}_{l,m,n} = k_{\mathrm{z}n} - k_{\mathrm{z}m} - \ell\,\gamma^{(+)}_{m,n} \tag{4.26}$$

The second equation in (4.25) has been found by exchanging m and n where $K_{n,m} = K^*_{n,m}$ has to be obeyed.

Contradirectional coupling

Modes that travel in opposite directions are attributed to $\exp\{i(k_{zm} - k_{zn})z\}$ in (4.18) and (4.19) and are so called contradirectional waves. Even for low order modes the $z-$dependence is very fast. As above we expand the coupling-coefficients into a Fourier-series

$$K^{(-)}_{m,n} = K_{tm,n} - K_{zm,n} = \sum_{\ell=1}^{\infty} \kappa^{(-)}_{\ell,m,n} \exp\{i\,\ell\,\gamma^{(-)}_{m,n} z\} + \kappa^{(-)*}_{\ell,m,n} \exp\{-i\,\ell\,\gamma^{(-)}_{m,n} z\} \quad (4.27)$$

and find neglecting the fast varying part

$$\begin{aligned} \frac{\partial}{\partial z} A_m &\simeq i \sum_{n=1}^{\infty} \sum_{\ell=1}^{\infty} A_{-n} \kappa^{(-)}_{\ell,m,n} \exp\{i\,(k_{zn} + k_{zm} - \ell\,\gamma^{(-)}_{m,n}) z\} \\ \frac{\partial}{\partial z} A_{-m} &\simeq -i \sum_{n=1}^{\infty} \sum_{\ell=1}^{\infty} A_n \kappa^{(-)*}_{\ell,m,n} \exp\{i\,(k_{zn} + k_{zm} - \ell\,\gamma^{(-)}_{m,n}) z\} \quad . \end{aligned} \quad (4.28)$$

Two modes m and n are coupled via

$$\begin{aligned} \frac{\partial}{\partial z} A_m &= i\,\kappa^{(-)}_{\ell,m,n} \exp\{i\,2\delta^{(-)}_{l,m,n} z\} \cdot A_{-n} \\ \frac{\partial}{\partial z} A_{-n} &= -i\,\kappa^{(-)*}_{\ell,m,n} \exp\{-i\,2\delta^{(-)}_{l,m,n} z\} \cdot A_m \end{aligned} \quad (4.29)$$

with

$$2\delta^{(-)}_{l,m,n} = k_{zn} + k_{zm} - \ell\,\gamma^{(-)}_{m,n} \quad . \quad (4.30)$$

4.1.3 Coupled-Mode Theory

As we have seen in the preceding section despite very complicated coupling mechanisms the coupling between two modes follows two similar very simple differential equations.

Let us here consider two modes $m = a$ and $n = b$ with amplitudes $A_m = A$ and $A_n = B$ and wavevectors $k_{zm} = \beta_a$ and $k_{zn} = \beta_b$. The electric field of the superposition follows then

$$\vec{E} = A \cdot \exp\{\pm i\,\beta_a z\} \cdot \vec{\bar{E}}_a + B \cdot \exp\{\pm i\,\beta_b z\} \cdot \vec{\bar{E}}_b \quad . \tag{4.31}$$

Under coupling-conditions we have from (4.25) or (4.29)

$$\begin{aligned} \frac{\partial}{\partial z} A &= \kappa_{a,b} \cdot B \cdot \exp\{i\,2\delta z\} \\ \frac{\partial}{\partial z} B &= \kappa_{b,a} \cdot B \cdot \exp\{-i\,2\delta z\} \quad . \end{aligned} \tag{4.32}$$

The coupling coefficients $\kappa_{a,b}$ and $\kappa_{b,a}$ are connected to $\kappa_{a,b} = i\kappa^{(+)*}_{\ell,m,n}$ and $\kappa_{b,a} = i\kappa^{(+)}_{\ell,m,n} = -\kappa^*_{a,b}$ for codirectional coupling where the coupling-phase is $\delta = \delta^{(+)}_{\ell,m,n} = \beta_b - \beta_a - \ell\,\gamma^{(+)}_{a,b}$ from (4.26). For contradirectional coupling we find $\kappa_{a,b} = i\kappa^{(-)}_{\ell,m,n}$ and $\kappa_{b,a} = -i\kappa^{(-)*}_{\ell,m,n} = \kappa^*_{a,b}$ and coupling-phase $\delta = \delta^{(-)}_{\ell,m,n} = \beta_b + \beta_a - \ell\,\gamma^{(-)}_{a,b}$ from (4.30).
In lossless waveguide we have energy-conservation and therefore

$$\frac{\partial}{\partial z}\left(|A|^2 \pm |B|^2\right) = 0 \tag{4.33}$$

where the sum is valid for codirectional coupling and the difference for contradirectional coupling. This means that any change in the energy of a mode is transferred directly to the other mode. Solutions of (4.26) are found when the system is regarded as a matrix-differential-equation. With eigenvalues and eigenvectors of such a system the solution is found very easy. Unfortunately the matrix in (4.32) is $z-$dependent and must be transformed into a $z-$independent form. We take the vector $\vec{u}\{z\} = (A\{z\}, B\{z\})^T$ and $\vec{v} = \begin{bmatrix} \exp\{-i\,\delta z\} & 0 \\ 0 & \exp\{i\,\delta z\} \end{bmatrix} \vec{u}$ which transforms (4.32) into

$$\frac{\mathrm{d}}{\mathrm{d}z}\vec{v} = \begin{bmatrix} -i\,\delta & \kappa_{a,b} \\ \kappa_{b,a} & i\,\delta \end{bmatrix} \cdot \vec{v} \quad . \tag{4.34}$$

Now the matrix is z- independent and has eigenvalues $\lambda_{1,2} = \pm\lambda$, $\lambda = \sqrt{\kappa_{a,b} \cdot \kappa_{b,a} - \delta^2}$ and eigenvector matrix $[w] = \begin{bmatrix} \kappa_{a,b} & -(\lambda + i\,\delta) \\ \lambda + i\,\delta & \kappa_{b,a} \end{bmatrix}$. The solution of (4.34) is

$$\vec{v}\{z\} = [w]\,[\exp\{\lambda_{1,2}(z - z_0)\}]\,[w]^{-1}\vec{v}\{z_0\} \tag{4.35}$$

where $\exp\{\lambda_{1,2}(z-z_0)\}$ is a diagonal matrix. Back transformation of the solution gives the solution of (4.32)

$$\begin{pmatrix} A\{z\} \\ B\{z\} \end{pmatrix} = \begin{bmatrix} m_{AA} & m_{AB} \\ m_{BA} & m_{BB} \end{bmatrix} \begin{pmatrix} A\{z_0\} \\ B\{z_0\} \end{pmatrix} \tag{4.36}$$

with matrix elements

$$\begin{aligned} m_{AA} &= \frac{1}{\lambda}(\lambda \cdot \cosh\{\lambda z\} - i\,\delta \sinh\{\lambda z\}) \cdot \exp\{i\,\delta z\} \\ m_{AB} &= \frac{1}{\lambda}\kappa_{a,b} \cdot \sinh\{\lambda z\} \exp\{i\,\delta z\} \\ m_{BA} &= \frac{1}{\lambda}\kappa_{b,a} \cdot \sinh\{\lambda z\} \exp\{-i\,\delta z\} \\ m_{BA} &= \frac{1}{\lambda}(\lambda \cdot \cosh\{\lambda z\} + i\,\delta \sinh\{\lambda z\}) \cdot \exp\{-i\,\delta z\} \end{aligned}$$

Let us consider a perturbed region of length L. A forward traveling mode b of amplitude B_0 excites forward or backward traveling modes a by co- and contradirectional coupling as sketched in figure 4.4.

In the case of co-directional coupling we have with $\kappa = |\kappa_{a,b}|$ $\lambda = i\sqrt{\kappa^2 + \delta^2}$ due to $\kappa_{b,a} = -\kappa^*_{a,b}$. Taking the case that at $z = 0$ mode a carries no energy $A\{0\} = 0$ and mode b has amplitude $B\{0\} = B_0$ we find in the perturbed region $z > 0$

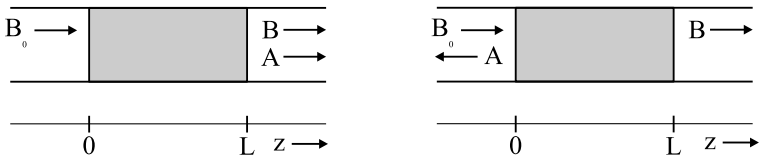

Figure 4.4: Co- and contradirectional coupling in a perturbed region of length L. Mode b of amplitude B_0 is incident and excites mode a in the coupling region.

$$A\{z\} = B_0 \cdot \frac{\kappa_{a,b}}{\sqrt{\kappa^2+\delta^2}} \cdot \exp\{i\,\delta z\} \cdot \sin\left\{\sqrt{\kappa^2+\delta^2}\,z\right\}$$
$$B\{z\} = B_0 \cdot \left(\cos\left\{\sqrt{\kappa^2+\delta^2}\,z\right\} + i\frac{\delta}{\sqrt{\kappa^2+\delta^2}} \cdot \sin\left\{\sqrt{\kappa^2+\delta^2}\,z\right\}\right) \cdot \exp\{-i\,\delta z\} \tag{4.37}$$

which has a very simple form for synchronous waves $\delta = 0$

$$A\{z\} = B_0 \cdot \frac{\kappa_{a,b}}{\kappa} \cdot \sin\{\kappa z\} \tag{4.38}$$
$$B\{z\} = B_0 \cdot \cos\{\kappa z\} \quad . \tag{4.39}$$

In figure 4.5 the absolute values of A and B are depicted for both cases $\delta = 0$ and $\delta \neq 0$.

The energy-transfer between modes a and b is for the first time maximal at the so called coupling length

$$L_C = \frac{\pi}{2} \cdot \frac{1}{\sqrt{\kappa^2+\delta^2}} \tag{4.40}$$

where the energy exchange is proportional to

$$\mathrm{Max}\left\{\left|\frac{A}{B_0}\right|^2\right\} = \frac{\kappa^2}{\kappa^2+\delta^2} \quad . \tag{4.41}$$

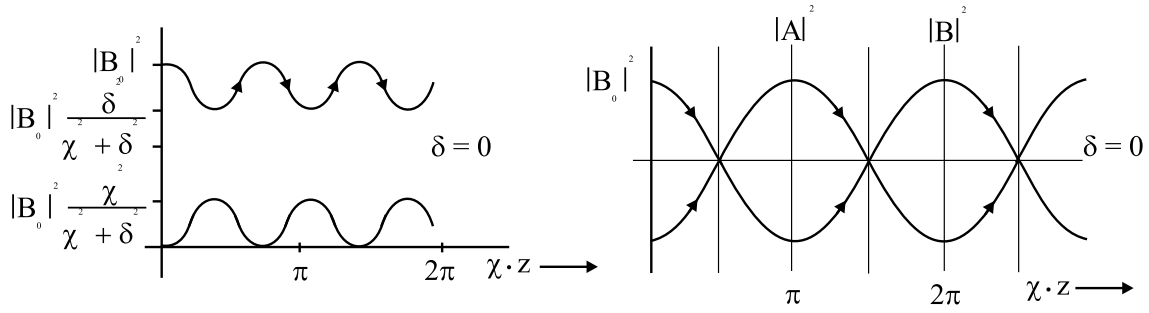

Figure 4.5: Asynchronous and synchronous co-directional coupling. As can be seen in the case of asynchronous coupling $\delta \neq 0$ only a small amount of energy is exchanged between the modes a and b.

As we see only for synchronisation $\delta = 0$ the total power can be exchanged between the two modes. For strong asynchronous waves nearly no energy is transferred as we already assumed in the preceding section.

Let us now consider contra directional coupling. Again we take $\kappa = |\kappa_{a,b}|$ and have $\lambda = \sqrt{\kappa^2 - \delta^2}$ due to $\kappa_{b,a} = \kappa^*_{a,b}$. No backward traveling wave is incident at the end of the perturbed region we set $A\{L\} = 0$. At the begin of the coupling region we have again amplitude $B\{0\} = B_0$ in the forward traveling wave. Taking this in (4.36) we find

$$A\{z\} = B_0 \cdot \frac{\kappa_{a,b}}{\lambda} \cdot \sinh\{\lambda L\} \cdot \left(\frac{\sinh\{\lambda z\}}{\sinh\{\lambda L\}} - \frac{\lambda \cosh\{\lambda z\} - i\,\delta \sinh\{\lambda z\}}{\lambda \cosh\{\lambda L\} - i\,\delta \sinh\{\lambda L\}} \right) \cdot \exp\{i\,\delta z\} \tag{4.42}$$

$$B\{z\} = B_0 \cdot \frac{\lambda \cosh\{\lambda(L-z)\} - i\,\delta \sinh\{\lambda(l-z)\}}{\lambda \cosh\{\lambda L\} - i\,\delta \sinh\{\lambda L\}} \cdot \exp\{-i\,\delta z\}$$

which for synchronisation $\delta = 0$ reduces to

$$A\{z\} = B_0 \cdot \frac{\kappa_{a,b}}{\kappa} \cdot \sinh\{\kappa L\} \left(\frac{\sinh\{\kappa z\}}{\sinh\{\kappa L\}} - \frac{\cosh\{\kappa z\}}{\cosh\{\kappa L\}} \right) \tag{4.43}$$

$$B\{z\} = B_0 \cdot \frac{\cosh\{\kappa(L-z)\}}{\cosh\{\kappa L\}} \quad . \tag{4.44}$$

Figure 4.6 shows the energy exchange between the input wave B and excited wave A for synchronization. In the case of asynchronous coupling the exchange amount of energy is lowered and for $\delta > \kappa$ also backcoupling from A to B occurs but the amount of coupled energy still decreases with growing phase mismatch.

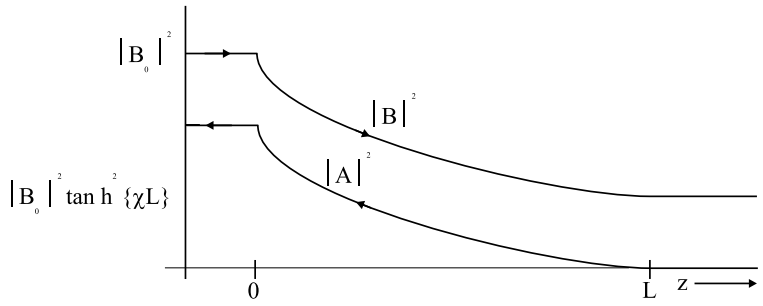

Figure 4.6: Synchronous contradirectional coupling in a perturbed region of length L. Energy is coupled from wave b to a nearly exponentially

As we see from figures 4.5 and 4.6 for synchronization contradirectional coupling is not as efficient as codirectional coupling. Only for long coupling regions $\kappa \cdot L \gg 1$ a remarkable amount of energy is coupled to the contradirectional traveling mode. This is a strong design-criterion for DFB- and DBR-lasers.

4.1.4 Lateral Waveguide-Mode Coupling

If two waveguides are placed very near to each other the evanescent fields of the modes extend remarkably into the neighboring waveguide. In this case there is a perturbation polarization in that waveguide present that can excite a mode. On the other hand the exciting mode is slightly perturbed such that the phase-constant k_z changes to $k_z + \delta k_z$. The perturbation polarization that is caused for a mode in waveguide a in the presence of a waveguide b can be calculated from the change in the dielectric constants.
The way to find the coupling is sketched in figure 4.7. in the presence of a second waveguide which interacts with the incident wave the interaction leads to a virtually increased effective index (increased k_z) and a perturbation polarization which replaces

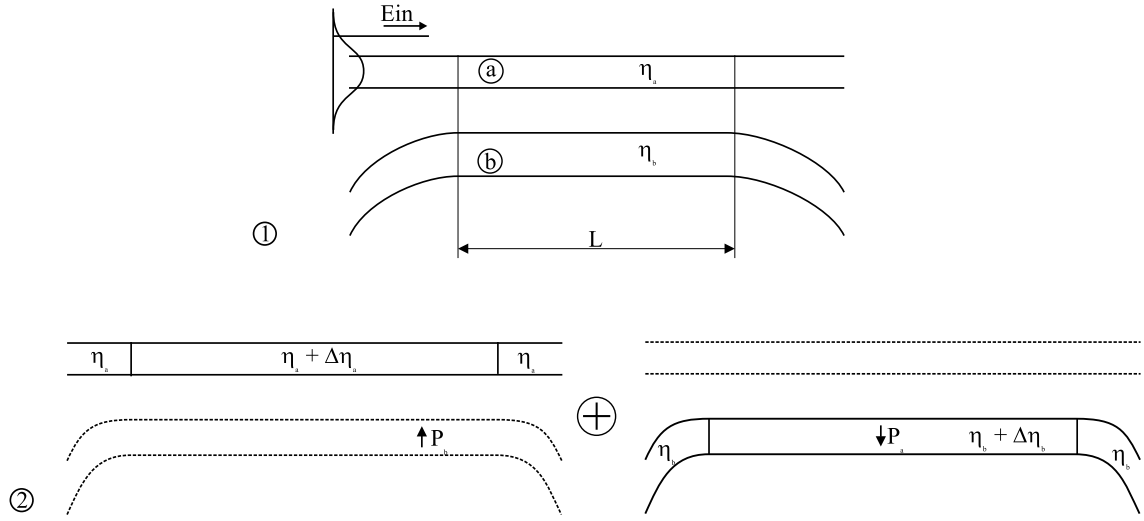

Figure 4.7: Lateral waveguide mode coupling. The presence of waveguide b is accounted by an increase propagation constant and a perturbation polarization in a caused by b and vice versa.

the first waveguide. The polarization generates modes in the second waveguide. After traveling distance dz the mode in waveguide b generates a field in waveguide a which interferes with the original field in the waveguide. At the same time waveguide a produces a new field in waveguide b which as well interferes with the already existing field. After coupling length L generally in both waveguides a field exists despite the fact that only one waveguide was exited at the begin of the coupler. Efficient coupling can only occur when k_z in both waveguides has the same value and both waves are assumed to travel in the same direction. Contradirectional coupling is not likely because there is no z- dependent coupling in the coupler which could generate back traveling waves.

Let us consider a waveguide that can be described by $\eta_a\{x,y\} = \eta_s + \Delta\eta_a\{x,y\}$ and $\eta_b\{x,y\} = \eta_s + \Delta\eta_b\{x,y\}$ as shown in figure 4.8.

The dielectric constant follows from $\varepsilon = \eta^2 = (\eta_s + \Delta\eta_a\{x,y\} + \Delta\eta_b\{x,y\})^2$ and the perturbation for waveguide a is $\Delta\varepsilon_a = \varepsilon - \eta_a^2 = (\eta_s + \Delta\eta_a + \Delta\eta_b)^2 - (\eta_s + \Delta\eta_a)^2$ and for waveguide b follows from $\Delta\varepsilon_b = \varepsilon - \eta_b^2$. The whole perturbation polarization is

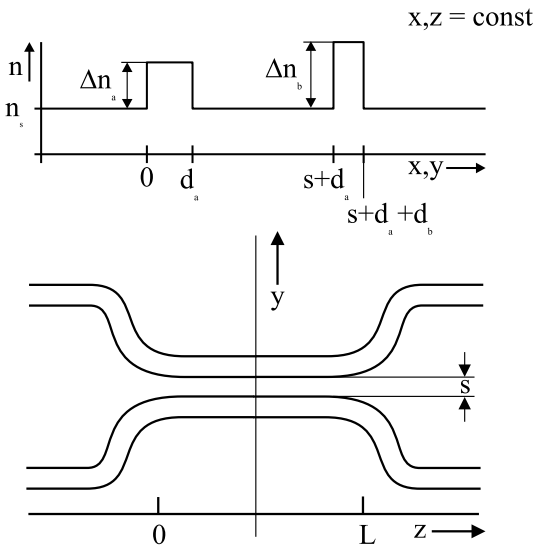

Figure 4.8: Lateral waveguide coupling for closely neighboring waveguides. The upper figure shows the index-distribution assumed and the lower figure gives an overview of the coupler.

given by

$$\vec{P} = \Delta\varepsilon_{\mathrm{a}} \cdot \varepsilon_0 \cdot A_m \cdot \vec{\vec{E}}_m^{(a)} \cdot \exp\{i\, k_{zm}^{(a)} z\} + \Delta\varepsilon_{\mathrm{b}} \cdot \varepsilon_0 \cdot B_n \cdot \vec{\vec{E}}_n^{(b)} \cdot \exp\{i\, k_{zn}^{(b)} z\} \quad (4.45)$$

when we assume that only one mode m in waveguide a and n in waveguide b is present. Application of (4.13) for both modes results into

$$\frac{\mathrm{d}}{\mathrm{d}z}\begin{pmatrix} A_m \\ B_n \end{pmatrix} = i \begin{pmatrix} \delta k_{zm}^{(a)} & \kappa_{a,b} \cdot \exp\{-i\,(k_{zm}^{(a)} - k_{zn}^{(b)})z\} \\ \kappa_{a,b} \cdot \exp\{i\,(k_{zm}^{(a)} - k_{zn}^{(b)})z\} & \delta k_{zn}^{(b)} \end{pmatrix} \begin{pmatrix} A_m \\ B_n \end{pmatrix} \quad (4.46)$$

with abbreviations

$$\begin{aligned} \delta k_{\mathrm{z}} &= \omega\varepsilon_0 \iint\limits_S \Delta\varepsilon \left|\vec{\vec{E}}\right|^2 \,\mathrm{d}x\,\mathrm{d}y \\ \kappa_{a,b} &= \omega\varepsilon_0 \iint\limits_S \Delta\varepsilon_{\mathrm{b}} \cdot \vec{\vec{E}}_n^{(b)} \circ \vec{\vec{E}}_m^{(a)*} \,\mathrm{d}x\,\mathrm{d}y \\ \kappa_{b,a} &= \omega\varepsilon_0 \iint\limits_S \Delta\varepsilon_{\mathrm{a}} \cdot \vec{\vec{E}}_m^{(a)} \circ \vec{\vec{E}}_n^{(b)*} \,\mathrm{d}x\,\mathrm{d}y \quad . \end{aligned} \quad (4.47)$$

As in the sections before (4.52) must be transformed into a system with $z-$independent matrix elements. With

$$\begin{aligned} \tilde{A}_m &= A_m \cdot \exp\{i(\delta k_{zm}^{(a)} + \delta)z\} \\ \tilde{B}_n &= B_n \cdot \exp\{i(\delta k_{zn}^{(b)} + \delta)z\} \end{aligned} \tag{4.48}$$

(4.52) is transformed into the required system that looks like (4.34). For δ in (4.48) we have to take $2\delta = k_{zn}^{(b)} - k_{zm}^{(a)} + \delta k_{zn}^{(b)} - \delta k_{zm}^{(a)}$. In the case of two identical waveguides we find phase-matching $\delta = 0$ and total energy transfer can occur over the coupling-length $L_\mathrm{c} = \frac{\pi}{\kappa}$ as we have seen in section 3.6.4 for codirectional coupling. When the incident wave is split into two equal portions on both waveguides such a coupler is called a 3dB coupler.

As an example let us consider two identical waveguides of width $d_\mathrm{a} = d_\mathrm{b} = d$ with spacing s between them. Additionally let us assume strong guiding. Due to the geometric symmetry from (4.47) follows

$$\kappa = \omega\varepsilon_0 \iint\limits_S \Delta\varepsilon \cdot \vec{\bar{E}}_m\{x,y\} \cdot \vec{\bar{E}}_m^*\{x,y+s\}\ \mathrm{d}x\ \mathrm{d}y \tag{4.49}$$

with $\Delta\varepsilon = (\eta_\mathrm{s} + \Delta\eta)^2 - \eta_\mathrm{s}^2 = \eta_\mathrm{f}^2 - \eta_\mathrm{s}^2$ at the position of the neighboring waveguide and $\Delta\varepsilon = 0$ elsewhere such that for

$$\kappa = \omega\varepsilon_0(\eta_\mathrm{f}^2 - \eta_\mathrm{s}^2) \int\limits_0^d \vec{\bar{E}}\{y\} \cdot \vec{\bar{E}}^*\{s+y\}\ \mathrm{d}y \tag{4.50}$$

follows. In strongly guiding waveguides we find $\vec{E} \simeq \vec{E}_\mathrm{t}$ and the field inside $E_m\{y\} = E\{\frac{d}{2}\} \cdot \cos\{\beta_\mathrm{y} y\}$ is periodic. The field $E_n\{y+s\}$ is assumed to be the evanescent field which is proportional to the field in the middle of the waveguide by $E_n\{s+y\} = E\{\frac{d}{2}\} \cdot \exp\{-\alpha_m(s + \frac{d}{2} + y)\}$ with $\alpha_m \simeq \sqrt{\beta^2 - \eta_\mathrm{s}^2 k_0^2} \simeq k_0\sqrt{2\eta_\mathrm{f}(\eta_\mathrm{f} - \eta_\mathrm{s})} \simeq k_\mathrm{y}$ for

strong guiding. Taking this we may set E_n constant ($y = 0$) under assumption that the evanescent field E_n varies very slowly compared to the field E_m in the waveguide. For the coupling coefficient follows

$$\begin{aligned}\kappa &\simeq \omega\varepsilon_0(\eta_\mathrm{f}^2 - \eta_\mathrm{s}^2) \cdot \exp\{-\alpha_m(s + \tfrac{d}{2})\} \cdot \int_x \left|\vec{E}\{x, \tfrac{d}{2}\}\right|^2 \,\mathrm{d}x \cdot \int_0^d \cos\{\alpha_m y\}\,\mathrm{d}y \\ &\simeq \frac{2\sqrt{2(\eta_\mathrm{f} - \eta_\mathrm{s})}}{\sqrt{\eta_\mathrm{f}}\, d} \cdot \exp\left\{-\frac{2\pi}{\lambda} \cdot \sqrt{2\eta_\mathrm{f}(\eta_\mathrm{f} - \eta_\mathrm{s})} \cdot (s + \tfrac{d}{2})\right\} \quad . \end{aligned} \tag{4.51}$$

Taking $\lambda = 1\mu\mathrm{m}$, $\eta_\mathrm{f} = 3.5$, $\Delta\eta = \eta_\mathrm{f} - \eta_\mathrm{s} \simeq 5\,\mathrm{E}{-3}$ and $s = d = 3\mu\mathrm{m}$ we find $\kappa \simeq \frac{2}{\mathrm{cm}}$ and for total energy transfer from a to b we need a coupling length of $L_\mathrm{c} \simeq 3\mathrm{cm}$. Stronger guiding Δn or more spacing lower the coupling coefficient and make longer coupling zones necessary.

4.1.5 Anisotropic Media Mode-Coupling

In anisotropic media the electrical displacement $\vec{D}$ and the electric field are connected via the dielectric tensor $[\varepsilon_\mathrm{E}]$

$$\vec{D} = \varepsilon_0[\varepsilon_\mathrm{E}] \cdot \vec{E} \tag{4.52}$$

which can be field dependent. Historically the field dependence was written for $[g_\mathrm{E}] = [\varepsilon_\mathrm{E}]^{-1}$ in the form

$$[g_\mathrm{E}] = [g] + [\Delta g] = [g] + \begin{bmatrix} \Delta g_1 & \Delta g_6 & \Delta g_5 \\ \Delta g_6 & \Delta g_2 & \Delta g_4 \\ \Delta g_5 & \Delta g_4 & \Delta g_3 \end{bmatrix} \tag{4.53}$$

where for the linear electrooptic effect

$$\begin{pmatrix} \Delta g_1 \\ \vdots \\ \Delta g_6 \end{pmatrix} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ \vdots & \vdots & \vdots \\ r_{6,1} & r_{6,2} & r_{6,3} \end{pmatrix} \cdot \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} \tag{4.54}$$

results.

For the main materials used in optoelectronical applications $[g]$ is a diagonal tensor

$$[g] = \begin{bmatrix} \frac{1}{\varepsilon_{1,1}} & 0 & 0 \\ 0 & \frac{1}{\varepsilon_{2,2}} & 0 \\ 0 & 0 & \frac{1}{\varepsilon_{3,3}} \end{bmatrix} \tag{4.55}$$

with the proper choice of the coordinate system and all elements of Δg are small compared to the elements of $[g]$ such that with

$$g_{i,j} \simeq \begin{cases} \frac{1}{\varepsilon_{i,j}} & \text{for } i = j \\ -\frac{\varepsilon_{i,j}}{\varepsilon_{i,i}\varepsilon_{j,j}} & \text{for } i \neq j \end{cases} \tag{4.56}$$

and $[\varepsilon_{\mathrm{E}}] = [\varepsilon] + [\Delta\varepsilon]$ can be calculated from

$$[\Delta\varepsilon] \simeq -\begin{bmatrix} \Delta g_1\, \varepsilon_{1,1}^4 & \Delta g_6\, \varepsilon_{1,1}^2\, \varepsilon_{2,2}^2 & \Delta g_5\, \varepsilon_{1,1}^2\, \varepsilon_{3,3}^2 \\ \Delta g_6\, \varepsilon_{1,1}^2\, \varepsilon_{2,2}^2 & \Delta g_2\, \varepsilon_{2,2}^4 & \Delta g_4\, \varepsilon_{2,2}^2\, \varepsilon_{3,3}^2 \\ \Delta g_5\, \varepsilon_{1,1}^2\, \varepsilon_{3,3}^2 & \Delta g_4\, \varepsilon_{2,2}^2\, \varepsilon_{3,3}^2 & \Delta g_3\, \varepsilon_{3,3}^4 \end{bmatrix} \quad . \tag{4.57}$$

As an example we take GaAs- or InP-based material systems. Along the crystal-axes $a = 1$, $b = 2$ and $c = 3$ all elements of $[\varepsilon]$ are equal $\varepsilon_{1,1} = \varepsilon_{2,2} = \varepsilon_{3,3} = \eta^2$. In the electrooptic tensor $[r]$ only three elements are present $r_{4,1} = r_{5,2} = r_{6,3} \neq 0$.

The crystal-axes are $\vec{a} = \vec{e}_{\mathrm{x}}$, $\vec{b} = \cos\{\frac{\pi}{4}\} \cdot \vec{e}_{\mathrm{y}} - \sin\{\frac{\pi}{4}\} \cdot \vec{e}_{\mathrm{z}}$ and $\vec{c} = \cos\{\frac{\pi}{4}\} \cdot \vec{e}_{\mathrm{y}} + \sin\{\frac{\pi}{4}\} \cdot \vec{e}_{\mathrm{z}}$ as depicted in figure 4.9 where the coordinates x, y and z are aligned to the edges of the crystal which are obeyed when the crystal grows or is broken. Usually the $(100)-$surface is the surface of a wafer.

Application of an electric field normal to the surface $\vec{E} = E_1 \cdot \vec{e}_{\mathrm{x}}$ results into a perturbation $\Delta g_4 = r_{4,1} \cdot E_1$. The refractive index splits into three parts $\eta_{\mathrm{x}} = \eta$, $\eta_{\mathrm{y}} \simeq \eta + \eta^3 r_{4,1} E_1$ and $\eta_{\mathrm{z}} \simeq \eta - \eta^3 r_{4,1} E_1$. An optical wave traveling into $z-$direction is only influenced when it has an electric field into y- direction. Such a wave is transversal electric with

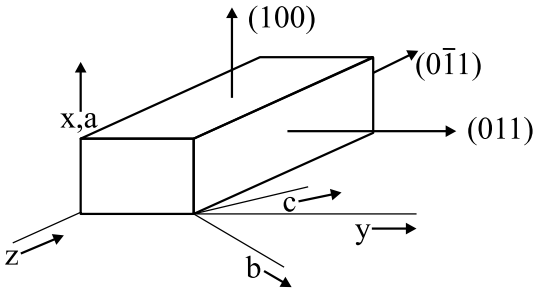

Figure 4.9: Axes in zincblende crystals. The axes a, b, c denote the optical main axes where $[\varepsilon]$ is diagonal. The cartesian coordinates are aligned to the natural breaking edges.

respect to the wafer surface and therefore only TE-modes are perturbed by the applied electric field E_1.

If the applied field shows in y- direction $\vec{E} = E \cdot \vec{e}_{\mathrm{y}} = \frac{1}{2}\sqrt{2} \cdot E \cdot (\vec{b} + \vec{c})$ both Δg_5 and Δg_6 are occurring $\Delta g_5 = \Delta g_6 = \frac{1}{2}\sqrt{2} \cdot r_{4,1} \cdot E$ and a TE-wave gives a perturbation for fields in x- direction and z- direction whereas a TM-wave produces a polarization into y- direction and z- direction. This means that a TE wave excites a TM one and vice versa. With applied field in y- direction the polarization of a traveling wave can be switched.

4.2 Mode Radiation into Free Space

In section 3.6.1 looked at the end-face coupling of two waveguides and calculated the coupling efficiency by means of excited modes. In free space there are no longer modes that we can take for a field-expansion. As is observed in nature the light radiated from a fiber or laser into free space propagates more or less like a beam with low divergence. This fact can be described with a paraxial wave-equation. Solutions of that equation are Hermite-Gauß or Laguerre-Gauß waves depending on the coordinate system used (cartesian or cylindrical coordinates). The waves exhibit orthogonality and can be considered to be the modes of free space and the radiation into free space can be calculated as in the preceding sections. A very nice behaviour of that modes is that near- and far-field have the same spatial distribution. The characteristic parameters are the beam-waist diameter and radius of divergence and their evolution under propagation can be calculated by means of a matrix transformation.

4.2.1 Paraxial Waves in Free Space

A light beam in free space diverges only by a certain angle. This fact can be described by waves that propagate nearly parallel to the axis of the beam. Such waves are called to be paraxial. For a mathematical description we start from the scalar wave-equation $\Delta E + k^2 E = 0$ which holds for monochromatic time-harmonic fields. Let the $z-$axis be the direction of propagation with starting point z_0 we start with a Bloch wave $E = u\{\vec{r}\} \cdot \exp\{i(k(z - z_0) - \omega t)\}$ and find for the field distribution $\Delta u + i\,2k\frac{\mathrm{d}}{\mathrm{d}z}u = 0$. Paraxial waves with small divergence-angels fulfill $\frac{\mathrm{d}^2}{\mathrm{d}z^2}u \ll \left\{ \frac{\mathrm{d}^2}{\mathrm{d}x^2}u;\ \frac{\mathrm{d}^2}{\mathrm{d}y^2}u;\ 2k\frac{\mathrm{d}}{\mathrm{d}z}u \right\}$ and we find the so called paraxial wave-equation

$$\nabla_{\mathrm{s}}^2 u\{\vec{s}, z\} + i\,2k\frac{\mathrm{d}}{\mathrm{d}z}u\{\vec{s}, z\} = 0 \tag{4.58}$$

where $\vec{s}$ may be noted as $\vec{s} = x \cdot \vec{e}_{\mathrm{x}} + y \cdot \vec{e}_{\mathrm{y}}$ in cartesian coordinates or $\vec{s} = \varrho \cdot \vec{e}_{\varrho}$ in cylindrical coordinates with corresponding ∇_{s}- operators. In cartesian coordinates the solution of (4.58) are Hermite-Gauß-waves

$$u = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} A_{m,n} \cdot u_{\mathrm{x}m} \cdot u_{\mathrm{y}n} \tag{4.59}$$

with amplitudes $A_{m,n}$ and field-distribution

$$\begin{aligned} u_{\mathrm{x}m} &= \left(\frac{p_{\mathrm{x}}^2 k^2}{p_{\mathrm{x}}^2 k^2 - ik(z - z_0)}\right)^{\frac{m}{2}-1} \cdot \mathrm{H}_m\{\tfrac{x}{p_{\mathrm{x}}}\} \cdot \exp\left\{-\tfrac{x^2}{2p_{\mathrm{x}}^2}\right\} \\ u_{\mathrm{y}n} &= \left(\frac{p_{\mathrm{y}}^2 k^2}{p_{\mathrm{y}}^2 k^2 - ik(z - z_0)}\right)^{\frac{n}{2}-1} \cdot \mathrm{H}_n\{\tfrac{y}{p_{\mathrm{y}}}\} \cdot \exp\left\{-\tfrac{y^2}{2p_{\mathrm{y}}^2}\right\} \quad . \end{aligned} \tag{4.60}$$

where

$$p_{\mathrm{x}} = \sqrt{\frac{v_{\mathrm{x}}}{k} + i\,\frac{z - z_0}{k}}$$

$$p_{\mathrm{y}} = \sqrt{\frac{v_{\mathrm{y}}}{k} + i\,\frac{z - z_0}{k}} \tag{4.61}$$

and H is the Hermite polynom

$$\begin{aligned} \mathrm{H}_0\{x\} &= 1 \\ \mathrm{H}_1\{x\} &= 1 \\ \mathrm{H}_{m+1}\{x\} &= x \cdot \mathrm{H}_m\{x\} - \mathrm{H}_{m-1}\{x\} \quad . \end{aligned} \tag{4.62}$$

The values of v_{x} and v_{y} in (4.60) can be free chosen and therefore adjusted such that the observed beam can be described with a small number of Hermite-Gauß waves. When we look at the argument of the exp $-$function we can split it into real- and imaginary part by the expansion

$$-\frac{1}{2p^2} = -\frac{1}{w^2} + i\,\frac{k}{2R_{\mathrm{b}}} \tag{4.63}$$

with beam waist width w and bending-radius R_{b}. By comparison we find

$$\begin{aligned} \frac{1}{w^2} &= \frac{v}{v^2 + (z - z_0)^2} \cdot \frac{k}{2} = \frac{1}{w_0^2} \cdot \frac{1}{1 + \left(\frac{z - z_0}{v}\right)^2} \\ \frac{1}{R_{\mathrm{b}}} &= \frac{z - z_0}{v^2 + (z - z_0)^2} = \frac{1}{z - z_0} \cdot \frac{1}{1 + \left(\frac{v}{z - z_0}\right)^2} \end{aligned} \tag{4.64}$$

In the beam waist at $z = z_0$ R is infinite and $w = \sqrt{\frac{2v}{k}} = w_0$ is minimal. Here we can adjust v in such way that the $\frac{1}{e}-$width d_0 of the observed beam coincides with $2w = d_0$ such that $v = \frac{k d_0^2}{8}$. At big distances from the waist $z - z_0 \gg v$ we can simplify $w \simeq w_0 \cdot \frac{z - z_0}{v}$ and $R_{\mathrm{b}} \simeq z - z_0$. Often v is referred as Rayleigh distance $z_{\mathrm{R}} = v$. For an arbitrary field profile $u\{x, y, z\}$ at any position z we can calculate the amplitudes $A_{m,n}$ of the Hermite-Gauß-waves in (4.59) by expansion of that field. This is possible due to

the fact that the Hermite-Gauß-modes are a complete set of orthogonal functions with orthogonality

$$\frac{1}{\sqrt{\pi}\,n!}\int_{-\infty}^{\infty}\mathrm{H}_m\{\xi\}\cdot\exp\left\{-\frac{\xi^2}{2}\right\}\cdot\mathrm{H}_n\{y\}\;\mathrm{d}\xi=\delta_{m,n} \tag{4.65}$$

such that we find with substitution $\xi=\frac{x}{p_\mathrm{x}}$ and $\xi=\frac{y}{p_\mathrm{y}}$

$$A_{m,n}=\frac{1}{\pi\,m!\,n!\,p_\mathrm{x}\,p_\mathrm{y}}\cdot\left(\frac{1}{1-i\frac{z_1-z_0}{kp_\mathrm{x}^2}}\right)^{1-\frac{m}{2}}\cdot\left(\frac{1}{1-i\frac{z_1-z_0}{kp_\mathrm{y}^2}}\right)^{1-\frac{n}{2}}\cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}u\{x,y,z\}\cdot\mathrm{H}_m\{\tfrac{x}{p_\mathrm{x}}\}\cdot\mathrm{H}_n\{\tfrac{y}{p_\mathrm{y}}\}\;\mathrm{d}x\;\mathrm{d}y\quad .$$

The position of the beam waist and the waist-diameter can be taken as arbitrary values. If they meet the real physical data often only a small number of modes are present ($A_{m,n}\neq 0$) and a reasonable interpretation of the physical behaviour can be deduced from (4.59).

The Hermite-Gauß-modes explain very well the beam-propagation emitting from a rectangular area as we find it at the edge of edge-emitting laser-diodes. If we want to calculate the radiation from a circular area as we find in fibers, we better switch to the representation of (4.58) in cylindrical coordinates. Solutions of (4.58) in cylindrical coordinates are Laguerre-Gauß-waves with

$$u\{\varrho,\varphi,z\}=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\;A_{m,n}\left(\frac{1}{1-i\frac{z-z_0}{kp^2}}\right)^{-(m+n+1)}\cdot\left(\frac{\varrho}{\sqrt{2}\,p}\right)^n \cdot\mathrm{L}_m^{(n)}\left\{\frac{\varrho^2}{2p^2}\right\}\cdot\exp\left\{-\frac{\varrho}{2p^2}\right\}\cdot\exp\{i\,n\varphi\} \tag{4.66}$$

where $\mathrm{L}_m^{(n)}$ denote the Laguerre-polynomials

$$\mathrm{L}_0^{(n)}\{x\}=1$$

$$\begin{aligned} \mathrm{L}_1^{(n)}\{x\} &= 1+n-x \\ \mathrm{L}_{m+1}^{(n)}\{x\} &= \frac{1}{m+1}\left[(1+n+2m-x)\,\mathrm{L}_m^{(n)}\{x\} - (m+n)\,\mathrm{L}_{m-1}^{(n)}\{x\}\right] \end{aligned} \tag{4.67}$$

and p defined like (4.61) with $p = p_\mathrm{x}$ for $v_\mathrm{x} = v$. The expression for $\mathrm{L}_{m+1}^{(n)}$ is only valid for $m \geq 1$. Again the argument of the exp −function can be split into real and imaginary part and we find the same interpretation as in (4.64) with the exception that w now is the beam waist radius instead of width. Again the Laguerre-Gauß-waves are a complete set of orthogonal functions which follow

$$\frac{1}{\Gamma\{1+n\}\cdot\binom{m+n}{m}}\int_0^\infty \mathrm{L}_m^{(n)}\{x\}\cdot x^n\cdot\exp\{-x\}\cdot\mathrm{L}_{m'}^{(n)}\{x\}\;\mathrm{d}x = \delta_{m,m'} \tag{4.68}$$

and therefore we can calculate the amplitudes $A_{m,n}$ of the Laguerre-Gauß-waves in (4.66) by expansion

$$\begin{aligned} A_{m,n} = {} & \frac{1}{\binom{m+n}{n}\Gamma\{1+n\}}\cdot\left(\frac{1}{1-i\frac{z-z_0}{kp^2}}\right)^{m+n+1}\cdot \\ & \int_0^{2\pi}\int_0^\infty u\{\varrho,\varphi,z\}\cdot\exp\{-i\,n\varphi\}\cdot\left(\frac{\varrho}{p\sqrt{2}}\right)^n\cdot\mathrm{L}_m^{(n)}\left\{\frac{\varrho^2}{2\,p^2}\right\}\;\mathrm{d}\varphi\;\mathrm{d}\left(\frac{\varrho^2}{2\,p^2}\right) \end{aligned} \tag{4.69}$$

4.2.2 Near- and Farfield

Paraxial beams satisfy the Fresnell-condition

$$k_\mathrm{z}^2 \gg k_\mathrm{x}^2 + k_\mathrm{y}^2 \tag{4.70}$$

where $\vec{k} = k_\mathrm{x}\cdot\vec{e}_\mathrm{x} + k_\mathrm{y}\cdot\vec{e}_\mathrm{y} + k_\mathrm{z}\cdot\vec{e}_\mathrm{z}$ is the wavenumber-vector which is found when the field is decomposed into plane waves by spatial Fourier transformation. The point under observation is at distance $R = z - z_0$ from the emitting area. The field in the emitting area which exites the radiating waves is called the near-field. The farfield can be

calculated from the nearfield-distribution not only for paraxial waves by decomposition of the exiting nearfield into plane waves. The propagation of that waves over distance R is calculated an after recomposition of the propagated plane waves we find the farfield-distribution. In the case that the Fraunhofer-condition

$$\lambda \cdot R \gg (x - x_0)^2 + (y - y_0)^2 \tag{4.71}$$

is satisfied for any point of the emitting area nearfield and farfield are simply related to each other by spatial Fourier-transformation. Let the field distribution of the emitting area be $u\{x, y, z\}\Big|_{z=z_0} = u_0\{x_0, y_0\}$ we find the farfield with spatial Fourier-transformation.

$$u\{x, y, z = z_0 + R\} = \frac{1}{i\,\lambda R} \cdot \exp\{i\,2\pi\tfrac{R}{\lambda}\} \cdot \exp\left\{-i\,\pi\,\frac{x^2 + y^2}{\lambda R}\right\} \cdot \tag{4.72}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_0 \cdot \exp\left\{-i\,2\pi\frac{x \cdot x_0 + y \cdot y_0}{\lambda R}\right\}\ \mathrm{d}x_0\ \mathrm{d}y_0 \tag{4.73}$$

Here x_0 and y_0 are coordinates in the emitting area at $z = z_0$ and x and y are points under observation at distance R from the emitting area. We have to note that the $z-$axis has to be aligned normal to the emitting area.

4.3 Fourier- and Gauß-Optics

In (4.72) the integral is a spatial Fourier transformation with frequencies $\nu_\mathrm{x} = \frac{x}{\lambda R}$ and $\nu_\mathrm{y} = \frac{y}{\lambda R}$. From the physical point of view this means that at a given point $(x, y, z = z_0 + R)$ far away from the source we find the amplitude at frequency ν_x and ν_y of the spatial Fourier spectrum only modified by a phaseshift and the observed intensity is directly proportional to the power-density spectrum. This knowledge is used in the so called $4f-$setup where an optical beam is modified via spectral filtering as sketched in figure 4.10.

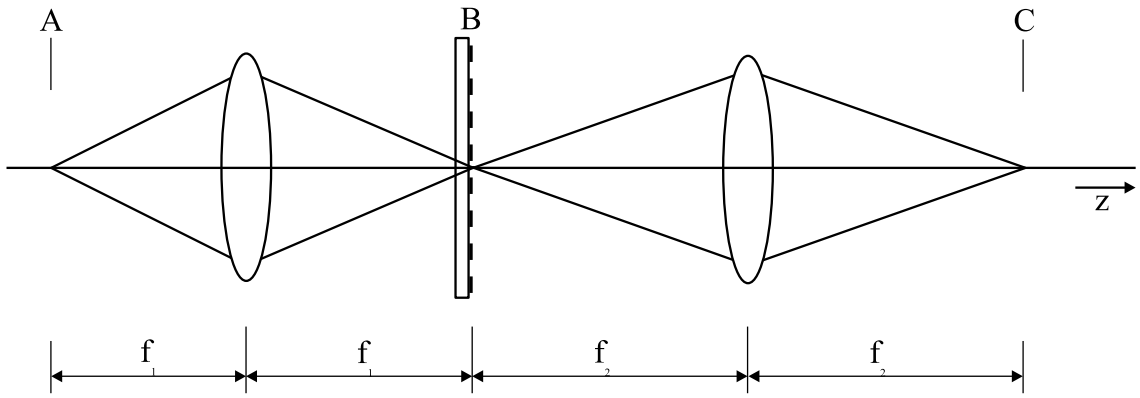

Figure 4.10: $4f-$setup. With the left lens the field in plane A is Fourier transformed with Fourier-spectrum in plane B. An optical filter mode of a coated glass plate or hologram changes the spectrum and the back transformation is done by the right lens in plane C. When both focal lengths are equal $f_1 = f_2 = f$ the distance between A and C is $4f$

Two lenses with focal lengths f_1 and f_2 transform the field distribution in plane A into Fourier spectrum in plane B and back into the spatial distribution in plane C. When both focal lengths are equal $f_1 = f_2 = f$ the distance between plane A and C is $4f$ which is the origin of the setup name. When a coated glass plate is introduced in plane B the Fourier spectrum of the incident light in plane A is modified and therefore a modified spatial distribution is found in plane C. Due to the fact that the modification takes place in the spectral region the effect is the same as spectral filtering well known form electrical circuits. If the filter is an aperture with transmission only in a small region around the $z-$axis only low frequencies are transmitted because x and y are small quantities and we have a low passfilter. On the other hand a high pass filter is a small opaque dot on the $z-$axis and a band pass filter is a transparent ring around the $z-$axis as sketched in figure 4.11.

The optical output in plane C is directly the Fourier transform of the filter when the incident wave in plane A can be regarded as a Dirac distribution. For each element in an optical setup the Fourier transformation can be calculated and with sufficient distance

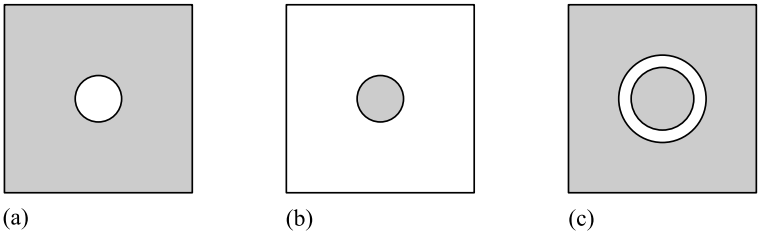

Figure 4.11: Low pass (a), high pass (b) and band pass filter (c) for optical filtering in the $4f-$setup.

between the elements the output-beam can be calculated by simply multiplying the Fourier transforms of the elements. This kind of calculation for an optical setup is the so called Fourier optics.

In Gauß-optics the input beam of an optical system is regarded as a Gaussian beam which is true for the paraxial approximation where each beam can be expanded into Gaussian beams as we have seen in section 3.7.2. The Gaussian beams exhibit the same spatial distribution in every plane z and only the beam waist diameter w and the radius of curvature R are changed with growing distance z. In the case of big distances from the beam waist $z - z_0 \gg v$ find in free space $w\{z\} = w_0 \cdot (z - z_0)$ and $R\{z\} = z - z_0$. A beam traveling in a homogeneous medium with refractive index η follows $\frac{w}{\eta} = \frac{w_0}{\eta} \cdot (z - z_0)$. Instead of calculating both w and R we go back to (4.61) and take

$$q = -i\frac{2p^2}{k} = \frac{2}{k}(z - z_0 - iv) = \frac{1}{R} - i\frac{\lambda}{\pi w^2}$$

and find for any given distance $z - z_0 = L$

$$\frac{q}{\eta_2} = \frac{A\frac{q}{\eta_1} + B}{C\frac{q}{\eta_1} + D} \tag{4.74}$$

where we assumed that the refractive index may have changed from η_1 to η_2 after distance L. In free space we find $A = 1$, $B = L$, $C = 0$ and $D = 1$. This can be written

in a matrix form as so called $ABCD-$matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \tag{4.75}$$

For any element in an optical system such a $ABCD-$matrix can be found and the output beam of such a system then is calculated by simple multiplication of all $ABCD-$matrices and calculating the beam waist and bending radius via (4.74) with the resulting overall $ABCD-$matrix for the output beam. This kind of calculation of an optical setup is called Gauß-optics.

4.4 Spatial Coherence and Beam-Quality

An optical beam is not only described by its wavelength and transverse profile. As it travels through space the beam diameter growth. When the growth is compared to an ideal gaussian beam it is found that the real beam widens faster than the ideal gaussian one. This behaviour is described by an empirical factor called M^2.
A second characteristic measure for an optical beam is the ability that two different parts of it can interfere. The coherence length is a measure at which distance between that two parts the ability of interference drops to a certain amount.
As a special kind of spatial coherence in optical nonlinear matter solitons can build up. An optical pulse of certain shape and with high power travels without dispersion over big distances giving the possibility of higher bit-rate transmission.

4.4.1 The M^2-Factor

The diameter of an gaussian beam is measured at the lateral points where the maximum intensity has dropped by a factor of e^2. This is for the lowest order Laguerre-Gauß or Hermite-Gauß modes a width of $2w$. With growing distance from the light source the waist radius grows according to

$$w^2 = \frac{2v}{k}\left[1+\left(\frac{z-z_0}{v}\right)^2\right] = w_0^2\left[1+\left(\frac{\lambda(z-z_0)}{\pi\cdot w_0^2}\right)^2\right] \tag{4.76}$$

as we know from (4.64). If the beam consists of higher order gaussian modes the measure of the e^{-2}-distance gives no longer the beam waist radius w because the Hermite or Laguerre polynomials give an additional lateral dependence. A light beam is called to be ideal when its beam waist diameter grows as that of the lowest order gaussian mode. In reality there are higher order modes with relative little amplitudes present that widen the beam diameter. As a first approximation the measured radius $w_{\text{meas}} = M\cdot w$ is believed to be by a factor of M bigger than the ideal one such that

$$w_{\text{meas}}^2 = M^2\cdot w_0^2\left[1+\left(\frac{\lambda(z-z_0)}{\pi\cdot w_0^2}\right)^2\right] \quad . \tag{4.77}$$

From the measured data the beam waist radius $w_{0\text{meas}} = M\cdot w_0$ can be extracted. Due to the fact that M is an unknown parameter (4.77) is written in the alternative form

$$w_{\text{meas}}^2 = w_{0\text{meas}}^2\left[1+\left(M^2\,\frac{\lambda(z-z_0)}{\pi\cdot w_{0\text{meas}}^2}\right)^2\right] \tag{4.78}$$

and the factor M^2 inside the parenthesis can be easily determined by simply measuring the beam diameter at different positions z and approximating the measured data with the expected dependence (4.78). For $M^2 = 1$ the beam is called to be diffraction limited because the beam radius at any point on the ray axis grows according to (4.76). This is the smallest diameter that a beam can have. Beams with arbitrary intensity distribution can be expanded into gaussian modes. As can be shown all higher order gaussian modes w_{meas} grows faster than (4.76) would suggest giving a $M^2 > 1$.
When we compare (4.76) and (4.78) we find nearly the same $z-$dependence. We define an opening angle $\tan\{\frac{\theta}{2}\} = \frac{w}{z-z_0}$ for big $z - z_0$ as sketched in figure 4.12.
In this case the approximation $\tan\{\frac{\theta}{2}\} \approx \frac{\lambda}{\pi}\cdot\frac{1}{w_0}$ is valid. For a real beam follows an opening angel of $\tan\{\frac{\theta}{2}\} = \frac{w_{\text{meas}}}{z-z_0} \simeq M^2\cdot\frac{\lambda}{\pi}\cdot\frac{1}{w_{0\text{meas}}}$. Due to the fact that the theory is only

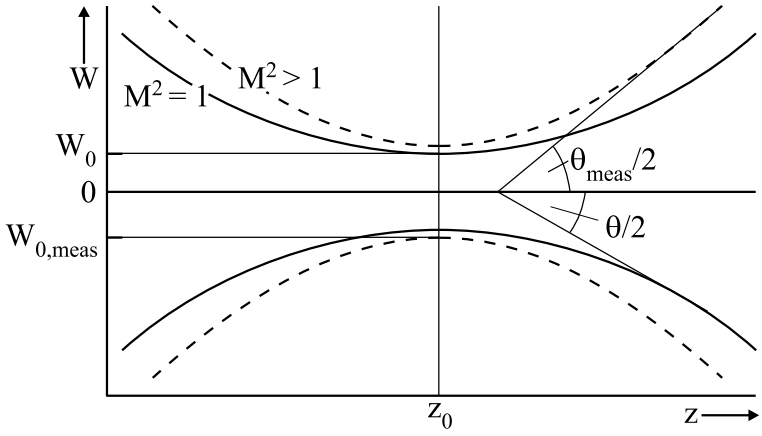

Figure 4.12: Beam width of real and ideal gaussian beams.

valid for paraxial beams the opening angles are small and we can further approximate $\tan\{\theta\} \simeq \theta$ and therefore

$$\frac{\theta}{2} \simeq M^2 \cdot \frac{\lambda}{\pi} \cdot \frac{1}{w_{0\mathrm{meas}}} \quad . \tag{4.79}$$

From (4.79) we directly see that beams with smaller beam waists exhibit bigger opening angels and vice versa. This tells us that the focal spot size of such a beam gets smaller when the focal length shortens at fixed lens diameter. The ratio of lens diameter and focal length is $\tan\{\varphi\} = \frac{d}{2f} \simeq \sin\{\varphi\} = \mathrm{NA}$. With growing numeric aperture NA of a lens the opening angle increases giving a smaller possible spot size.

4.4.2 Coherence Length

Until now we expected that an optical beam is coherent over arbitrary lengths or times. This means that portions taken at any position give a interference pattern when superimposed as sketched in figure 4.13.

The distance d between the portions is equivalent to a time delay $\tau = \eta \frac{d}{c}$. As a measure for the coherence the modulation depth V_c [1] of the interference pattern intensity I is

[1] The modulation depth is also called contrast or visibility.

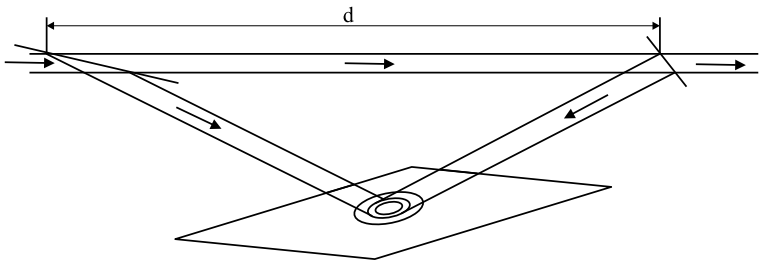

Figure 4.13: Interference between two portions of an optical beam. The portions are taken at distance d which is similar to time delay $\tau = \eta \frac{d}{c}$.

measured

$$V_{\mathrm{c}} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{4.80}$$

where $I_{\min}$ and $I_{\max}$ denote minimal and maximal intensities. Completely incoherent beams show no interference pattern $I_{\min} = I_{\max}$ and are therefore described by $V = 0$. Ideal coherent beams have $I_{\min} = 0$ and therefore $V = 1$. The contrast decreases with growing distance between the interfering portions $V_{\mathrm{c}} = V_{\mathrm{c}}\{d\}$. An often taken definition for the coherence length ℓ_{c} is the distance where V_{c} decrease to $V_{\mathrm{c}}\{\ell_{\mathrm{c}}\} = \frac{1}{e}$. A mathematical description can be given when the electrical field of the beam is regarded as $E = (E_0 + e\{t\}) \cdot \exp\{-i\,\omega t\}$ where $e\{t\}$ denotes a small disturbance like noise and E_0 can be taken as real amplitude. It can be stated here that e is always present and gives rise to a finite line width in the spectrum of the beam. The intensity follows from the time averaged quadratic field $I = |E|^2$ such that the interference pattern follows from

$$\begin{aligned}
I &= \overline{|E\{t\} + E\{t+\tau\}|^2} = \overline{(E\{t\} + E\{t+\tau\})(E\{t\} + E\{t+\tau\})^*} \\
&= \overline{|E\{t\}|^2} + \overline{|E\{t+\tau\}|^2} + 2\,\mathrm{Re}\left\{\overline{E\{t\} \cdot E^*\{t+\tau\}}\right\} \\
&= 2\,I_0 + 2\,\mathrm{Re}\left\{E_0^2 + E_0 \cdot \overline{(e\{t\} + e^*\{t+\tau\})} \cdot \exp\{i\,\omega\tau\}\right\} \\
&= 2\,I_0 + 2\,\mathrm{Re}\left\{I_0 + \sqrt{I_0} \cdot \overline{(e\{t\} + e^*\{t+\tau\})} \cdot \exp\{i\,\omega\tau\}\right\} \\
&= 2\,I_0\,(1 + \mathrm{Re}\,\{1 + \varphi_{\mathrm{e}}\{\tau\} \cdot \exp\{i\,\omega\tau\}\}) \\
&= 2\,I_0\,(1 + \mathrm{Re}\,\{\psi_{\mathrm{e}}\{\tau\} \cdot \exp\{i\,\omega\tau\}\})
\end{aligned} \tag{4.81}$$

with the autocorrelation function $\varphi_e\{\tau\} = \frac{1}{e_0}\overline{(e\{t\} + e^*\{t+\tau\})}$ of the distortion amplitude. The noise intensity $I_N = \overline{(e\{t\}e^*\{t+\tau\})}$ was neglected in 4.81 Without any distortion e the interference pattern would oscillate with time delay τ between zero and $4\,I_0$ such that $V = 1$ results. Real sources exhibit stochastic noise e and for infinite time delay $\psi\{\tau\}$ vanishes such that only the intensity of the single portions are added without any interference resulting into $V_c = 1$.

4.4.3 Solitons

In section 3.6.4 we already considered anisotropic media and especially the linear electrooptic effect. The next higher order of field dependence is the quadratic electrooptic effect which is present for example in quartz fibers. This effect gives a refractive index that is dependent on the intensity of the wave. Under this circumstances a wave of sufficient power can build a graded index in such a way that the chromatic dispersion can be completely compensated. An equivalent picture for this effect is given in figure 4.14. Three jogging man run with different speeds on a plain area. But if they are running on a rubber mate the mate bends and the fast one always has to run up a hill speeding down, the middle one runs as on a plain area with the same speed, and the slow one runs down a hill therefore speeding up.

Under this conditions the group stays together over arbitrary distances. If only one of them is missing (the intensity is lower) or one more is accompanying them or the initial distance changes the bending of the mate changes and the group spreads out in a moment. The same holds for an optical pulse of certain power and length. A pulse that travels dispersion free over arbitrary distances is called a soliton. For the existence of a soliton a certain material dependent pulse-width power product has to be met.

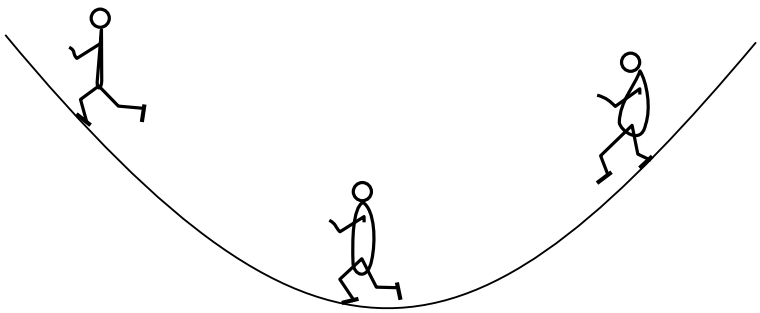

Figure 4.14: A group of three persons runs on a nonlinear ground. Their speed difference is compensated by the ground form such that they stay together.

Chapter 5

Modulation of Light

A pure light beam as emitted from a source carries no information and is only useful for illumination. To provide information transport the light has to be modulated. This can be achieved by direct modulation of the source which is considered in a later chapter. Here we want to concentrate on the modulation of a light beam as emitted from the source. Modulation in this context shall be understood as a change in the irradiance or direction when the light passes a modulator.

Such a modulation can be done in the simplest way mechanically with a shutter which switches the beam on and off. Another modulation scheme is a passive modulation with dyes where the transparency can be varied by a control beam. This modulation schemes cannot be integrated monolithically with lasers and photodetectors. This can be done when electrooptic effects like the Pockels- and Kerr-effect are used. Both effects lead to a varying refractive index as a function of applied electric field. Also magneto-optic effects are discussed, the big problem is the integration of a magnet with sufficient field strength. With magneto-optic effects effective optical isolators and circulators can be realized. Another way to modulate light is the application of the acousto-optic effect where the refractive index is modulated by an acoustic wave. Another way to control the refractive index is to utilize the plasma effect which describes the dependence of the refractive index depends on free carrier density. As a last point in quantum wells special

effects usable for modulation i.e. the quantum confined Stark effect are available. Also the nonlinear behaviour of materials can be used for example in saturable absorbers.

5.1 Mechanical Modulation

Mechanical modulation is the most popular modulation type. It can be achieved by cutting the light beam with an opaque or semitransparent shutter or deflecting the beam with a mirror or combinations of them. In both cases the light intensity on a receiver is changed. Applications are found in high speed cameras where a circular assembly of planar mirrors is rotated with high speed and the film is transported parallel to the axis of rotation. With this configuration the film is illuminated only for very short times. An equal setup with rotating mirrors was first used to generate very short laser pulses in a gas laser. Here one mirror of the laser resonator was replaced by a rotating mirror. Only for a short time the resonator gets into resonance and laser light is emitted. In the mean time pumping produced a highly population inverted material such that the power emitted in the pulse becomes extremely high.

A new technique is the application in video projection systems or video beamers as they are called. Here the light is splited on an array of micromirrors and deflected into two positions. One position is an absorbing wall, the other one is a lens system. In the first case the regarded dot of the picture remains black in the other case the desired colour of light appears on the screen. The principle has been invented by Texas Instruments in 1977. They call the key device a Digital Micromirror Device (DMD). When it is filled with the surrounding electronics for projection it is called a Digital Light Processor (DLP). The function principle described above is sketched in figure 5.1.

Another currently discussed mechanical light modulator is the Grating Light Valve (GLV). Here the first order diffraction of a grating is used as sketched principally in figure 5.2. Under exitation the planar grating is deformed and the first order diffraction can pass a lens.

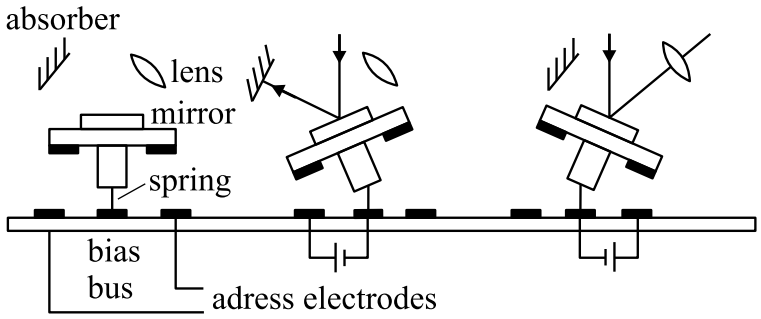

Figure 5.1: Functional principle of a DMD which is the core device in a DLP invented by Texas Instruments.

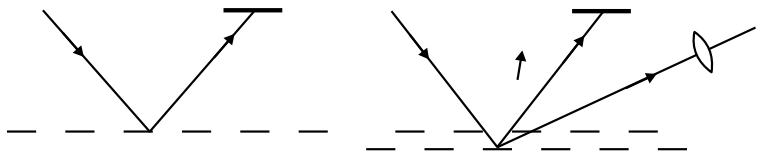

Figure 5.2: Grating light valve. Under exitation the grating is deformed and the first order diffraction beam passes a lens. Depending on the height of the deformation the colour of the light that satisfies the angular condition can be adjusted.

Only for a small spectral range the diffracted light meets the conditions to pass the lens and therefore by changing the deformation the used colour of light is determined.

5.2 Passive Modulation with Dyes

Optical dyes are used with low concentration in a solvent. When light passes the dye it is absorbed and generates free carriers. Under high excitation such dyes can act as an optical amplifier. In this case special lasers, so called dye lasers, can be set up. Another interesting characteristic is the fact that the absorbance changes with light intensity as sketched in figure 5.3. Under high exitation all possible states are occupied such that no more free carriers can be generated.

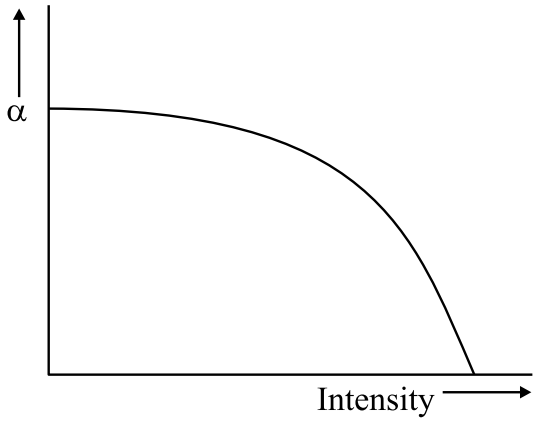

Figure 5.3: Absorption coefficient of a dye as function of incident light intensity. Above a certain intensity the dye becomes transparent such that additional light passes the modulator without any loss.

Additional light can pass through the dye without any change in intensity. Such a absorbing behaviour leads to the so called saturable absorbers. Applications are found again in laser sources that produce short pulses with high intensities. This sources run without any mechanical detuning of the resonator making them simpler to handle as such described in the preceding section.

5.3 Electro-Optic Effects

As already mentioned in section 3.6.4 the dielectric tensor can be field dependent. For the linear electro-optic effect a linear change of the inverted dielectric constant resulted.

As a next higher approximation a quadratic field dependence can be taken into account such that an outer field E results into

$$\Delta\left(\frac{1}{\eta^2}\right) = r \cdot E + p \cdot E^2 \quad . \tag{5.1}$$

Here we assumed again nonmagnetic media such that $\varepsilon = \eta^2$ applies. For the discussion in 3.6.4 we took $p \cdot E \ll r$ and found

$$\Delta\eta = \pm\tfrac{1}{2} \cdot \eta^3 \cdot r \cdot E \tag{5.2}$$

where the sign is dependent on the crystal and the light propagation with respect to the electric field. Two configurations for electrooptic modulators using the Pockels effect are common, i.e. the longitudinal and the transversal setup as sketched in figure 5.4. In both cases such modulators are also called Pockels cells.

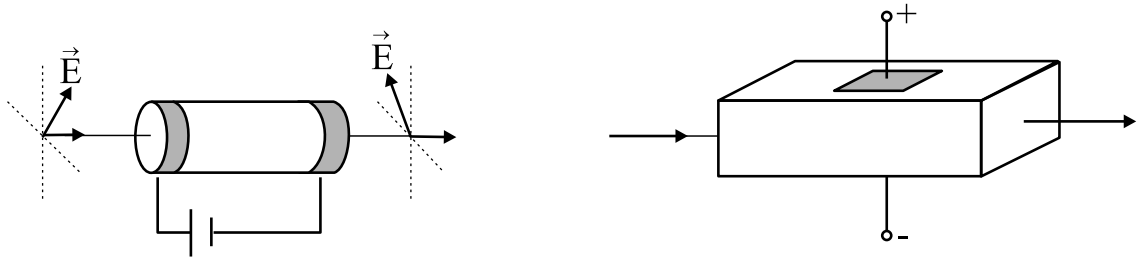

Figure 5.4: Longitudinal and transversal electrooptic cell. In the longitudinal Pockels cell the light travels parallel to the applied electric field. The direction of polarization changes due to the electric field induced birefringence. In the transversal Pockels cell the applied field is perpendicular to the propagation of light. The induced birefringence creates a phase shift only for waves that are perpendicular to the applied field polarized.

The quadratic field dependence in (5.1) is called electrooptic Kerr effect. When $r = 0$ applies which is true for most material (5.1) can be written in the form

$$\Delta\eta = K \cdot \lambda \cdot E^2 \tag{5.3}$$

where K denotes the Kerr constant. K is in the range of $10^{-12}\ \frac{\mathrm{m}}{\mathrm{V}^2}$ but can be much higher for ferro electric materials near the Curie temperature or in semiconductors including quantum wells near the band edge. Normally the applied electric field causes a molecular displacement which in turn gives rise to the changed refractive index. This molecular rearrangement takes some time which is found as a delay in Kerr modulator varying from some seconds in polar media like water to 10^{-11}s in nonpolar media. Due to the small Kerr constant high electric fields must be applied. In Fabry Perot resonators the light itself can reach this values giving rise to self modulation. Also intense optical pulses exhibit high electric fields thus enabling solitons to build up in quartz glass fibers.

5.4 Magneto-Optic Effect

The refractive index is not only subject to change by an electric field but also to a magnetic field in special materials. Like with the linear electrooptic effect the refractive index perpendicular to the applied magnetic field changes linear with the applied field. A linear polarized wave can be expanded into two circular polarized waves of similar amplitude but opposite rotation. The difference for the refractive indices of the left and right hand rotating waves is

$$\Delta\eta = \eta_{\mathrm{r}} - \eta_{\ell} = \frac{\lambda_0}{\pi} \cdot V\{\lambda_0\} \cdot B \tag{5.4}$$

with the wavelength dependent Verdet constant V which is for flint glass $V\{\lambda = 589\mathrm{nm}\} = 13.9 \cdot 10^{-3} \frac{1}{\mathrm{T\,mm}}$. When the light travels a distance L inside the material the linear polarization has turned by

$$\theta = V \cdot B \cdot L \quad . \tag{5.5}$$

This rotation of polarization was observed first by Faraday and therefore the effect is called Faraday effect. The main difference to the linear electrooptic effect is that the

rotation is in the same direction regardless if the wave travels parallel or antiparallel to the applied magnetic field. The Faraday effect is used in optical isolators. Here the Faraday rotator is placed between two polarizers that are rotated 45° against each other. The magnetic field is adjusted such that the forward traveling wave is rotated by 45° and therefore passes the output polarizer without losses. The backward traveling light is rotated by the same amount in the same direction such that it is 90° rotated against the input polarizer and blocked there.

If light is reflected from a magnetic birefringent material the polarization of the reflected wave can be rotated as well. This effect is the so called magnetic Kerr effect.

Both magnetooptic effects require big magnetic fields of some Tesla or big interaction length due to the small proportional constants. This is the reason why they are not integrable for integrated optic devices. The second reason is that the required materials cannot monolithically grown on the normally used III-V materials used for optoelectronics.

An application of the magnetic Kerr effect is found in optical storage systems. Writing is done by heating the material with a laser beam above the Curie temperature in the presence of a magnetic field. When the temperature decreases under the Curie point the magnetization is frozen. To read the information the laser is driven at lower power level and the change in polarization of the reflected beam gives an information about the stored orientation of magnetization.

5.5 Acousto-Optic Effect

The acousto optic effect is based on a refractive index change caused by mechanical strain. In this special case the strain is induced by an acoustic wave traveling through the medium generating a pattern of higher and lower refractive index regions as sketched in figure 5.5. This pattern can be regarded as a grating. An incident optical wave is diffracted at that grating.

For the explanation of the diffraction two different cases have to be regarded namely

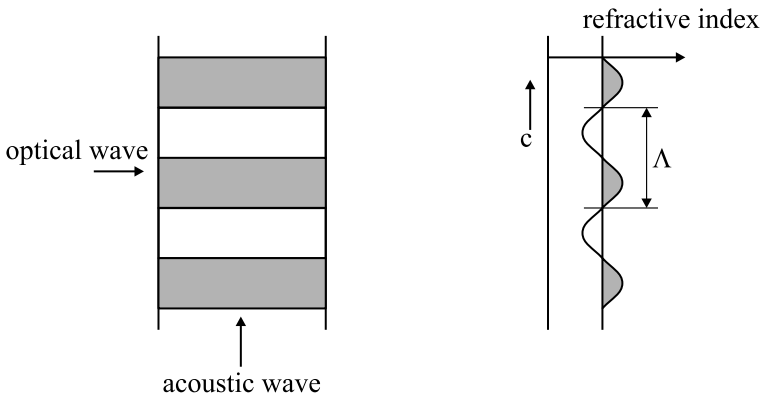

Figure 5.5: An acoustic wave produces a refractive index grating which in turn leads to diffraction of incident light.

the Raman- Nath regime and the Bragg regime. The difference is found in the length the optical beam interacts with the acoustic wave as sketched in figure 5.6. In the Raman- Nath regime the interaction length is so small that the diffracted light suffers no further redistribution before leaving the modulator. The diffraction occurs as at a simple plane grating to the law

$$m \cdot \lambda = \Lambda \cdot \sin\{\theta_m\} \tag{5.6}$$

where m denotes the order of diffraction. In the Bragg regime the interaction length is that long that the diffracted light itself is diffracted again. In this case the diffraction law follows $m \cdot \lambda = \Lambda \left(\sin\{\theta_\mathrm{i}\} + \sin\{\theta_\mathrm{d}\}\right)$ as is found at plane gratings. A rigorous treatment taking a sinusoidal index variation shows that the angle of incidence and of the diffracted light must equal each other and only the lowest order of diffraction $m = 1$ gives a contribution such that

$$\lambda = 2\,\Lambda \cdot \sin\{\theta_\mathrm{B}\} \tag{5.7}$$

follows. The discrete solution $\theta_\mathrm{i} = \theta_\mathrm{d} = \theta_\mathrm{B}$ is called the Bragg angle.

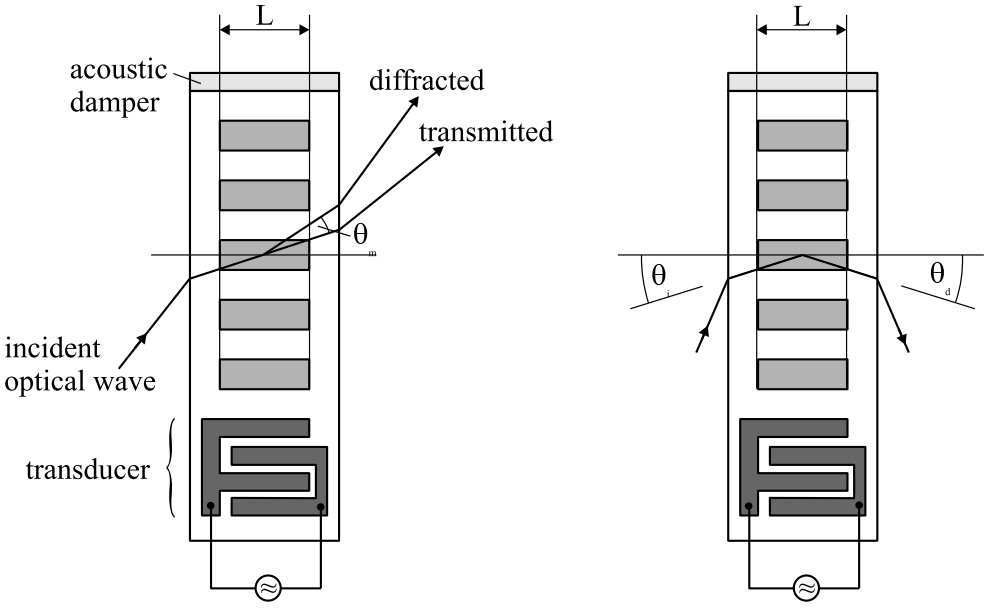

Figure 5.6: Raman- Nath and Bragg modulator. In the Raman- Nath modulator the interaction length between optical and acoustic wave is small. In a Bragg modulator L is long compared to the acoustic wavelength Λ.

The difference between the Raman- Nath and the Bragg diffraction lies in the amount of power that can be diffracted and is expressed by the diffraction efficiency η_{D} which can theoretically be 34% in the Raman- Nath regime and 100% in the Bragg-Regime. In the later case the efficiency follows $\eta_{\mathrm{D}} = \sin\left\{\pi \frac{\Delta\eta \cdot L}{\lambda \cdot \cos\{\theta_{\mathrm{B}}\}}\right\}$ where $\Delta\eta$ is the maximum index change produced by the acoustic wave.

5.6 Quantum-Well Effects

In bulk semiconductors free electrons and holes can build bounded pairs called excitons. The binding lowers their total energy by some meV such that their energy levels are found inside the bandgap of the semiconductor. With incident light they can be separated again under absorption of energies that are smaller than the bandgap energy. In quantum wells the excitons haver bigger binding energies and therefore exist at room temperature. Under an electric field the band edges are tilted and the electron and hole wavefunctions are separated spatially as sketched in figure 5.7. With the bigger distance their binding energy is lowered and therefore the absorption disappears under application of the electric field. The effect can be some 20meV at $100\,\frac{\mathrm{kV}}{\mathrm{cm}}$ field and is called quantum confined Stark effect (QCSE). With a QCSE modulator the absorption of light is controlled by the applied voltage.

5.7 Plasmaeffect of Free Carriers

In gas lasers the emitted wavelength changes when higher pumping levels are applied even under constant operation temperatures. A closer look reveales a linear dependence between refractive index of the plasma and the free carriers in the gas plasma leading to the name free carrier plasma effect. The effect can be explained with a resonance of free carriers leading to increased absorption with increasing carrier density. This not only true for gas plasma but for any given material. It is well known that a resonance not only affects the amplitude but as well the phase. This can be attributed to a change of

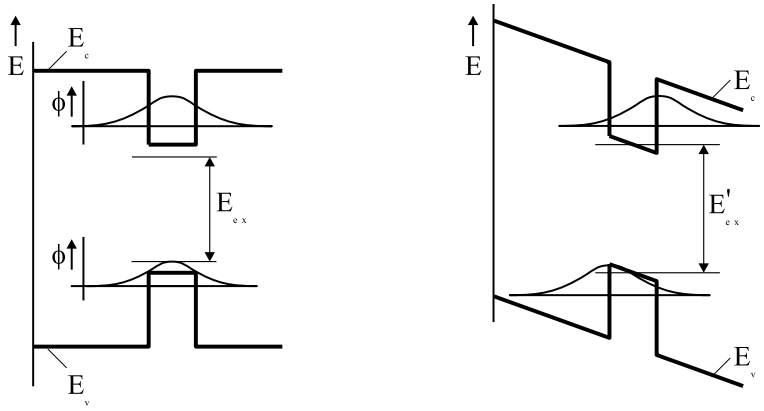

Figure 5.7: Under application of an electric field the bandedges are tilted separating free electrons and holes and lowering the binding energies of excitons. This leads to reduced absorption at λ_{ex}.

the dielectric constant and therefore or the refractive index. With increasing resonance effects the dielectric constant change increases. The spectral dependence of absorption and phase constant is sketched in figure 5.8. Far away from the resonance frequency the absorption as well as its change are negligible. But the change in refractive index that is attributed to the phase is clearly measureable.

Figure 5.8: Absorption and phase as a function of frequency. With increasing resonance amplitude the phase above resonance frequency is decreased.

oe2f4801.eps

Mathematically the correspondence between real and imaginary part of the dielectric constant is described by the so called Kramers Kronig relation (see appendix D) which can also be used for the refractive index. As a result in semiconductors a refractive index change of

$$\Delta\eta/\Delta n \simeq -10^{-3}/\left(10^{18}\mathrm{cm}^{-3}\right)$$

is found. The negative sign can be understood in the way that the carrier resonance is far below the optical frequency.

5.8 Thermooptic Effect

Not only mechanical properties are dependent on temperature. Also the dielectric constant and therefore the refractive index exhibits a strong temperature behaviour. With increasing temperature the lattice vibrations increase leading to a increased refractive index. The temperature coefficient in the typical semiconductors for optical components is about

$$\Delta\eta/\Delta T = 3 - 4 \cdot 10^{-4}/\mathrm{K}.$$

5.9 Non-linear Optics

With the invention of lasers non-linear optics have become important due to the high power densities and fields the laser beam can exhibit. For comparison sunlight has a power density of $20\,\frac{\mathrm{W}}{\mathrm{m}^2}$ and electric field of about $100\,\frac{\mathrm{V}}{\mathrm{m}}$ where a 1 mW laser beam of about $1\,\mu$m diameter shows $10^9\,\frac{\mathrm{W}}{\mathrm{m}^2}$ power density and a field of $1069\,\frac{\mathrm{kV}}{\mathrm{cm}}$. This is not entirely negligible compared to the inner field of an crystal which is of the magnitude of $10^{11}\,\frac{\mathrm{V}}{\mathrm{m}}$. Large non-linear effects have been observed already for power densities of $10^5\,\frac{\mathrm{V}}{\mathrm{m}} = 1\,\frac{\mathrm{kV}}{\mathrm{cm}}$ in semiconductors due to free carrier effects.

The optical nonlinearity is caused by an oscillation of the nuclei in the solid the light travels through. This oscillation in turn gives rise to emission of radiation as follows

from simple electrodynamics. If the oscillation amplitude is small only radiation at the same frequency as the incident light occurs. But for stronger exitation the mechanical oscillation of the nuclei becomes nonlinear and therefore their contribution to the total radiated field is nonlinear. The resulting polarization can be written as

$$P = \varepsilon_0 \left(\chi E + \chi_2 E^2 + \chi_3 E^3 + \ldots \right) \tag{5.8}$$

where χ is the linear susceptibility and χ_2, χ_3, ... are the nonlinear electrooptic coefficients. If we only assume $\chi_2 \cdot E^2$ to be of remarkable strength it is well known that in such a medium the incident wave of frequency ω generates a wave of frequency $2\,\omega$. Energy conservation requires in the photon model two photons of frequency ω to produce one of energy $2\,\omega$ due to $\hbar\omega + \hbar\omega = \hbar\,(2\,\omega)$. Without any further interaction with the solid the momentum is conserved as well resulting into $\hbar k_1 + \hbar k_1 = \hbar k_2$ where $k_1 = \frac{\omega \cdot \eta_1}{c}$ and $k_2 = \frac{2\,\omega \cdot \eta_1}{c}$ are the wavenumbers at the both frequencies respectively and $\eta_1 = \eta\{\omega\}$ and $\eta_2 = \eta\{2\,\omega\}$ are the corresponding refractive indices. From the momentum corresponding follows $\eta_1 = \eta_2$. If this so called phase matching condition is not met no efficient second harmonic generation can be reached. Equation (5.8) is only an expansion of the resulting polarization into a series of the exiting field implying that only harmonic oscillation can occur. In practice also nonharmonic fields can be produced. Energy and momentum conservation require

$$\begin{aligned} \omega_3 &\rightleftharpoons \omega_1 + \omega_2 \\ k_3 &= k_1 + k_2 \quad . \end{aligned} \tag{5.9}$$

In (5.9) nothing is said about the sign of the frequencies such that under exitation with ω_3 both lower and higher frequencies ω_1 can be produced. The resulting outputs are called signal (ω_1) and idler (ω_2). The effect is the same as was developed in electronic engineering for parametric amplifiers where the interaction between the exiting

frequency and the two other ones was adjusted by variation of some circuit parameter (typically the capacitance). The generation of new frequencies with non-linear optical media is therefore called parametric oscillation. On the other hand two signals at ω_1 and ω_2 can produce via (5.9) the new frequency ω_3. In this case the process is called sum or difference generation and the second and higher harmonic generation is a special case of sum generation. Parametric oscillation and sum and difference generation are sketched in figure 5.9.

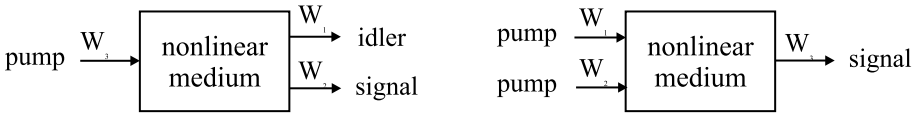

Figure 5.9: Parametric oscillation and sum and difference generation in non-linear optical media.

The sum generation in its special representation as second harmonic generation is used to extend the available emission range of lasers into the UV.

Chapter 6

Non Communication Applications of Optoelectronics

A lot of applications that are not used for communication technology are based on optical fibers where by external effects the properties of the fibers are changed which in turn lead to a change in the guided light. Another application is the light distribution as is found in automobiles, planes and other applications.

The non communication application of lasers is well known for welding and cutting where high optical energies are used. The next popular application is found in Compact Disc (CD) and Digital Versatile Disc (DVD) players and writers where the laser provides at least reading of the disc. The last familiar application is holography where the laser is used for three dimensional storage of pictures. Optical measurement technology is a more less unknown application of lasers beside alignment of constructions and sometimes illumination which we will not deal with here.

As a last non communication application of optoelectronics be have to look at optical displays which are briefly discussed at the end of this chapter.

6.1 Optical Fiber Applications

Optical fibers are used for illumination, in traffic signs, automobiles and various others. A special type of fiber is the light pipe which provides illumination of industrial buildings. Optical fiber sensors are favorable over their electronic and mechanical counterparts in critical atmospheres as are found on offshore platforms, in turbines, and others due to low energy level transported and the fact that quartz or plastic fibers withstand aggressive media. There is a wide field of applications and we will concentrate only on a small selection here.

6.1.1 Optical-Fiber Sensors

Optical fiber sensors can be divided into two groups i.e. the external and internal sensors. In external sensors the fiber is only a transport medium for light and the measurable effect occurs outside the fiber. Typical applications are displacement, distance, and temperature measurements as depicted in figure 6.1.

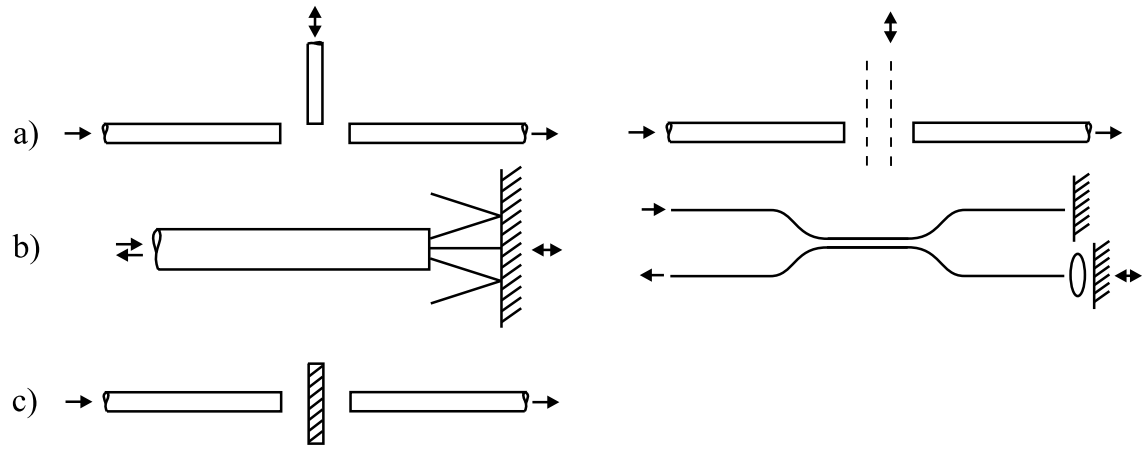

Figure 6.1: External fiber sensors: (a) displacement sensor, (b) distance sensor, (c) temperature sensor.

In the displacement Sensor (a) the light transmission from one fiber to the other is modulated ether by a movable shutter or a grating pair with a fixed and a movable

grating. The latter leads to a higher displacement sensitivity. Distance measurement (b) employs the divergent output beam of a fiber. The portion coupled back into the fiber after reflection depends on the distance between reflector and fiber. Higher sensitivity can be obtained with an interferometric setup where phase differences are measured giving resolution lower than the wavelength used for measurement. The temperature sensor (c) works with light transmission at the band edge of a semiconductor. With increasing temperature the bandgap shrinks and therefore the absorption at a given wavelength increases.

For internal fiber sensors the fiber itself is employed as the sensing medium. Due to the fact that the properties of a fiber can only be measured by relatively complicated and expensive techniques internal fiber sensors are not very common. The only important is a speckle sensor. In a multi mode fiber exited by a laser the modes gather different phases along their way through the fiber. With interference at the fiber end a spotty light distribution called speckle pattern can be observed. Bending of the fiber at any point changes the pattern. The change is not very predictable but can be used to determine if there is an additional bending or not. Such sensors can therefore be employed for example in traffic density measurements where a fiber buried in the road is bended by the cars rolling over.

A very special fiber optic sensor is the optical gyroscope which employs the Sagnac effect as depicted in figure 6.2.

The clockwise traveling wave needs time $t_\mathrm{c} = \frac{2\pi RN}{c+\Omega R}$ and the counterclockwise traveling one $t_\mathrm{r} = \frac{2\pi RN}{c-\Omega R}$ in a coil with N turns.

6.1.2 Light Guiding Fibers

Light guiding fibers are used since 1920 for illumination. At that time they consisted of polished acryl class and exhibited high losses due to surface roughness. 1953 a fiber with core and cladding was suggested and led to increased performance. Nearly all light guiding fibers are multimode and normally bundles of fibers called light guides are used.

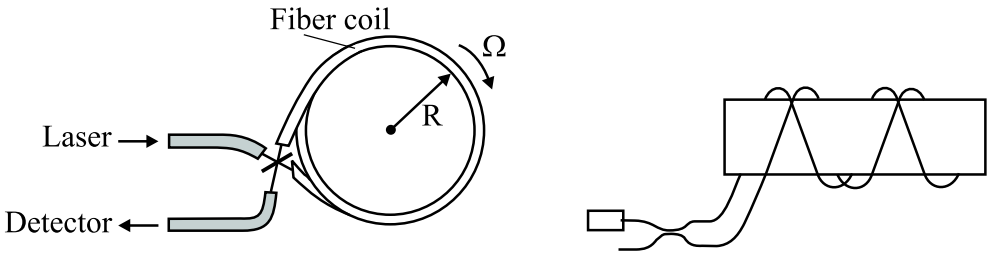

Figure 6.2: Fiber optic gyroscope. Rotation of the coils leads to a phase difference between the left hand and the right hand circulating waves.

The advantage is that they can provide illumination in hazardous environments where the light source can be placed a part from it. Applications of light guiding fibers are found in traffic signals and illumination of swimming pools etc. as sketched in figure 6.3.

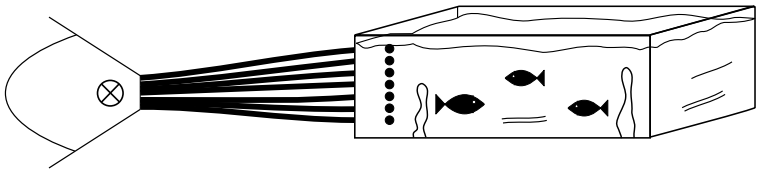

Figure 6.3: Light guiding fiber for illumination. Especially in hazardous environments are when lamp heating is a problem the light guiding fibers are favorable. Another application is found for positions where exchange of the lamps is difficult as in swimming pools.

When illumination shall occur over long distances one could argue that fibers with rough surface can be used. But they have the disadvantage that the out coupled light decreases the more the longer the fiber is when the out coupling is constant. This is because a constant fraction of the intensity present is scattered leaving a smaller intensity for the next out coupling. Mathematically this can be described by

$$\frac{\partial}{\partial z} I = -\alpha I$$

where $I' = \frac{\partial}{\partial z} I$ is the fraction of light per unit length which is out coupled. With constant α (6.1.2) gives the decreasing intensity

$$I = I_0 \exp\{-\alpha z\} \quad .$$

For fixed length of the fiber a constant out coupling can be achieved with z- dependent α such that αI =const follows. In this case we find in the guide

$$I = I_0 \frac{L - z}{L}$$

under assumption that all light has been extracted over length L resulting into a hyperbolic extraction coefficient

$$\alpha = \alpha_0 \frac{L}{L - z} \quad .$$

This kind of extraction is not easy to be produced especially at the end where a grows infinite. A possible way out is to take constant low extraction α such that the light coupled out at $z = L$ is nearly the same as at $z = 0$. The ratio is $\exp\{-\alpha L\}$. In this ease nearly all light is leaving the fiber at the end with no illumination effect resulting into poor performance. When at each end a mirror is placed as sketched in figure 6.4 the light travels back and forth again until all light is coupled out. In this case the light extraction at a given point z results from the forward traveling part

$$I'_{\mathrm{F}} = \sum_{n=0}^{\infty} \alpha I_{\mathrm{F}n} \exp\{-\alpha z\}$$

and the backward traveling part

$$I'_{\mathrm{R}} = \sum_{n=0}^{\infty} \alpha I_{\mathrm{R}n} \exp\{-\alpha (L - z)\}$$

where

$$I_{\mathrm{F}n} = I_0 \exp\{-2n\alpha L\}$$

and

$$I_{\mathrm{R}n} = I_0 \exp\{-(2n+1)\alpha L\}$$

are the amplitudes at the ends after n roundtrips such that the total light extraction is

$$\begin{aligned} I' &= I'_{\mathrm{F}} + I'_{\mathrm{R}} \\ &= \alpha I_0 \left[\frac{\exp\{-\alpha z\}}{1-\exp\{-\alpha 2L\}} + \frac{\exp\{-\alpha(2L-z)\}}{1-\exp\{-\alpha 2L\}} \right] \\ &= \alpha I_0 \exp\{-\alpha z\} \frac{1+\exp\{-\alpha 2L\}}{1-\exp\{-\alpha 2L\}} \simeq I_0 \frac{1}{L} \exp\{-\alpha z\} \end{aligned} \tag{6.1}$$

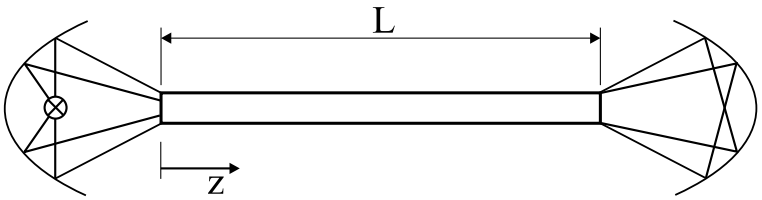

Figure 6.4: Illumination fiber setup. The transverse light extraction is $\exp\{-\alpha z\}$ for small extracting coefficients when the mirrors couple the light emitting from the ends back into the fiber.

The principle of transverse illuminating fibers has been extended to high power applications where the lamp heating detoriates the fiber. The fiber damaging can be overcome when the fiber is replaced by a tube which guides the light. This are so called light pipes or prism light guides are installed in high buildings as trade halls. The advantage is that exchange of the light bulb can be done from save maintenance areas. Illumination lengths of 20 m are state of the art. The principle of prism light guides is sketched in figure 6.5.

Figure 6.5: Prism light guide for high illumination intensities. The light is reflected inside the tube. The sawtooth prisms at the walls reflect up to 99 % of the light and couple out the rest. Illumination lengths of 20 m and more are obtainable.

6.1.3 Image fibers

In endoscopy an image for example of the stomach is taken by image fibers. Their operation principle is that the picture is a combination of a lot of pixels each represented by a fiber end as sketched in figure 6.6.

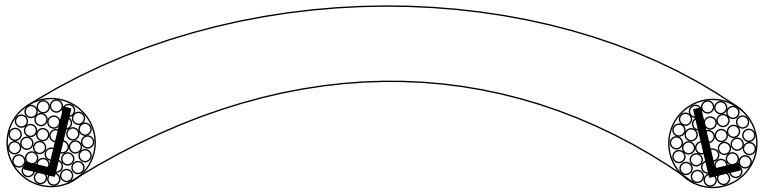

Figure 6.6: Image fiber for flexible endoscopy.

The trick is to place the fiber ends on each end of the bundle into the same position. To achieve high pixel density for good resolution the fiber must be thin. With decreasing cladding thickness the pixel density and therefore the resolution increases. But on the other hand at small cladding thickness lateral coupling occurs and the resolution decreases again. Another application with very coarse resolution is found in traffic signs where the desired picture is generated by a fiber bundle as sketched in figure 6.7. The pictogram can be exchanged and light may be switched on and off giving a high grade of flexibility. Maintenance can be done in a safe maintenance area apart from the road.

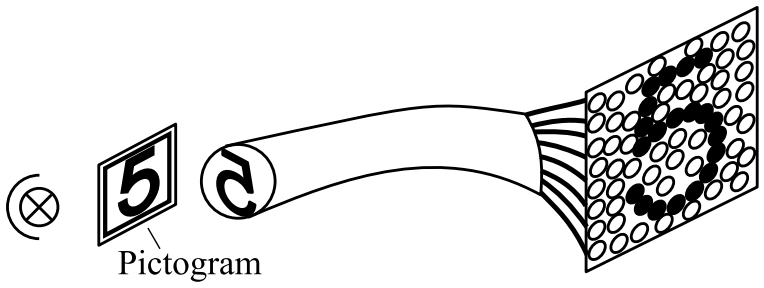

Figure 6.7: Application of image fiber in traffic signs. The pictogram can be exchanged to produce a new traffic sign.

6.2 Laser Applications

Lasers are found in a big variety of applications. Here we will make an arbitrary division into applications in measurements, picture reproduction, high energy applications and data storage. There may be more applications that do not fit into this division and are omitted here.

6.2.1 Measurement Technology

When laser applications in measurements are discussed we can note that the most common are distance measurement, contour and bar code scanning, and leveling in conjunction with spirit levels. Leveling applications are found in a broad spectrum from building construction to pipeline laying. Here the small beam size over long distances is employed.

In bar code scanning the light is deflected by a modulator and the reflection from the label is read. A more sophisticated scanning applications comes in conjunction with distance measurements which enables contour recognition.

Distance measurement with lasers follows three different principles namely interferometric methods, beam modulation telemetry, and pulse echo technique. Interferometric distance measurements makes use of a Michelson interferometer as sketched in figure 6.8.

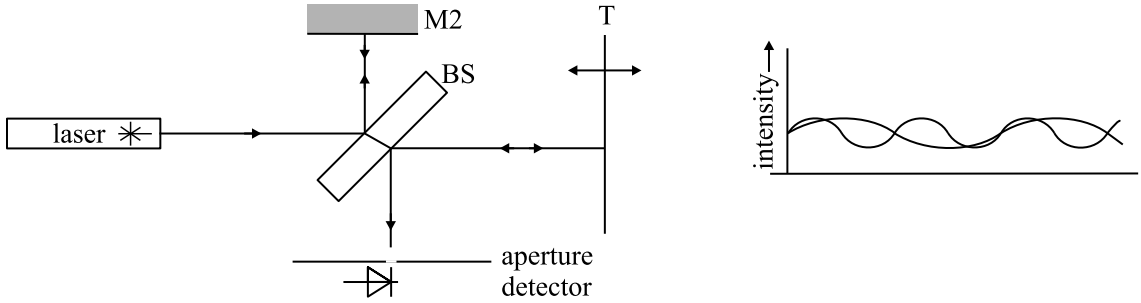

Figure 6.8: Michelson interferometer. The laser beam is splitted into a reference and a measuring beam. After traveling to a fixed mirror M and the target T respectively the returning light is recombined in the beam splitter and is incident on the screen producing interference fringes. A moving target results into moving fringes. The detector is illuminated through a small aperture thus resulting into a modulated intensity when the target is moved.

A moving target results into moving interference fringes on the screen. When the interference fringes are observed through a sufficient small aperture the fringe motion results into a modulation of the detected intensity. When the target and the mirror are parallel a circular fringe pattern results whereas in the other case straight line fringes occur. With a tilting angle of θ the maximum in a fringe pattern is found when the condition

$$p\lambda = 2\eta D \cos\{\theta\}$$

is met. Here p is an integer number, λ the measurement wavelength and D the difference of the optical ways to the target and the reference mirror. Moving the target results into other positions of the fringe maximum. Despite the fact that no absolute distance can be measured the movement can me measured very accurate. Up to 1 part into 10^6 can be achieved limited by changes in the refractive index of air due to pressure and temperature variations. Applications are found in machine tool control, length standard and for seismic and geodetic purposes. To prevent the reflected

light reentering the laser leading to unwanted laser modulation cube corner reflectors as sketched in figure 6.9 are used instead of plane mirrors. The reflected light beam travels always parallel to the incident one with a lateral offset.

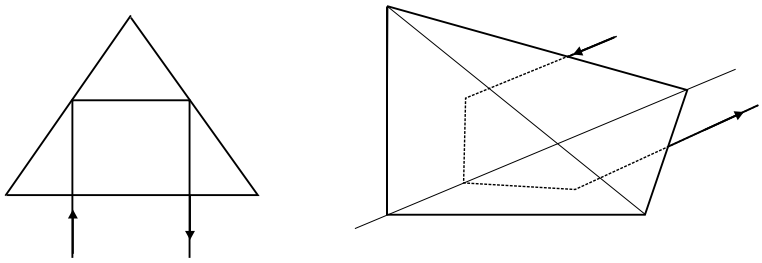

Figure 6.9: Prism and corner cube reflector. In a prism the incident and reflected beam are parallel in one transverse dimension, in the corner cube reflector the beams are always parallel.

Absolute distance measurement can be achieved by beam modulation telemetry and pulse echo technique. In both cases the laser beam is modulated ether by a sinusoidal signal or by pulses. In beam modulation telemetry the phases of the emitted signal and the reflected signal are compared. The phase change due to the time the light travels is

$$\phi = 2\pi f \frac{\eta}{c} 2L = (p+q)2\pi$$

where L is the target distance. Again p is an unknown integer and q a faction less than unity. For one modulation frequency q can be measured, with a set of modulation frequencies also p can be determined. Accuracies up to 1 mm in 1 km distance have been reached and are used in big constructions or for airborne instruments in band profiling and geodetic surveying. The measurement principle is sketched in figure 6.10. Pulse echo distance measurement follows nearly the same principle as the modulation telemetry. The difference is that the emitted beam is pulse modulated and the pulse delay of the reflected beam is measured. The delay is the time the light travels to the target and back giving the name 'time of flight'. For 100 m distance the delay $\Delta t = \frac{\eta}{c} 2L$

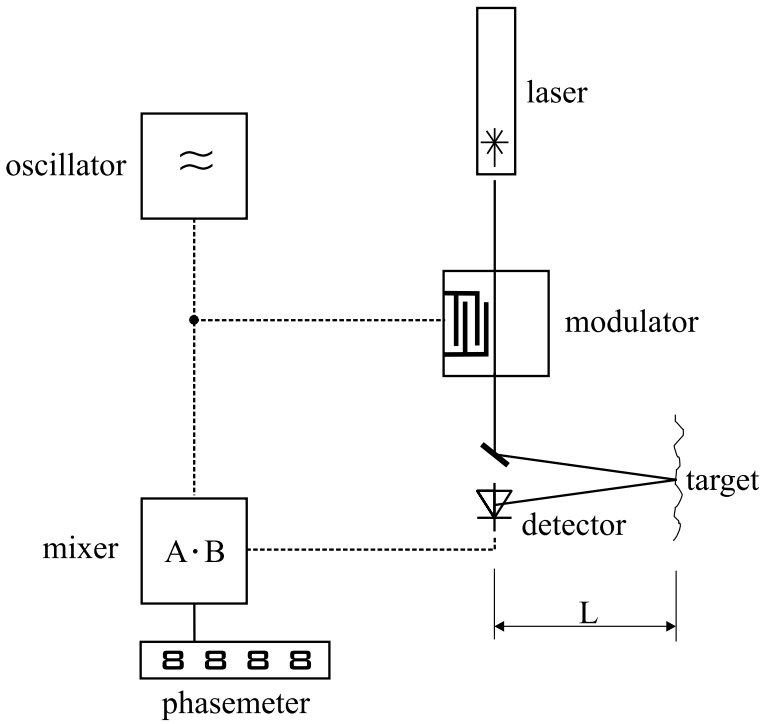

Figure 6.10: Beam modulation distance measurement. The phase difference between emitted and reflected signal is measured at different modulation frequencies giving the distance to the target.

is in the range of 0.6 μs. For 1 mm resolution the time delay must be measured with an accuracy of 6 ps. Accuracies of 1 mm in 10 km (10^{-7}) for military purposes are usual and can be extended as has been done in the lunar Apollo missions where the distance of the moon has been measured with an accuracy of $\pm$15 cm. The technique is also known as LIDAR (Light Detection And Ranging) and has been extended to atmospheric studies where the amount of back scattered light is dependent on the amount of pollutants such as CO_2 and SO_2.

6.2.2 Picture reproduction

Picture reproduction can be done by ether projection of a picture as is found in TV or holography. The latter one is widely used today for security applications as on money notes or labels.

6.2.2.1 Holography

The classical way of hologram fabrication is to take a sort of photograph of the desired object whereas today security holograms are calculated and embossed into a metal surface. To understand how a hologram works we first look at the classical setup as sketched in figure 6.11.

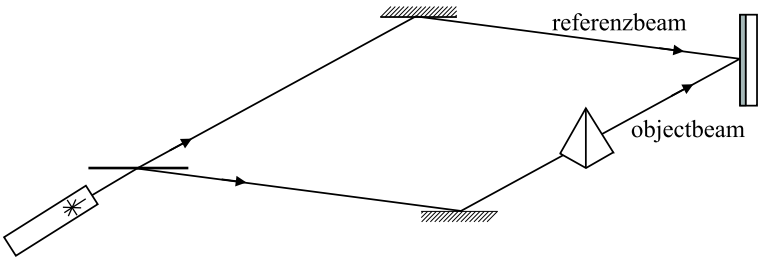

Figure 6.11: Holographic photography of an object

A laser beam is splitted into reference beam which is directly incident on the holographic plate and the object beam which is modulated by the object before interfering

with the reference beam on the plate. The plate is covered with a photosensitive film where the interference pattern is recorded ether as grey scale pattern like a classical photograph or as a refractive index pattern. The intensity of the interference pattern follows from the fields of the reference beam E_{O} and the object beam E_{R} after time averaging to

$$I_{\mathrm{rec}} = \langle E_{\mathrm{R}} E_{\mathrm{R}}^* \rangle + \langle E_{\mathrm{O}} E_{\mathrm{O}}^* \rangle + \langle E_{\mathrm{R}} E_{\mathrm{O}}^* \rangle + \langle E_{\mathrm{O}} E_{\mathrm{R}}^* \rangle$$

and is stored as transmissivity $T = a_1 + a_2 I_{\mathrm{rec}} t_{\mathrm{rec}}$ where a_1 and a_2 are film parameters and t_{rec} is the recording time. Reproduction of the hologram follows by direct illumination of the developed film as sketched in figure 6.12.

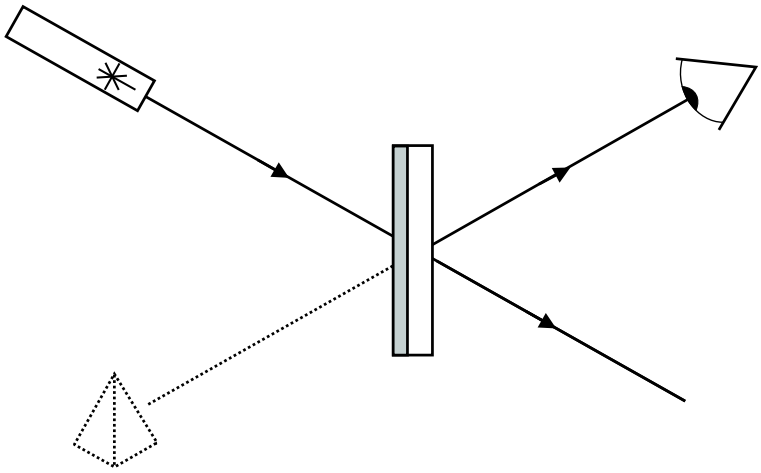

Figure 6.12: Reproduction of a classical hologram

In the observers eye a three dimensional picture is reconstructed. The intensity of the reconstructed hologram pattern follows (omitting the time average) from

$$I_{\mathrm{rec}} = a_1 + a_2 t_{\mathrm{rec}} (E_{\mathrm{R}} E_{\mathrm{R}}^* E_{\mathrm{R}} + E_{\mathrm{O}} E_{\mathrm{O}}^* E_{\mathrm{R}} + E_{\mathrm{R}} E_{\mathrm{O}}^* E_{\mathrm{R}} + E_{\mathrm{O}} E_{\mathrm{R}}^* E_{\mathrm{R}}) \quad .$$

The first three terms are found in the zero order (straight) transmission. The fourth term can be found in the reflected beam. When it is observed the object is inverted which means that front and backside are exchanged and the object seems to hide away.

The last term is the reconstructed object beam. When it is observed the object is virtually reconstructed in three dimensions. Here we discussed a so called amplitude hologram where part of the light is absorbed in the grey scale pattern and therefore the reconstructed object intensity is small. When a so called phase hologram is used where the recorded intensity pattern is stored as refractive index changes nearly no light is absorbed and a bright object is reconstructed. In both case the reconstruction must be made with the same wavelength as the recording otherways a changed object is reconstructed.

Applications of optically stored holograms are not covering the full potential of the method. The main application is found in non destructive testing of materials and components where the reconstructed picture of an object is superimposed with the actual one as sketched in figure 6.13. If the actual object is the same as the recorded no interference fringes occur. In case of deformations or refractive index changes fringes occur at the positions of distortion showing hair breaks or others.

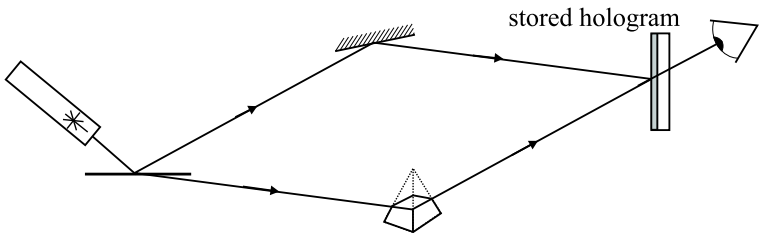

Figure 6.13: Real time holography. The picture of a recorded distorted object and the actual one are superimposed producing interference fringes at distorted positions of the object.

This kind of holography is called real time holography. Another special application is the recording of a vibrating object. Only non vibrating parts lead to reconstruction of the object thus indicating the nodes of the vibration. The technique is called time averaged holography and has been used for example to find the ideal position for the rear view mirror in automobiles that is glued to the front window. An application

under research is optical data storage in holograms which provide the advantages of high data read rates due to parallel stored information and robustness against scratches due to the fact that the bit information is spread over the entire hologram.

For some purposes we want the reconstruction by white light and in this case so called white light holography has to be employed. Recording and reconstruction of a white light hologram is sketched in figure 6.14.

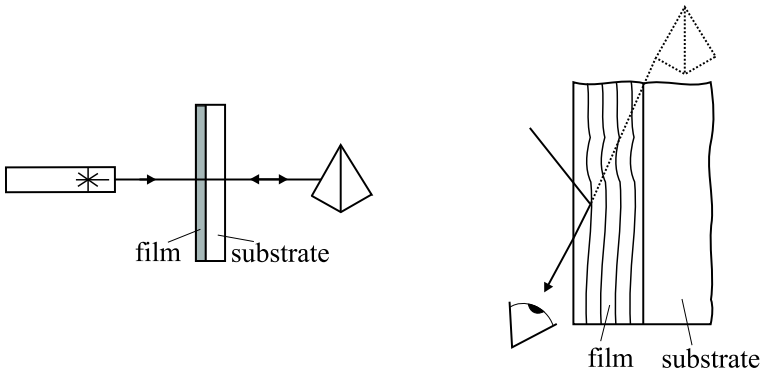

Figure 6.14: Recording and reconstruction of a white light hologram. Recording is done as usual with a laser. The interference between incident and reflected beam forms a specific index or transmission profile in the film. When the film is observed under shallow incidence the object is reconstructed even with white light.

Recording is done with a laser under normal incidence on the holographic film. The transmitted light is reflected from the object and interferes with the incident beam in the film producing a refractive index profile. Reconstruction occurs with white light which is scattered at the index profile. Only that wavelengths interfere constructively in the observers eye which reconstruct the object. Also a surface reflectivity profile can provide a similar phase change in white light. This profiles can be calculated and embossed into the surface of a metal. Applications are found on security labels or for toys.

6.2.2.2 LASER-TV

Picture projection as found in TV systems can be made not only with cathode ray tubes but as well with laser beams. They can provide mobile presentation on big displays. The problems today are to generate laser beams of the desired wavelength, esp. blue and green semiconductor lasers are needed to build a compact system. The second problem is that the reflected light from the screen which is reconstructed to the picture in the eyes takes slightly different ways and due to interference produces spotty dots instead of the desired homogeneous ones. The spots are called speckle and they are moving with any changes in the optical path length as produced by eye movement of flowing tear liquid on the eye surface. The effect can be minimized with a broad spectrum of the light beam or by pulsed light beams.

6.2.3 High Energy Applications

The high energy applications make use of concentrated high intensity laser beams that interact with a surface by spatial heating. Functions as welding, drilling, cutting, marking (scribing) and surface treatment are now known for several years. The problem is to deliver the high intensity laser beam to the workpiece without detoriating the involved optical elements. This is done by a beam expander and cooled mirrors as sketched in figure 6.15.

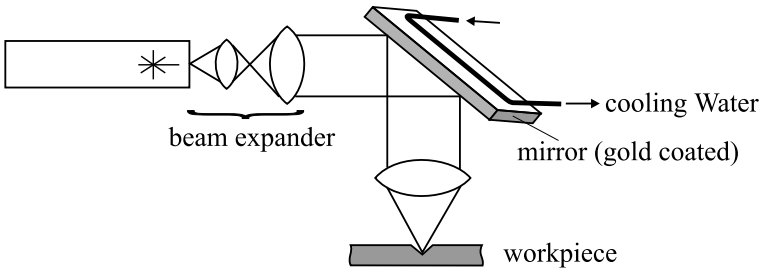

Figure 6.15: High energy laser processing system.

With the big beam diameter the power densities fall under certain limits and parasitic

absorption as occurring in the mirrors don't lead to distorting heating. For good heat transfer the mirrors are made of copper with a gold surface. The gold provides high reflectivity and low absorption and the copper heat transfer. In modern systems the beam delivery is also provided by optical fibers with core diameters up to $600\,\mu$m. Bigger diameters are not useful because the fiber looses its flexibility with increasing diameter. The main advantages of fiber delivery systems are the flexibility and the possibility to observe the treatment process of the material through the fiber. The fibers are made of silica for Nd:YAG lasers but cannot be used for CO_2 lasers that emit a $10,6\,\mu$m where silica exhibits high losses. For this purpose several other materials have been developed i.e. zirconium fluoride, sapphire, germanium, chalcogenide glasses and silver halide as well as hollow light guides where the laser is guided inside a tube.

6.2.3.1 Industrial applications

For a brief discussion we will limit the description of industrial applications here to welding and cutting. Normally the processes are further detailed by the nature of the laser beam to be continuous wave (CW) or pulsed and the repetition rate and duty cycle for pulsed applications. In a welding process two metal parts (similar or dissimilar metals) are placed together and the contact region is heated by a laser beam until the metals melt and fuse together. The main problem is the high reflectivity of the metals which make an accurate control of delivered energy necessary. The main advantages of laser technology over conventional welding techniques is the low distortion of the combined pieces due to low heat input, faster welding rates and higher quality welds together with smaller heating affected zones due to higher cooling rates. With very high light intensities the laser beam first vaporizes the metals forming a so called keyhole as sketched in figure 6.16. The light is trapped in the keyhole and melts the surrounding parts and fills later on the keyhole. With this technique weld depth of several tenth of millimeters have been achieved with multikilowatt laser beams in contrast to the one millimeter range for normal $500\,\mathrm{W}$ CO_2 laser beam welding.

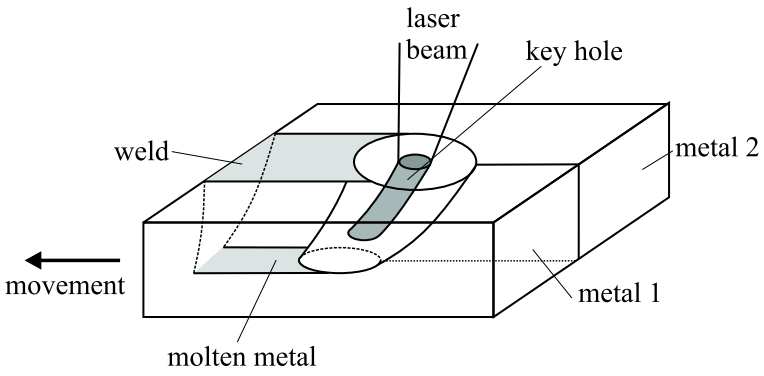

Figure 6.16: Formation of a 'keyhole' during high power laser beam welding.

The second application of high energy laser beams is cutting. The aim is to produce a heat effected zone (HAZ) where the material is vaporized as quick as possible. Under assistance of a forming gas which may prevent burning and provides cooling the molten parts are blown away thus producing the cutting kerf. Thicknesses of 5 mm in steal can be cut by 500 W CO_2 lasers. The kerf cross section is mainly dependent on the focal point of the laser beam with respect to the workpiece and the material to be cutted. For metals it is desirable to have the focus on the surface whereas for non metallic materials as paper or ceramic the spot should lie under the workpiece to provide straight cutting edges.

6.2.3.2 Medical applications

Medical high energy applications are found in surgery where the laser beam seals the cutted material and therefore provides low bleeding. The second advantage is the limited damage of tissue due to small cutting size. The cutting effect is based on light absorption in the water present which is high for CO_2 and Er:YAG laser light. Other applications are found in dentistry and ophthalmology. In the later one the treatment of detached retinas is a well known application. Also applications in the treatment of cancers are established.

6.2.4 Optical Storage

Optical data storage on Compact Discs (CD) and Digital Versatile Discs (DVD) is widely known. The principle is sketched in figure 6.17.

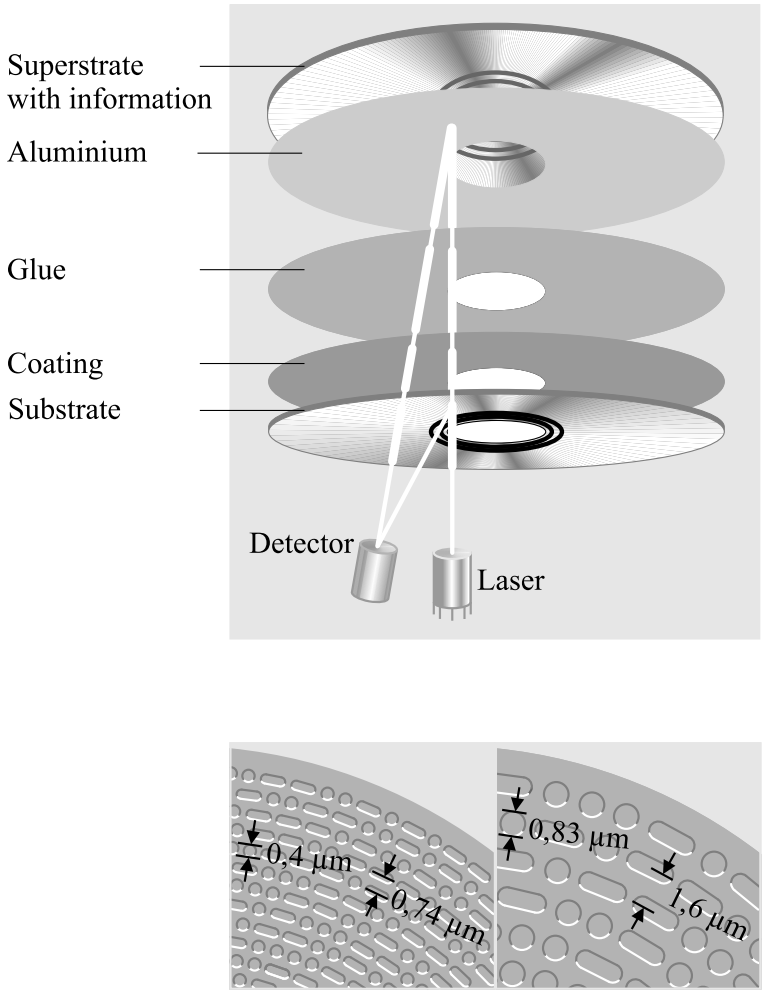

Figure 6.17: Optical data storage in a Compact Disc. The reflecting metal leads to a modulation in the reflected beam. The lower figures compare typical pitch and dot sizes in DVD (left) and CD (right).

A Laser beam is focused to the buried reflecting metal layer. The structure in the metal surface modulates the reflected laser beam. There are two types of DVD formats called HDD (High Density Disc: Sony, Philips) and SD (Super Density: Toshiba, Matsushita,

Thomson) and three types of access ROM (Read Only Memory), WORM (Write Once Read Many) and R/W (Read/Write). The SD-DVD has a six times higher storage capacity due to smaller track spacing and pitch size. The pitch sizes can be calculated from the diffraction limit of the gaussian beam waist to be $s = k\lambda/\text{NA}$ where λ is the operating wavelength, NA the numerical aperture of the optical system with the focal divergence angle NA= $\sin\{\alpha\}$ and k is a correction factor due to non ideal gaussian beams. For the SD the wavelength is reduced from 780 nm (CD) to 635 nm and the numerical aperture increased from 0,38 to 0,6 resulting into a two fold smaller pitch size thus enabling a four times higher storage capacity.

Higher storage capacity can be obtained with more information layers in the storage medium. Concepts for up to four layers are sketched in figure 6.18.

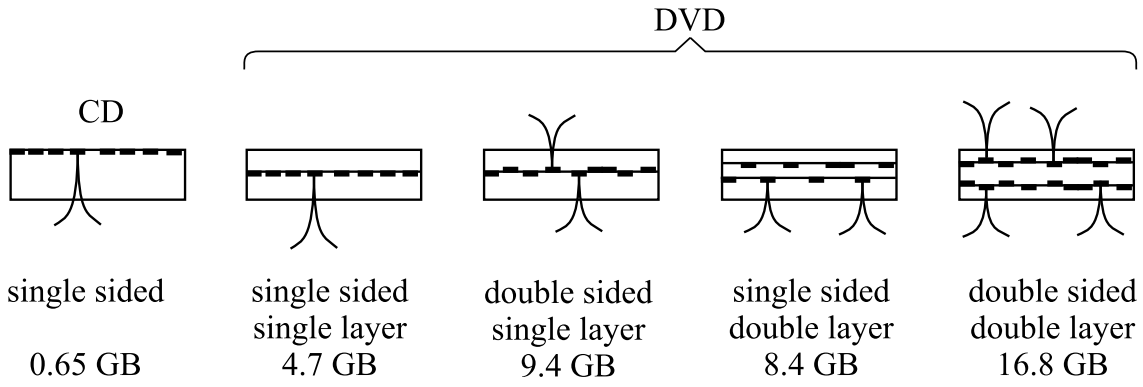

Figure 6.18: Comparison of standard CD-ROM and DVD storage concepts.

The structure of the information layer can ether be a surface deformation leading to beam deflection or the beam modulation is done by a change in refractive index (phase modulation), absorption or polarization. Possible writing principles are sketched in figure 6.19 for write once and rewritable discs. For the today's R/W CD's magneto optic films are used where the magneto optic Kerr effect is employed.

A new proposal for optical data storage is based on fluorescent pits. Figure 6.20 shows the cross section of a so called fluorescent multilayer disk (FMD). The disk consists of up to 50 storage layers. Each layer is structured with holes along the tracks

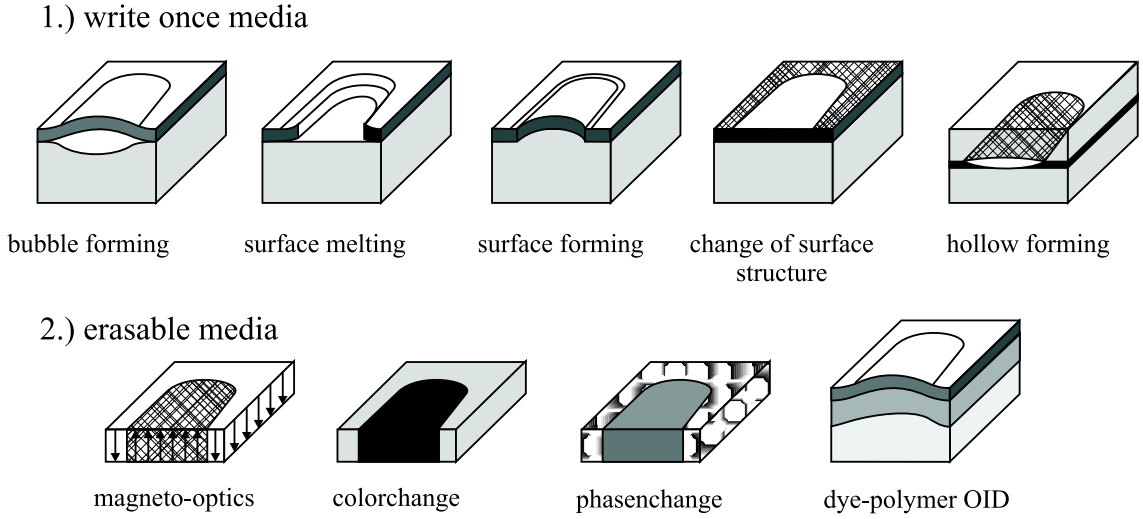

Figure 6.19: Writing principles for writable and erasable optical disc storage.

that are filled with fluorescent material. The pits are individually activated. Under laser exitation the activated pits (logical 1) are emitting light at a certain wavelength. Different wavelengths are chosen for each storage layer. With this technique 140 G byte, that is more than 3 hours film in real (non compressed) HDTV format, compared to 15 min. on a conventional two layer DVD-9 of the same size.

The idea of fluorescent storage goes back to the year 1994 when Dr. Joseph Malkin investigated photochemical storage media. Currently Constellation 3D (www.constellation3D.com) works on the realization of the FMD.

6.3 Displays

Optical displays can be classified into two broad categories (a) those that emit their own radiation (active devices) and (b) those that modulate radiation provided by other means (passive devices). In category (a) we will briefly discuss LED and plasma devices here and for category (b) liquid crystal displays (LCD) will be mentioned. The scope is to give an overview.

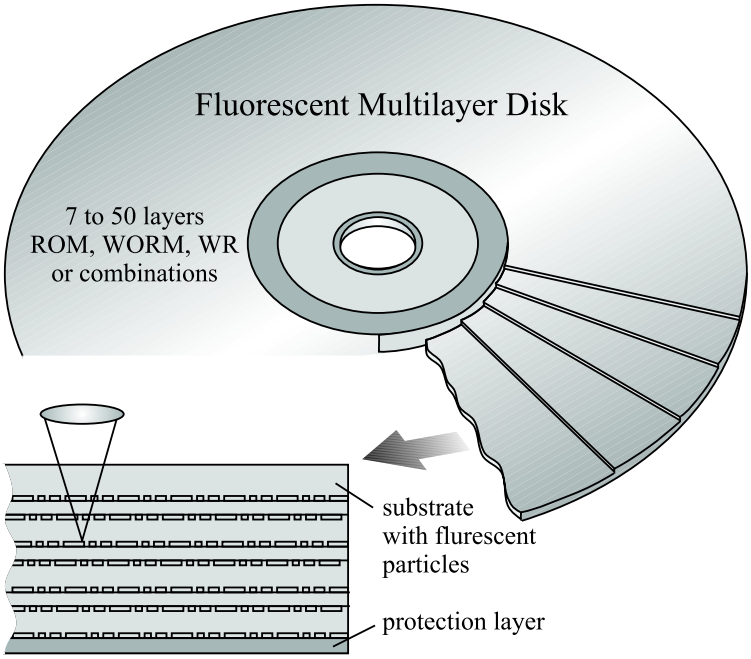

Figure 6.20: Cross section through a fluorescent multilayer disk (FMD). Under laser exitation activated pits (logical 1) of each layer emit fluorescent light. The dots of different layers are emitting at different wavelength.

6.3.1 LED Displays

LED displays have been widely used because of their good contrast. Nowadays they are more or less exchanged by LCD displays that can be tuned in spectral appearance by simple spectral filters. Nevertheless LCD Displays are found in some applications where high brightness and a big angle of view is desired. The constituting LED's are made of different materials for the desired colours (wavelength).

GaAs LEDs: GaAs is a direct bandgap semiconductor and therefore high efficient luminescence can be exited. The disadvantage is their emission in the near infrared spectral range at about 860 nm which makes them useful only for example for illumination purpose in night view goggles.

GaP LEDs: GaP is a indirect semiconductor with direct transition wavelength of 550 nm. The low radiation efficiency is enhanced by group V Doping with N or Bi replacing partly the phosphor. The dopands estate isoelectronic traps which assist radiative transitions. Codoping with Zn and O results into deeper traps and therefore higher emission wavelengths (red, 800 nm) and better efficiency compared to N doping where yellow emission (590 nm) is found.

$\mathbf{GaAs_{1-x}P_x}$ LEDs: With a phosphor content below 45 % GaAsP is a direct semiconductor. At $x = 0.4$ red emitting LED's are obtained (650 nm). Doping with N provides emission at 632 nm (orange) for the indirect 65 % material.

$\mathbf{Ga_{1-y}Al_yAs}$ LEDs: GaAlAs is direct with Al-content below 45 %. Deep red emission at 650 nm is obtained with 40 % aluminium content.

III-V Nitride (e.g. GaN, AlN) LEDs: This material provide an emission range from green to ultraviolet. A number of problems are associated for example lack of suitable substrate materials (lattice and thermal expansion matching) and difficulties in p-doping. With new growth techniques the problems are merely solved and this LEDs are found in traffic signs and giant displays.

InGaAs LEDs: InGaAs on GaAs substrate are commercially available with high brightness in the blue (450 nm) emission range. They have a 6 % In content and are

Zn doped.

SiC LEDs: Silicon carbide LEDs have been very promising for many years because their emission wavelength can be tuned by B, Al, Sc and Be doping to yellow, blue, green and red emission. Their main disadvantage is the low external quantum efficiency of 0,03 % which is 100 times lower than for the new InGaAs LEDs.

II-VI Semiconductor LEDs (e.g. ZnSe): ZnSe and related compounds provide direct blue and green emitting diodes. Their main disadvantage is that they are much softer than III-V semiconductors and degrade rapidly under normal atmosphere.

6.3.2 Plasma-Displays

Plasma displays rely on the effect that a glow is produced when an electrical discharge takes place in a gas. The energy of free electrons in the discharge is that high that they can excite the gas atoms such that they are raised into higher energy levels. They return to the ground state by radiating emission. This effect is used in plasma displays where the dots are gas filled transparent chambers that are exited by voltage application. The emitted radiation is normally in the ultra violet region and changed into the desired colour with fluorescent media on top of the chamber similar as is found in neon tubes. Typical chamber sizes are 100 μm spacing and a gas filling with $5 \cdot 10^4$ Pa. The discharge is started with a firing voltage of 150 V and held with 90 V. Plasma displays can be very bright and exhibit a big angle of view. Their disadvantage is the high voltage handling and the problem that the filling gas is lost over time.

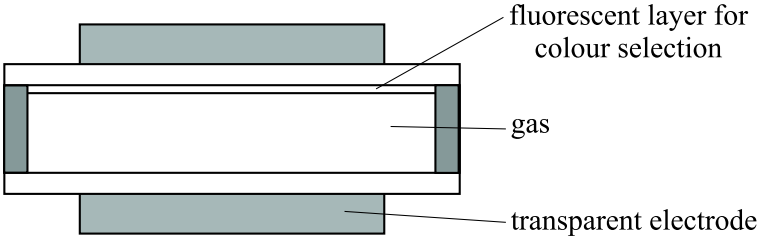

Figure 6.21: Plasma device chamber.

6.3.3 Liquid Crystal Displays

Liquid crystal displays (LCD) can be classified into the reflective and the transmissive type. Both utilize a cell that is formed between two glass plates with 10μ m spacing filled with liquid crystal material. This is an organic material in a special phase state that exists in a restricted temperature range. Below that range the material is a solid and above it changes to a clear liquid. The major characteristic of the material with its yellowish milky appearance is determined by its rod like shaped molecules. The molecules can have arbitrary orientation to each other and to the crystal surface. There are three types of ordering namely nematic, cholesteric and smectic. Only the first two are employed in LCD's because the smectic LCs have too long response times. In nematic ordering all molecules are parallel but can move independently as in a liquid. Cholesteric material can be regarded as consisting of an ensemble of nematic planes with each plane tilted against the others as sketched in figure 6.22.

When a nematic liquid crystal comes into contact with a surface the molecules tend to align perpendicular or parallel to the surface. The first case is called homeotropic the latter one homogeneous ordering. One of the most important characteristics of the liquid crystal is that it exhibits different dielectric constants ϵ_{p} and ϵ_{n} when an external field is aligned parallel or perpendicular to the molecule orientation. If $\epsilon_{\mathrm{p}} > \epsilon_{\mathrm{p}}$ we refer it to be positive material. Application of an electric field to positive material will cause the molecules to orient into field direction thus minimizing the stored energy. A homogeneous orientation can therefore be changed to a homeotropic one. This effect is sketched in figure 6.23 and occurs when a certain threshold filed E_{c} is exceeded.

Light that is traveling perpendicular to the molecule orientation is polarized parallel to the molecules. This effect is used for most LCD's where twisted nematic cells are employed. The cell is constructed such that the crystal orientation is 90^o tilted on opposite walls as sketched in figure 6.24. The polarization of light passing the cell is rotated 90^o. With applied electric field the 90^o rotation is switched off.

In a liquid crystal device a twisted nematic cell is combined with two polarizers on both

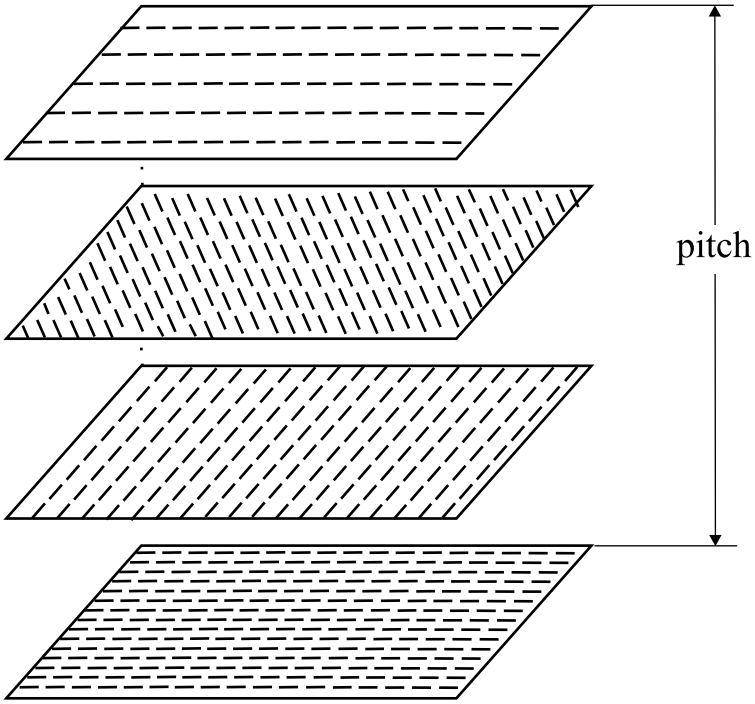

Figure 6.22: Cholesteric liquid crystal cross section. In a plane all molecules are oriented parallel to each other. In adjacent planes the orientation is tilted against each other. The same orientation is found in a certain distance called pitch.

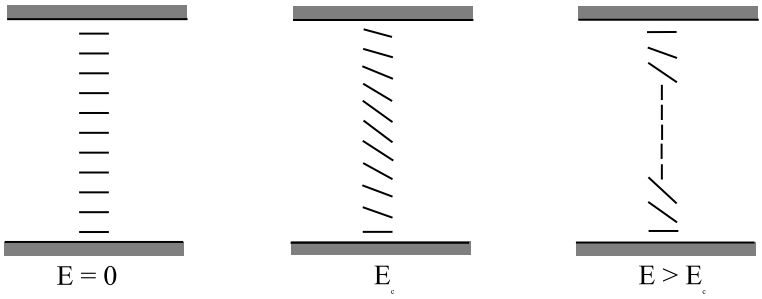

Figure 6.23: Under application of an external filed the homogeneous orientation is changed to homeotropic orientation.

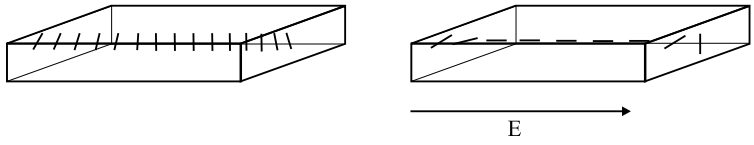

Figure 6.24: Twisted nematic cell. Light passing the cell undergoes a 90 0 polarization rotation. With applied field the polarization remains unchanged.

sides. In reflective devices additionally a mirror is placed on the backside as sketched in figure 6.25. Entering light passes the cell and undergoes the 90^o polarization rotation and passes the backside polarizer. After reflection the light leaves the front side. When the electric field is switched on the light is blocked by the backside polarizer due to unchanged polarization state and the display is off.

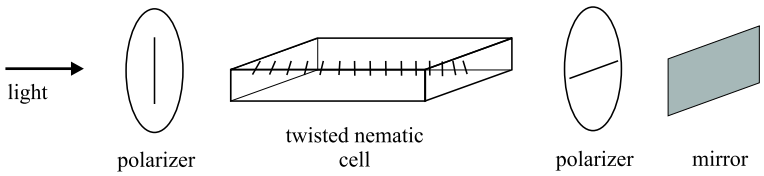

Figure 6.25: Typical LCD with twisted nematic cell.

The necessary applied voltage to turn off the LCD is about 3 V. Direct current operation tends to shorten the LCD lifetime owing to electrochemical reactions inside the cell and hence alternating current is invariably used. In transmissive devices the mirror is replaced by a light source. Colored devices are made by involving colour filters in the setup. The use of polarizers in the LCD limits their angle of view to about 45^o. Enhanced view angles are possible with supertwisted nematic cells where the rotation is 270^o. The switching to off position occurs at lower voltages but the switching time is much longer than for the 90^o device. Response time can be shortened by use of cholesteric materials.

Part II

Lasers, Detectors and Systems

Chapter 7

Lasers

This chapter covers a broad spectrum of laser relevant details beginning with some theoretical investigations on the way a laser works. This includes a closer look on noise of the laser light, the spectral range a single laser line covers, and the spectral distribution when a laser emits several 'colours'. The way how several laser lines can be coupled to produce ultra short pulses is discussed in the mode locking section followed by a brief discussion of a variety of laser classes which are currently used. Last but not least semiconductor heterostructure lasers are discussed in detail because of their big importance in today's research and applications.

7.1 Optical Feedback

The laser is an optical oscillator as pointed out already in chapter 2. The Fabry-Perot resonator discussed there consists of two well aligned plane mirrors reflecting the optical wave back and forth through the amplifying region. Laser action is found when the losses on the light roundtrip are just compensated by the amplification. Nevertheless the resonator concept with plane mirrors is not the best one when mirrors with finite extensions are investigated. Because of diffraction two plane-parallel mirrors of finite extent cannot maintain a perfectly collimated beam. The beam spreads out

and eventually reaches the edges of the mirrors giving additional losses that must be compensated by additional gain. These diffraction losses can be reduced by the use of concave or convex mirrors which form a so-called stable resonator. Figure 7.1 shows the stability diagram of optical resonators with infinite sized mirrors assuming that a diffraction-limited gaussian beam is supported.

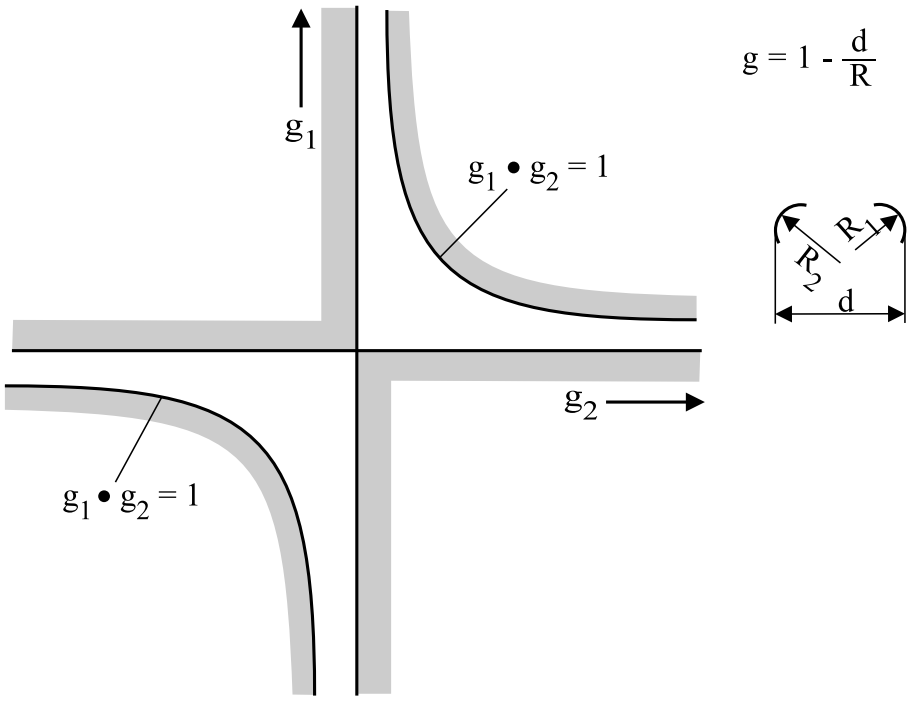

Figure 7.1: Stability diagram for an optical resonator supporting a gaussian beam. The nonshaded region satisfyiing $0 \leq g_1 g_2 \leq 1$ gives stable resonators meaning that a diffraction limited gaussian beam reproduces itself after one roundtrip. As can be seen the conventional plane-parallel (Fabry-Perot) resonator is metastable. Only infinite sized mirrors of unit reflectivity are considered in this figure.

As can be seen, a conventional Fabry-Perot resonator with plane-parallel mirrors is metastable. This means that small imperfections in the mirror surface may result in an unstable resonator. On the other hand, the beam waist diameter of the Fabry-Perot resonator mode is infinite whereas the resonator modes inside the stable regions exhibit

finite beam waist diameters. The losses that a mirror of finite dimensions introduces depend on the ratio of beam waist and mirror size and increase exponentially with smaller mirrors. Not only the diffraction losses at the mirrors have to be considered for laser action. The second aspect is the effective use of the gain medium inside the resonator. In a Fabry-Perot resonator the whole medium between the mirrors is homogeneously filled with light which leads to a proper use of the provided gain. The volume inside the resonator filled with light is called the mode volume. In stable resonators with bent mirrors (concave-concave or convex-concave pairs) the mode volume is drastically reduced. In this case some gain regions may not contribute to the amplification, reducing the overall laser efficiency. When such systems are misaligned, i.e. the optical axes of the mirrors do no longer coincide, the losses increase but the mode volume increases as well, making better use of the gain region. Such mirror systems are so-called unstable resonators because they exhibit losses even with ideal mirrors. In high-power lasers where the gain region is of finite lateral extent stable resonators lead to so-called filamentation. Because of the unused gain regions excess carriers introduce a gain profile that supports higher transverse order modes which in turn produce very bright lines on the output facets. The high intensity in those filaments can lead to mirror damage resulting in a totally inoperable laser. With unstable resonators (misaligned optical axes or combinations that are belong to the shaded regions of figure 7.1) the mode volume increases and the chance for the accumulation of excess carriers is minimized, resulting into lower probability of filamentation at the expense of higher driving power due to higher resonator losses.

7.2 Laser Threshold

As we have already seen in section 2.1, for laser action the threshold condition requires that the losses have to be balanced by the amplification on one roundtrip of light inside the resonator. The total loss of the laser resonator system is due to a number of different processes. Some of the most important losses are

1. Transmission of the mirrors: the transmission through the mirrors usually provides the useful light output. On the other hand it increases the pumping threshold. A compromise between output power and low threshold has to be chosen. For instance the reflectivity of one mirror can be made higher than the other to reduce the threshold and get high output at the lower reflective mirror at the same time.

2. Absorption and scattering at mirrors: especially the effect of absorption can lead to mirror damage in high power-lasers and must be avoided.

3. Absorption inside the resonator: some absorption always takes place inside the resonator but not all the free carriers generated by absorption contribute to lasing and therefore a net absorption has to be accepted.

4. Scattering at impurities of the laser medium: this effect applies mainly to solid state lasers.

5. Diffraction losses at mirrors: normally the mirror size is much bigger than the beam waist diameter. When both are of the same size diffraction at the mirror edges is considerable.

7.3 Noise in Lasers

When the output power of a laser in short time intervals is measured it is found that the power varies statistically. This effect originates in the generation process of photons by discrete transitions between two energy levels. In a simple picture the stimulated emission provides transitions of a certain frequency ν phase matched to the input beam but the time when the emission occurs is not predetermined, leading to an emission probability that is Poisson distributed as

$$p\{N_\mathrm{p}\} = \frac{m^{N_\mathrm{p}}}{N_\mathrm{p}!}\exp\{-m\} \tag{7.1}$$

where N_p denotes the number of photons generated within the time interval T, while assumming that the mean number of photons generated during that time is $\langle N_\mathrm{p} \rangle = m$ which follows from

$$\langle N_\mathrm{p} \rangle = \sum_{N_\mathrm{p}=0}^{\infty} N_\mathrm{p} p\{N_\mathrm{p}\} = m \quad . \tag{7.2}$$

The second mechanism that gives rise to noise is that the emitted light is not really fixed to a single frequency but is spread over a narrow spectrum as we will discuss in the following section. Together with the characteristics of the resonator transmission, frequency modulation transforms into amplitude modulation as sketched in figure 7.2.

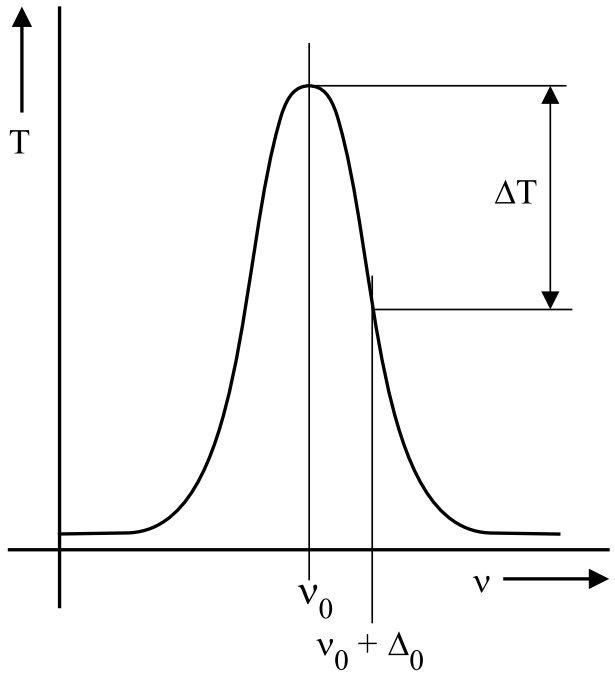

Figure 7.2: Transmission spectrum of a laser resonator. Light frequencies shifted by $\Delta\nu$ against the resonance frequency ν_0 result in lower output levels.

Due to the fact that the emission is drastically distributed as well in time as in frequency this effect gives rise to amplitude noise. The last effect considered here is the contribution of spontaneous emission to noise. Spontaneous emission is completely

decorrelated from the laser emission in frequency and phase. Only spontaneous emissions with the same frequency as the laser line may be considered here first. Imagine the stimulated emission to have $\langle N_\mathrm{p} \rangle = m$ photons. Normally the number of spontaneously emitted photons at the same frequency is very low. However their fields add with an arbitrary phase angle Φ to the field of stimulated emission, as sketched in figure 7.3, changing amplitude and phase of the total field by ΔE and $\Delta\Phi$, respectively.

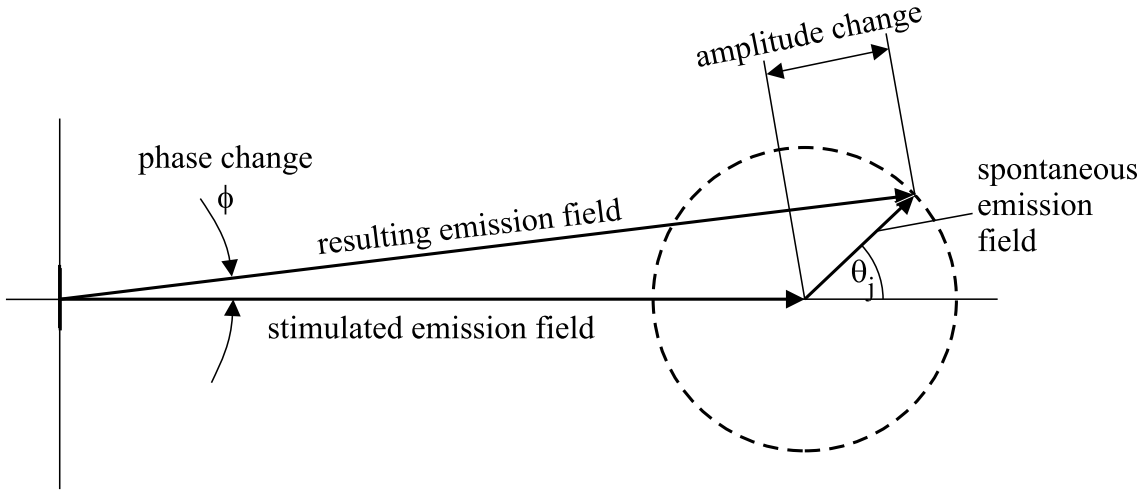

Figure 7.3: Noise contribution of one spontaneously emitted photon to the laser emission. The relative phase θ_j between stimulated and spontaneous emission changes with each individual spontaneous emission process.

In a time interval T the number of stimulated emitted photons may be M_p and the corresponding photon density is $M_\mathrm{phot} = M_\mathrm{p}/V_\mathrm{active}$. At the same time m spontaneously emitted photons are generated with the same frequency and polarization. Each individual photon contributes with electric field $\vec{E}_0$ to the total field. The electric field of the stimulated emission follows from the energy density with $\vec{E}_\mathrm{stim} = \vec{E}_0\sqrt{M_\mathrm{p}}$ whereas the spontaneously emitted field is $\vec{E}_\mathrm{spon} = \vec{E}_0 T \sum_{j=1}^{m} \exp\{i\theta_j\}\delta\{t - t_j\}$. Here we have assumed that the spontaneous emission processes take place at certain times t_j within the interval T. For the total field $\vec{E} = \vec{E}_\mathrm{stim} + \vec{E}_\mathrm{spon}$ we write $\vec{E}_\mathrm{tot} = \vec{E}_0\sqrt{N_\mathrm{p}}\exp\{i\Delta\Phi\} \simeq \vec{E}_0\sqrt{N_\mathrm{p}}(1 + i\Delta\Phi)$ assuming a small phase shift due to

spontaneous emission. When we compare this expression with the real part and the imaginary part of the total field

$$\vec{E}_{\rm tot} = \vec{E}_0\sqrt{M_{\rm p}} + \vec{E}_0 T \sum_{j=1}^{m} \exp\{i\theta_j\}\delta\{t-t_j\} \quad , \tag{7.3}$$

the total number of photons $N_{\rm p}$ results from

$$\sqrt{N_{\rm p}}(1+i\Delta\Phi) \simeq \sqrt{M_{\rm p}} + T\sum_{j=1}^{m}\left(\cos\{\theta_j\} + i\sin\{\theta_j\}\right)\delta\{t-t_j\} \quad . \tag{7.4}$$

After multiplication with $\sqrt{M_{\rm p}}$, from the real part of (7.4) we find for the change in the number of photons $\Delta N_{\rm p} = N_{\rm p} - M_{\rm p} \ll M_{\rm p}$ with $\sqrt{N_{\rm p}M_{\rm p}} = M_{\rm p} + \Delta N_{\rm p}/2$ and $M_{\rm p} \simeq N_{\rm p}$

$$\Delta N_{\rm p} \simeq 2T\sqrt{M_{\rm p}}\sum_{j=1}^{m}\cos\{\theta_j\}\delta\{t-t_j\} \simeq 2T\sqrt{N_{\rm p}}\sum_{j=1}^{m}\cos\{\theta_j\}\delta\{t-t_j\} \quad . \tag{7.5}$$

The value of $\Delta N_{\rm p}$ gives the amplitude change generated by spontaneous emission in the time interval T. It is common to define a noise generation rate

$$F_{\rm phot} = \frac{\Delta N_{\rm p}}{V_{\rm active}T} \simeq \sqrt{N_{\rm phot}}\frac{2}{\sqrt{V_{\rm active}}}\sum_{j=1}^{m}\cos\{\theta_j\}\delta\{t-t_j\} \tag{7.6}$$

which is called a Langevin force and gives the spontaneous emission photon density generation rate into a mode. For the description of laser operation by rate equations also the Langevin Force $F_{\rm N}$ for carrier noise is required. Each photon generated is caused by a carrier recombination. Besides spontaneous emission also non radiative recombinations have to be considered in the carrier Langevin force making it bigger than the photon Langevin force.

The phase change $\Delta\Phi$ caused by spontaneous emission follows under the same assumptions as the photon density change from (7.4) to be $\Phi \simeq \frac{T}{\sqrt{M_{\rm p}}}\sum_{j=1}^{m}\sin\{\theta_j\}\delta\{t-t_j\}$ and the Langevin force for the phase change is

$$F_\Phi = \frac{\Delta\Phi}{T} \simeq \frac{1}{\sqrt{N_{\text{phot}}}} \frac{1}{\sqrt{V_{\text{active}}}} \sum_{j=1}^{m} \sin\{\theta_j\}\delta\{t - t_j\} \tag{7.7}$$

The Langevin forces can be considered as small-signal modulation of the laser system. The Langevin force F_{phot} leads to the so-called relative intensive noise (RIN) and F_Φ gives rise to a spectral broadening of the spectrum of emitted light. It is worth to note that the amplitude noise grows whereas the phase noise decreases with the square root of the emitted power. This means that the spectral width of a laser line decreases with increasing output power.

Figure 7.4 illustrates how spontaneous emission acts as a driving force for laser action thereby broadening the spectrum. When we follow the electric field amplitude and phase on a round trip in an optical resonator the complex representation of the electric field amplitude $E\{z\} = E\{0\}\exp\{ik_z z\}$ follows a perfect circle trace in lossless resonators. When gain, internal absorption and mirror losses are considered the trace swings around the circle reconstructing itself after one roundtrip as follows from the resonance conditions for phase and amplitude. Additional spontaneous emissions modulate that trace arbitrary. Only when the field reconstructs itself after a roundtrip the resonance conditions are satisfied and stable operation follows. It has to be noted that the resonance conditions follow from boundary equations that have to be satisfied on the mirrors. Therefore the trace between the mirrors may be arbitrary. This leads to the effect that even for wavelength, where the phase condition is not satisfied in the coherent emission model with spontaneous emission the additional phase change $\Delta\Phi$ (see figure 7.3) can adjust the total phase and emission occurs. The probability that a phase change $\Delta\Phi$ is obtained from spontaneous emission decreases with growing $|\Delta\Phi|$ and therefore laser emission decreases apart from the center wavelength rapidly.

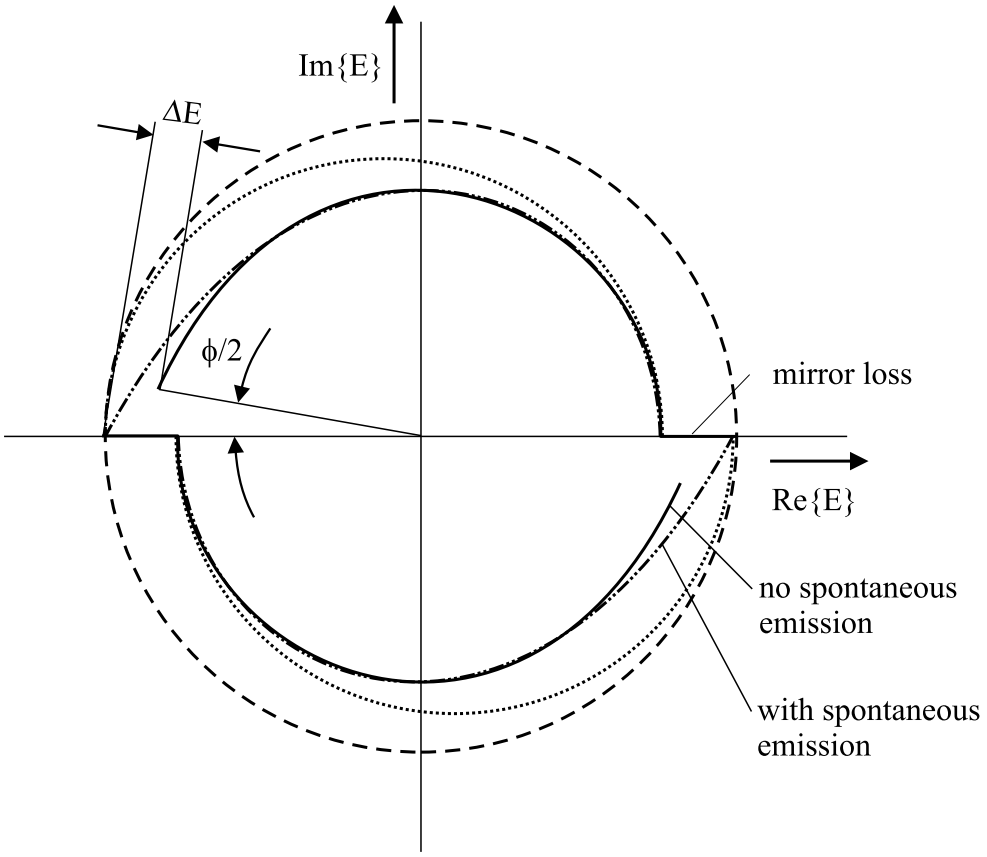

Figure 7.4: Trace of the complex electric field vector on one roundtrip through a laser resonator. With perfect mirrors and without losses a circle (dashed line) is results. When internal losses, gain and transmissive mirrors are taken into account the field reproduces itself after one roundtrip (dotted line). In both cases only for one wavelength and for a certain gain the resonance conditions are satisfied. With spontaneous emission the trace is modulated arbitrary and with a certain probability a spectral range of wavelength are able to satisfy the resonance conditions.

7.4 Lineshape

Normally laser lines are assumed to be perfectly monochromatic. A closer look shows that in fact there is a small spectral region covered. From section 1.2 we know that the emission and absorption spectra are closely connected to each other and measurements of light transmission through laser material show a bell shaped curve as sketched in figure 7.5. Using the knowledge that the transmission is an indirect (logarithmic) measure of the absorption we can derive the (spontaneous) emission spectrum also depicted in figure 7.5.

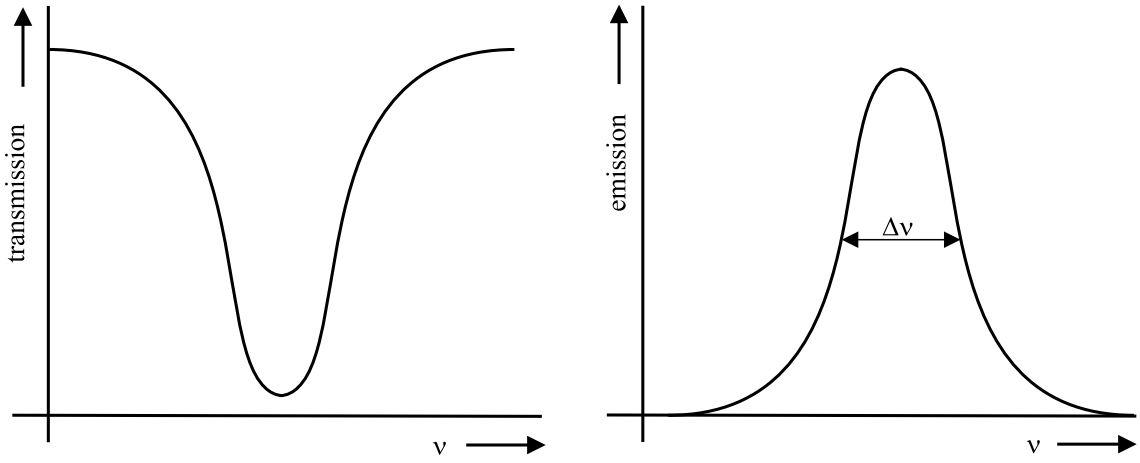

Figure 7.5: Transmission and spontaneous emission spectra for a laser material. The transmission is a logarithmic measure of absorption which gives a measure for the the emission spectrum via the van Roosbroeck-Shockley relation and therefore for the available gain (stimulated emission).

Due to the fact that the Einstein coefficients of spontaneous and stimulated emission are equal, the emission curve is a direct measure of the gain available in the material. The shape of an emission curve is called lineshape function $\tilde{g}\{\nu\}$. The spectral width $\Delta\nu$ is normally measured as Full-Width-at-Half-Maximum (FWHM) spread and called linewidth. Lineshape functions are normalized in the way $\int_{-\infty}^{\infty} \tilde{g}\{\nu\}\mathrm{d}\nu = 1$.
When a gain medium is placed inside a resonator, laser emission can take place only for

frequencies where the gain exceeds the losses (correctly: is close to the losses, see 7.8). Due to the bell shaped gain curve only in a small spectral region the available gain reaches the magnitude of the losses. The spectral width of the lasing line is therefore remarkably decreased compared to the spontaneous emission linewidth. The maximum available gain depends on the pumping strength and the width of the gain spectrum. The broader the linewidth is the lower the maximum gain will be at a certain pumping level. The spectral linewidth is caused mainly by Doppler broadening, collision broadening, natural damping, and spontaneous emission. Commonly the broadening effects are classified as homogeneous and inhomogeneous broadening effects which lead to different lineshape functions.

7.4.1 Homogeneous Broadening

Homogeneous broadening effects are collision broadening and natural damping. It will be shown later that also spontaneous emission leads to the same lineshape of the laser line in semiconductor lasers and is therefore classified as homogeneous broadening. Collision broadening is found in gas lasers. When photon emitting atoms or molecules suffer a collision, the wave train of the associated photon is shortened. Via Fourier analysis it can be shown that a shortened wavetrain is associated to a broadened spectrum. The magnitude of collision broadening clearly depends on the gas pressure and temperature and will increase with both. Natural damping of a wavetrain inside a material caused by several effects like scattering and absorption looks like a shortened wavetrain under Fourier transformation and leads to the same lineshape as collision broadening. Homogeneous broadening leads to a Lorentzian lineshape function

$$\tilde{g} = \frac{1}{2\pi} \frac{\Delta\nu}{(\nu - \nu_0)^2 + (\frac{\Delta\nu}{2})^2} \tag{7.8}$$

with center frequency ν_0 and linewidth $\Delta\nu$. The maximum at $\nu = \nu_0$ is $\max\{\tilde{g}\} = \frac{2}{\pi\Delta\nu}$. In semiconductor lasers the refractive index depends on the carrier density which is described by the alpha factor α_{H} also called Henry factor. Carrier modulation caused by

spontaneous emission together with resonator properties leads to spectral broadening of the emission with the same lineshape as is found in gas lasers with homogeneous broadening as we will see later.

7.4.2 Inhomogenous Broadening

Doppler broadening is normally the main broadening effect in gas lasers. The Doppler effect describes the frequency shift when source and observer move against each other. With movement v much slower than light velocity c the observed frequency is $\nu' = \nu(1 \pm \frac{v}{c})$, where the plus sign applies for approaching the source and the minus sign for receding. As all the atoms or molecules in a gas laser move in arbitrary directions, a spectrum of frequencies is observed despite the fact that all emitted frequencies are the same. This averaged output spectrum has a gaussian lineshape

$$\tilde{g}\{\nu\} = \frac{2}{\Delta\nu}\sqrt{\frac{\ln\{2\}}{\pi}}\exp\left\{-\ln\{2\}\left(\frac{\nu-\nu_0}{\Delta\nu}\right)^2\right\} \tag{7.9}$$

which also results, when the participating transitions have slightly different frequencies or lineshapes, as found for example with local temperature variations or crystal imperfections. In semiconductor lasers with good crystals inhomogeneous broadening is normally dominated by homogeneous broadening.

7.5 Emission Spectrum

The emission spectrum of a laser consists of one or more laser lines of spectral width $\Delta\nu$. The separation between the laser lines is determined by the resonance condition of the laser resonator. Assuming a resonator of length L the longitudinal phase condition requires $-\arg\{r_1\}-\arg\{r_2\}+2\pi\frac{\bar{n}_{\text{eff}}}{\lambda}2L = m2\pi$, where r_1 and r_2 are the mirror reflection coefficients. The effective refractive index $\bar{n}_{\text{eff}}$ differs in waveguide lasers from the bulk material refractive index.

For long cavities the spectral distances $\Delta\lambda$ between two neighboring (longitudinal) resonator modes can be calculated for long cavities (mode order $m \gg 1$) from

$$\frac{\Delta m}{\Delta\lambda} \simeq \frac{\mathrm{d}}{\mathrm{d}\lambda} m = -\frac{2L\bar{n}_{\mathrm{gr\ eff}}}{\lambda^2} - \frac{1}{2\pi}\frac{\mathrm{d}}{\mathrm{d}\lambda}(\arg\{r_1\} + \arg\{r_2\}) \quad . \tag{7.10}$$

Assuming no phase change at the mirrors, i.e. $\arg\{r\} = 0$, the mode spacing $\Delta\lambda$ follows with $\Delta m = 1$ as

$$\Delta\lambda = -\frac{\lambda^2}{2L\bar{n}_{\mathrm{gr\ eff}}} \quad . \tag{7.11}$$

When the mirror phase change is not neglected an effective laser length L_{eff} is introduced in (7.11) and comparison with (7.10) results into

$$L_{\mathrm{eff}} = L + \frac{\lambda^2}{4\pi\bar{n}_{\mathrm{eff}}}\frac{\mathrm{d}}{\mathrm{d}\lambda}(\arg\{r_1\} + \arg\{r_2\}) \quad .$$

Equation (7.11) implies an equally spaced line spectrum ($\Delta\nu = \frac{c}{2L_{\mathrm{eff}}\bar{n}_{\mathrm{eff}}}$) over the entire wavelength regime. In fact only in a small region laser emission can occur. This is due to the fact that not only the phase condition but also the amplitude condition for laser action has to be satisfied. The spectral range with sufficient gain in conjunction with the line spacing determines how many lines can lase. In (7.11) only the axial resonance is regarded and the effective index $\bar{n}_{\mathrm{eff}}$ implicitly considers the role of the transverse modes. Each transverse mode can lase when the longitudinal resonance condition is satisfied. This results into a line spectrum for each transverse mode, and the total spectrum results from the superposition of these individual comb-like spectra, as illustrated in figure 7.6 for four transverse modes.

Monomode laser operation therefore requires that there is only one transverse mode and in addition only one longitudinal mode inside the gain determined spectral region or less strictly only one mode does satisfy the resonance conditions. This can be reached in two ways. The simple way is to shorten the cavity so much that the line spacing $\Delta\lambda$ in (7.11) is slightly bigger than the useful spectral region. In this case all neighboring modes lie outside the gain region. The design of vertical-cavity surface-emitting lasers

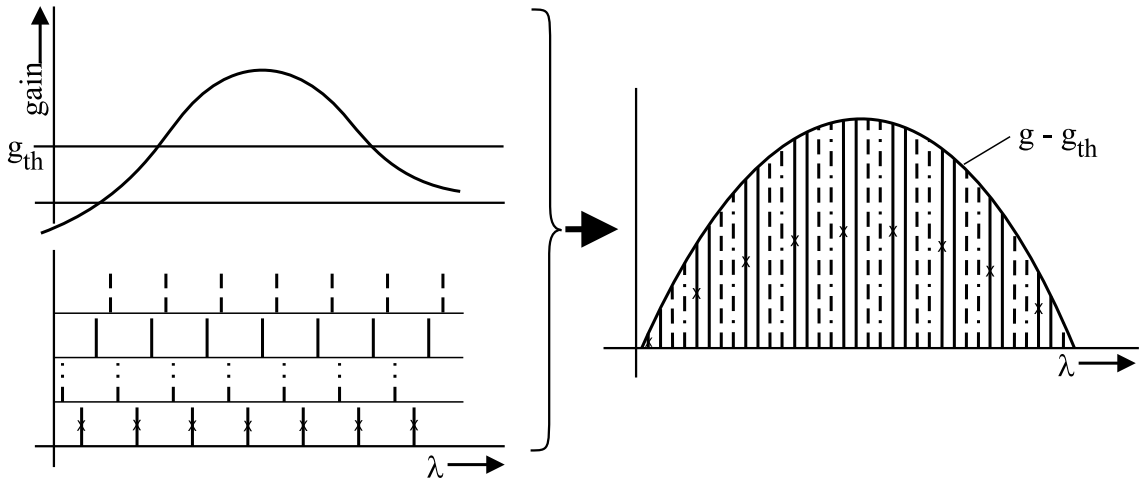

Figure 7.6: Spectral gain and line spectrum in a laser with four transverse modes. The resulting spectrum on the right-hand side is a superposition of the four axial spectra limited by the amplitude condition.

(VCSELs) relies on short cavities, resulting in single longitudinal mode operation. The other way is to modify the threshold gain curve that is normally flat. With DFB and DBR lasers only the central modes experience sufficient reflectivity for lasing, as is illustrated in figure 7.7.

7.6 Mode Locking

As we saw in the previous section, a laser normally emits in several spectral lines, the amplitudes and frequencies of which are not correlated to each other. Under certain circumstances it is possible to mutually couple them such that they superpose to short pulses in the light output. This case is called mode locking and is illustrated together with the unlocked case in figure 7.8.

A mathematical description of this correlation effect starts from the electric fields of each individual laser line (linewidth assumed as zero for simplicity) with angular frequency spacing between the lines $\Delta\omega = \frac{\pi c}{L}$ and individual emission frequency $\omega_n =$

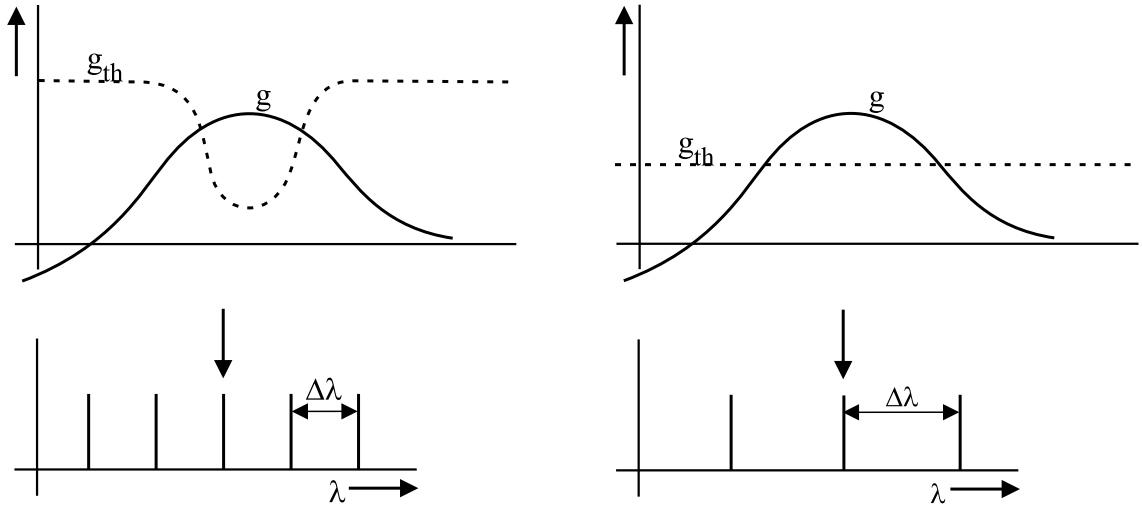

Figure 7.7: Single longitudinal mode selection via modification of the reflectivity as in DFB and DBR lasers (left) or with short cavity length as in VCSELs (right).

$\omega_0 + n\Delta\omega$ of mode n. The fields superpose to the total field

$$E_t = \sum_{n=1}^{N} E_n \exp\{i(\omega_n t + \delta_n)\} \quad , \tag{7.12}$$

where δ_n denotes the phase of the n-th mode. Without correlation, E_n and δ_n may change with time, giving rise to a total irradiance

$$I = |E_\mathrm{t}|^2 = \sum_{n=1}^{N} |E_n|^2 \quad . \tag{7.13}$$

With mode locking all phases δ_n are varying the same way – they are locked together. In this case the irradiance follows from the total field

$$E_\mathrm{t} = \exp\{i\delta\} \sum_{n=1}^{N} E_n \exp\{i(\omega_0 + n\Delta\omega)t\} \tag{7.14}$$

with equal phases δ assumed. Taking $E_n = E_0$ for simplicity, (7.14) further simplifies to

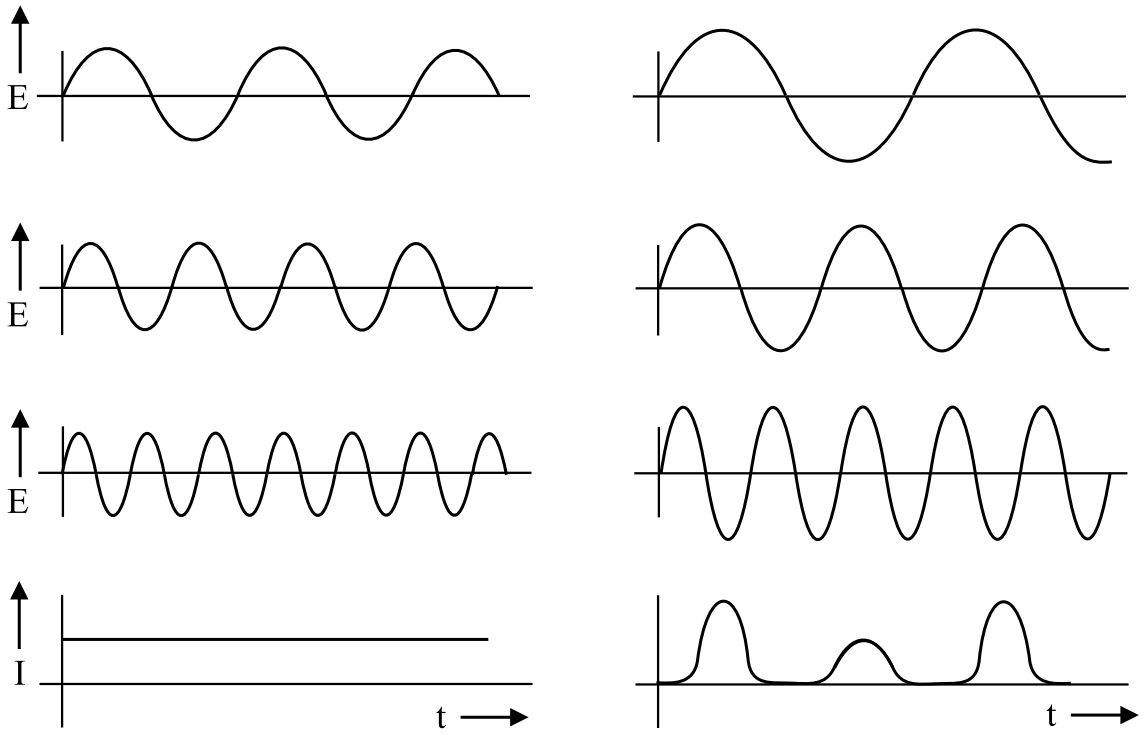

Figure 7.8: Mode locking of three modes. On the left-hand side the three modes are not correlated and the output intensity is nearly time independent. The modes on the right-hand side are strongly correlated and produce short pulses in the output intensity.

$$E_{\mathrm{t}} = E_0 \exp\{i(\omega_0 t + \delta\} \sum_{n=1}^{N} \exp\{in\Delta\omega t\} \quad , \tag{7.15}$$

which reduces to

$$E_{\mathrm{t}} = E_0 \exp\left\{ i \left[\left(\omega_0 + \frac{N+1}{2}\Delta\omega \right) t + \delta \right] \right\} \frac{\sin\{\frac{N}{2}\Delta\omega t\}}{\sin\{\frac{1}{2}\Delta\omega t\}} \quad , \tag{7.16}$$

giving the time dependent irradiance

$$I = E_0^2 \frac{\sin^2\{\frac{N}{2}\Delta\omega t\}}{\sin^2\{\frac{1}{2}\Delta\omega t\}} \quad . \tag{7.17}$$

In figure 7.8 the resulting irradiance for $N = 3$ is depicted. The maximum output in the pulses is $I_{\max} = N^2 E_0^2$ as can be seen from 7.15 in the cases $\Delta\omega t_{\mathrm{m}} = p2\pi$ or $\Delta\omega t_{\mathrm{m}} = (2p+1)2\pi$ with $p \in \mathbb{N}_0$. The pulses are periodic with $\Delta\omega t_{\mathrm{p}} = 2\pi$ giving $t_{\mathrm{p}} = \frac{2L}{c}$. This is just the time required for a complete for one round trip of the field in the resonator. The pulse drops to zero at times t_{z} when $N\Delta\omega t_{\mathrm{z}} = q2\pi$ and $\Delta\omega t_{\mathrm{z}} \neq p2\pi$, $\{p, q\} \in \mathbb{N}_0$ at the same time. The spacing $\Delta t = t_{\mathrm{z}} - t_{\mathrm{m}}$ between a maximum at $\frac{\Delta\omega}{2} t_{\mathrm{m}} = q\pi$ and the next zero $\frac{N}{2}\Delta\omega t_{\mathrm{z}} = p\pi$ requires $N\frac{\Delta\omega}{2}\Delta t = \pi$ which results into $\Delta t = (\frac{1}{N})(\frac{2L}{c}) = \frac{t_{\mathrm{p}}}{N}$. Here we see that the pulse width decreases with the number of modes locked together and simultaneously the power increases quadratically with N.

The question that arises is how can mode locking be achieved. Two methods namely active and passive mode locking can be used. The fixed phase relationship between longitudinal modes is achieved in active mode locking by ether gain or loss modulation. The difference between gain and loss is adjusted such that most of the time it is very small but for a short moment it is periodically increased. When the modulation periodicity and the periodicity of the pulses are equal, the pulses are much more enhanced than the rest of the output. Practically this is achieved by electrooptic switches where the losses are modulated and the gain stays constant or by modulation of the gain with constant losses. The first scheme is found mainly in gas lasers whereas the second one is preferred in semiconductor lasers where the driving current is easily modulated.

Passive mode locking makes use of saturable absorbing effects as discussed in section 5.2. Low intensity portions of light suffer more attenuation than parts with higher intensity due to a bleaching of the absorbing medium. This in turn enhances the mode locking once a pulse has established. In the initial operation phase a small pulse may be caused through spontaneous emission and with sufficient pulse energy mode locking evolves.

7.7 Laser Classes

This section deals with an overview of different laser classes beginning with the best known gas lasers where ionized gas or exited gas atoms serve the gain necessary for amplification. In liquid dye lasers special liquids which were traditionally used for coloring are able to generate light under exitation.
In parametric lasers nonlinear effects ensure the light amplification whereas in free electron lasers the radiation originates from charged particles that cross the light beam periodically at a speed close to the light velocity. In doped insulators the dopands can be excited and provide amplification whereas in semiconductor lasers the inversion is directly reached by current flow through pn-junctions. The inversion is extremely efficient in double heterostructures where carrier injection from wide gap materials helps.

7.7.1 Gas Lasers

Gas lasers are widely used, they range from the low power helium-neon (He-Ne) lasers commonly used in teaching laboratories to high power carbon dioxide (CO_2) lasers, which have many industrial applications. Gas lasers may be divided into three classes depending on the origin of the radiating transition: between electronic levels of atoms, or of ions, or between the vibrational/rotational levels of molecules. Pumping of those levels is not very easy because they are well defined, i.e. they have narrow absorption bands. For this reason optical pumping is nearly not possible and other methods as activation through collision in a gas discharge are used.

7.7.1.1 Atomic lasers

In atomic lasers the radiating transition takes place between the electronic levels of the atom. The best known atomic laser is the He-Ne laser. Here the neon provides the lasing transitions and the helium serves for pumping. The typical ratio of He:Ne is 10:1. The gas pressure is 100 Pa and discharge ignites at about 2 - 4 kV. As can

be seen from figure 7.9 mainly two exited states of the helium atoms transfer energy to the neon atoms resonantly which thermalize by emission of radiation and collision. The first transition contributes to lasing whereas the transition from 2p to 1s states is the well known neon emission as found in neon lights. The problem is that the 1s level fills rapidly which limits the inversion of energy states. Therefore the laser output not simply increases with driving current and He-Ne lasers are limited to low output power. The cross section of a typical He-Ne laser is depicted in figure 7.10.

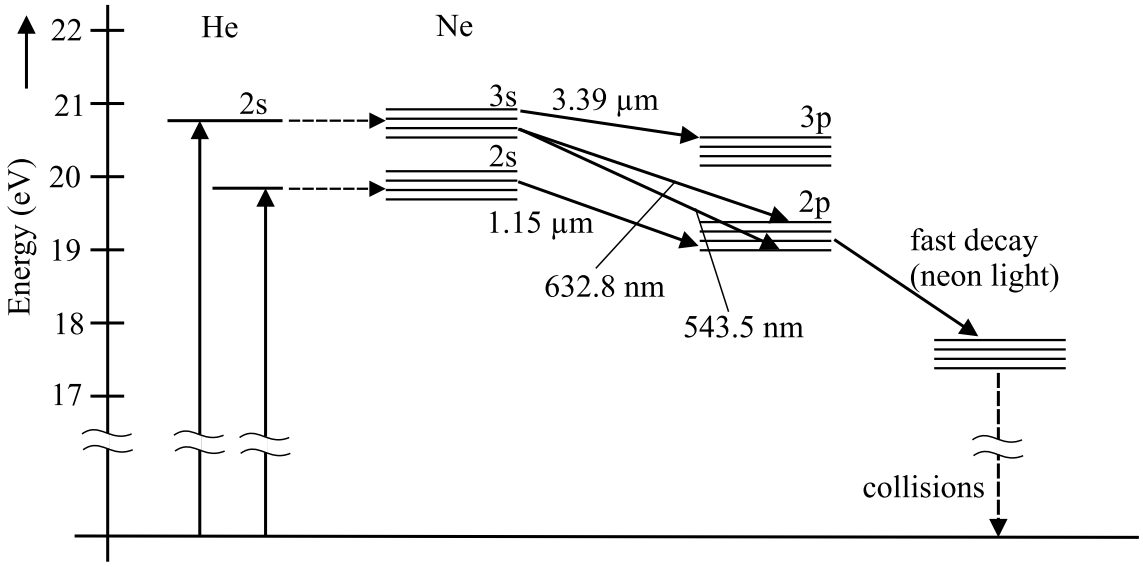

Figure 7.9: Energy states in He-Ne lasers. Excited He atoms transfer energy to the Ne atoms which in turn relax by radiative transitions and sidewall collisions.

The efficiency of He-Ne lasers is very low. The upper theoretical limit is given by the rates for lasing energy ($\simeq 1\,\mathrm{eV}$) and pumping energy ($\simeq 20\,\mathrm{eV}$) to be about 5 %. But in fact not all the excitation energy is transferred from the He atoms to Ne. The first reason is the Ne content of 10 % which leads to a maximum efficiency of 0.5 % and the second one is that resonant energy transfer is a tunneling process, the probability of which is determined by $\exp\{-\frac{\Delta E}{k_\mathrm{B}T}\}$ where ΔE is the energy difference of the exchanging states. Additionally also the He atoms thermalize by sidewall collision. In typical He-Ne lasers an efficiency of 0.02 % is found.

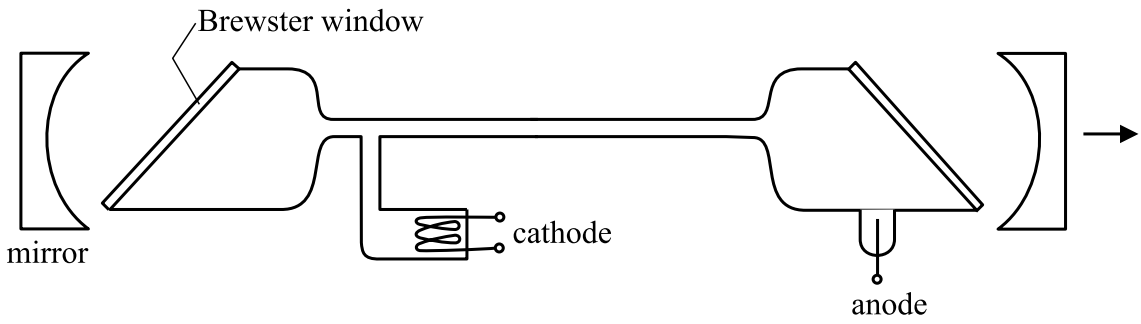

Figure 7.10: Cross section through a typical He-Ne laser.

Due to the fact that three of the four main lasing transitions start from the same level provisions must be taken to enforce only one line to lase. This is typically achieved by dielectrically coated mirrors in the form of Bragg mirrors that give enhanced reflectivity only for the desired wavelength.

7.7.1.2 Ion lasers

Typical representatives of ion lasers are the noble gas ion laser and metal vapor ion laser. A well known noble gas ion laser is the argon ion laser that emits in the visible spectral range. Krypton ion lasers are as well emitting visible light and are becoming increasingly used for physical and chemical investigations. In figure 7.11 the energy diagram of an Ar-ion laser is sketched. Pumping from the ground state with accelerated electrons of a few eV energy is achieved by many collisions. Several lines are emitted under transition between the 4p and 4s levels where inversion is generated. The 4s level is emptied by ultraviolet emission to the ion ground state. The recombination energy to the ground state is emitted as heating of the system.

The excitation is enforced when a longitudinal magnetic field is applied which transfers energy to the ions that spiral around the magnet field lines. Due to the strong heating the laser tube has to be cooled which is achieved by a tube made of refractory material as graphite or beryllium oxide with cooling discs inside as sketched in figure 7.12.

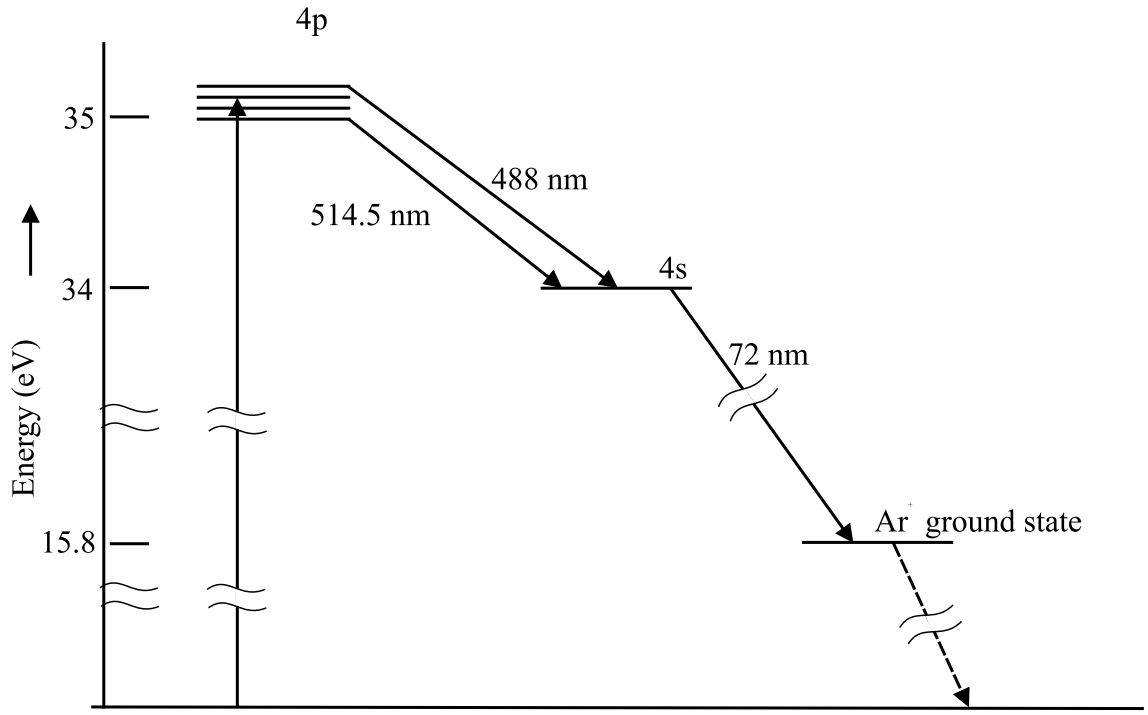

Figure 7.11: Energy diagram of an Ar-ion laser. Several lines are emitting under transition between the 4p and 4s levels. The two given above are the strongest lines.

The positively charged ions tend to collect at the cathode thereby preventing further excitation by shielding the field. A gas return path to the anode serves for pressure equalization and enables continuous wave operation.

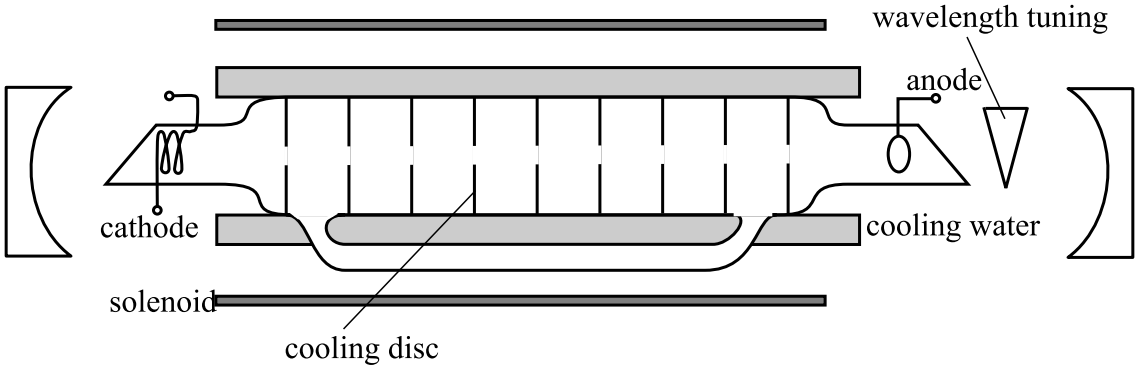

Figure 7.12: Cross section through a typical argon ion laser.

The laser is tuned to the desired wavelength by means of a prism that is tilted and determines which wavelength encounters the lowest losses in the optical resonator.

Metal ion lasers are becoming important due to the possibility to get ultraviolet laser emission which is useful for example in high resolution lithography. The operation principle is very similar to that of a He-Ne laser. Excitation of metal vapor is achieved via collision with excited gas atoms or molecules. The most demanding problem in metal vapor lasers is to prevent the laser windows from being coated with condensing metal. Due to the positive charge of ionized metal vapor the metal flows from the anode to the cathode under an applied electric field. In figure 7.13 a cross section through the well established He-Cd laser is sketched.

The tube between vapor source and condensor is heated to establish a uniform vapor density and prevent condensation on the sidewalls. Typically a He pressure of 460 Pa and evaporation temperatures of 220 ^{o}C are found in He-Cd lasers. The main emission lines are 441 nm and 325 nm which makes the He-Cd laser favorable for microelectronic lithography because the photoresist used is highly sensitive to these wavelengths.

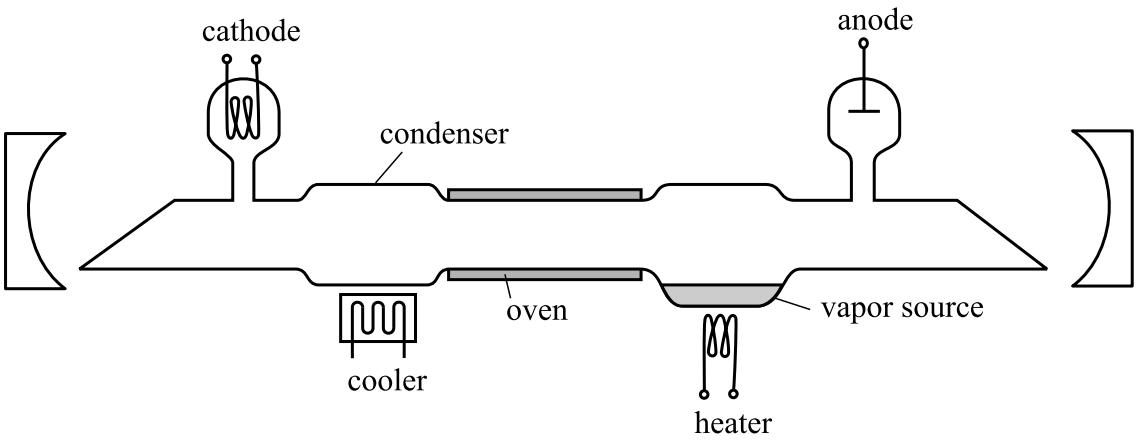

Figure 7.13: Cross section of a metal vapor ion laser.

7.7.1.3 Molecular Lasers

In molecular lasers the energy states follow from rotational and vibrational modes of the constituent gas molecules. The most prominent molecular laser is the carbon-dioxide (CO_2) laser. Today it is the most important laser from the standpoint of technological applications. Some of them have been discussed in section 6.2. Due to this special importance we will use the example of a CO_2 laser to explain the operation of molecular lasers.

The carbon dioxide molecule is basically a linear arrangement of oxygen-carbon-oxygen. Principally there exist three vibrational modes namely the symmetric, asymmetric and the bending mode. Each vibration state of the molecule can be described by superposition of these three basic modes that are sketched in figure 7.14. In the symmetric mode the oxygen atoms vibrate symmetrically to and fro the carbon atom. In the bending mode the oxygen atoms stand still and the carbon atom vibrates normally to the molecule axis. The asymmetric vibration is described as an axial vibration of one oxygen atom against the remaining CO-pair.

The vibration frequency of each mode is characterized by its frequency relative to the ground state frequency of that vibration. A set of three numbers (m_1, m_2, m_3) denotes the complete molecule vibration where m_1 denotes symmetric, m_2 bending, and m_3

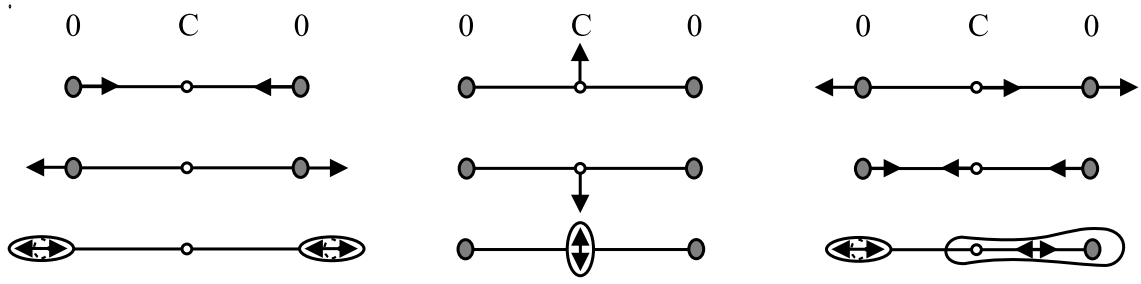

Figure 7.14: The three basic vibrational modes of ca CO_2 molecule: symmetric (left), bending (center), and asymmetric (right).

asymmetric vibration. For example (0,2,0) means that the molecule bends with twice the vibration frequency of the slowest bending vibration possible. Additionally the complete molecule may rotate inside the laser and this rotation splits the energy states of each vibration mode into closely spaced levels. In CO_2 lasers nitrogen serves for pumping and is therefore the major part of the filling gas. Additional He increases the efficiency and decreases Doppler broadening. The energy level diagram of a CO_2 laser is sketched in figure 7.15.

The efficiency enhancement through He is achieved via depletion of the lower energy states enabling higher population inversion. The Doppler broadening decreases because He has a higher thermal conductivity and enhances the cooling of CO_2 molecules. Typical gas contents are 1:4:5 = CO_2:N:He.

Not only vibrational modes of molecules contribute to the gain but also simple transitions between electron energy states as found in nitrogen lasers or by radiation of light through the dissociation of metastable states which is found in Excimer lasers. Nitrogen lasers give only pulse output because inversion is only achieved for short times due to the fact that the lifetime in the higher energy level is much shorter than the lifetime in the lower energy level. Hence laser output can only be stabilized for a few microseconds. Special provisions must be taken to ensure pumping of the upper level very fast. Excimer are associations of an excited and a non-excited atom (**Ex**cited Di**mer**) that attract each other on an atomic scale. If the excitation gets lost by radiation the

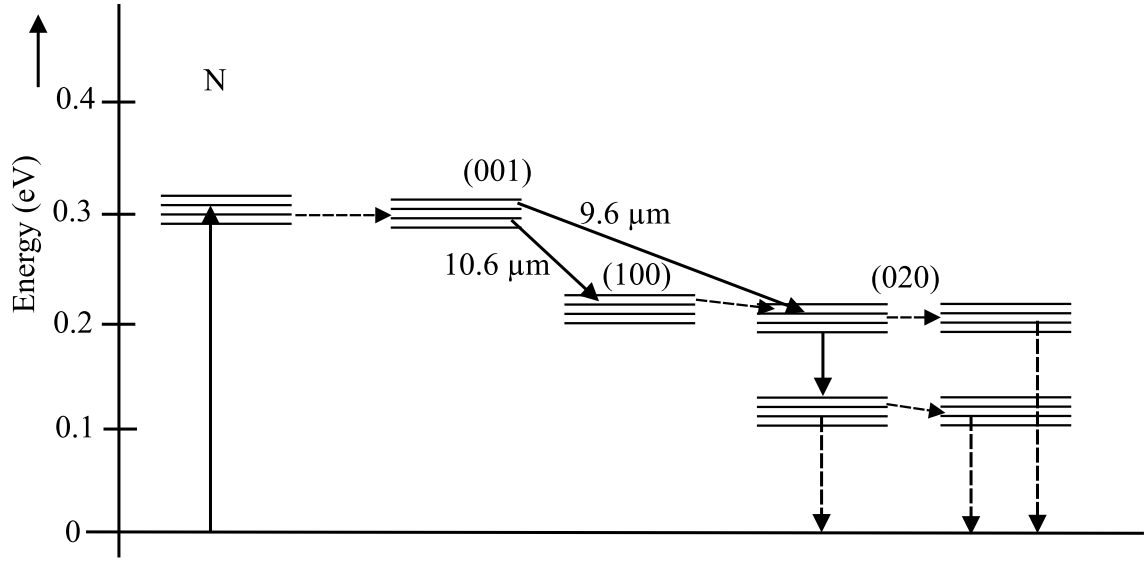

Figure 7.15: Energy level diagram of a CO_2 laser. Nitrogen serves for pumping and He decreases the population level at lower states thus enhancing population inversion.

dimer dissociates rapidly because the attraction changes to repulsion hereby reducing the population of the lower energy level. Excitation of Excimer lasers requires high energy which is provided by a so called transversal excited configuration as sketched in figure 7.16. The distance between cathode and anode is very small, producing high electric fields with acceptable voltage. Ions are produced by collision reactions in the high pressure gas. Excimer lasers work in pulsed mode and are able to emit from the visible down to the ultraviolet regime (500 nm - 120 nm) for XeF or KrF gas mixtures.

7.7.2 Liquid Dye Lasers

Lasers with liquid gain material have several advantages over solid state and gas lasers. In gas lasers the density of active material is relatively low and in solid state lasers it is difficult to ensure the required high degree of homogeneity. Several lasers with liquid active material have been developed and the most popular is the dye laser. The energy diagram of a dye laser is sketched in figure 7.17.

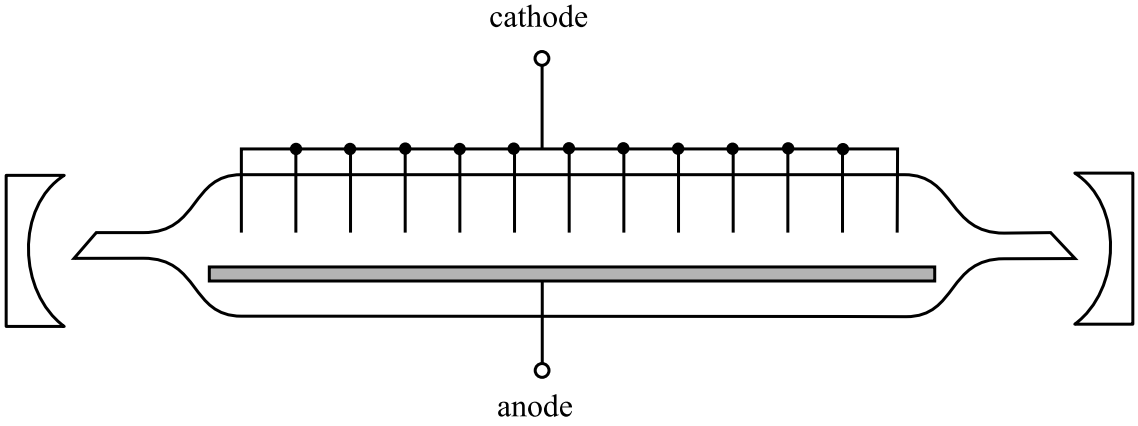

Figure 7.16: Transversely excited gas laser. This configuration is found in Excimer lasers and very high power CO_2 lasers because it allows excitation with relatively low voltages even in long cavities.

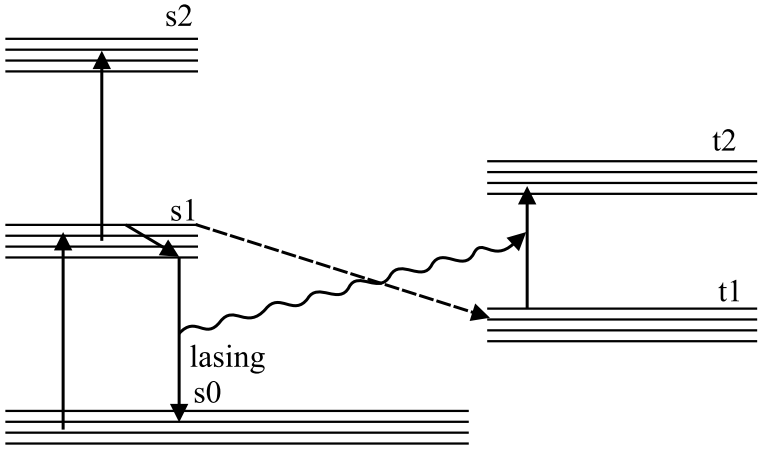

Figure 7.17: Energy level diagram of a typical liquid dye laser.

The energy levels are so called singlet and triplet states. In a singlet state the total electron spin of an excited molecule is zero whereas in a triplet state it is unity. Each electronic level is broadened by molecular vibration and rotation modes. Therefore absorption and emission spectra are broad enabling a significant laser emission range. Population inversion is easily reached with small excitation and high gain is possible. After excitation $S_0 \rightarrow S_1$ a fast decay is found for the transition back to the S_0 levels. At the same time T_1 is filled. The transition is not allowed one from the selection rules of quantum mechanics but nevertheless a small transition probability exists. Unfortunately the transition $T_1 \rightarrow T_2$ is at the same energy as $S_1 \rightarrow S_0$ and after some time of operation significant absorption of the emitted laser light occurs and limits the laser efficiency when enough electrons have piled up in T_1. To prevent the reabsorption the liquid dye is mechanically pumped through the active zone thereby refreshing the gain material and preventing the bleaching effect. Dye lasers are usually optically pumped and find application in spectroscopy and study of chemical reactions due to their superior tunability over significant wavelength ranges.

7.7.3 Parametric Lasers

Parametric lasers are based on the principle of nonlinear signal generation as described in section 5.9. Basically an optical nonlinearity produces two new frequencies from the pump signal. The idler frequency is not regarded here. When such a material is placed inside a resonator with resonance at the signal wavelength the system is called an optical parametric oscillator (OPO). The OPO resonator must be transparent for the incident pump signal and for the idler to prevent parasitic idler lasing. Figure 7.18 shows a schematic design of an OPO.

First examples of OPO's used nonlinear crystals. Typically BBO (barium beta borate), KTP (potassium titanyl phosphate), KNB (potassium niobate), and LBO (lithium triborate) are used. They are birefringent and the polarization of the pump is chosen to be on the extraordinary axis. Signal and idler occur on the ordinary axis and

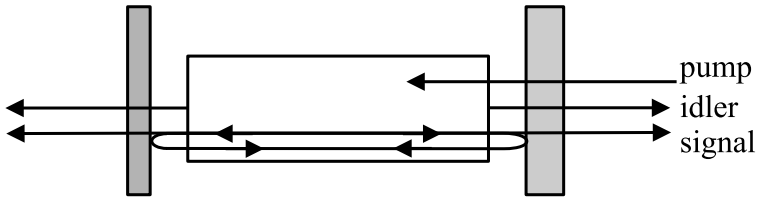

Figure 7.18: Setup of an optical parametric oscillator. The resonator is tuned to the signal wavelength and transparent for the pump and idler.

momentum conservation requires

$$\bar{n}_{\mathrm{p}}\omega_{\mathrm{p}} = \bar{n}_{\mathrm{e}}\omega_{\mathrm{p}} = \bar{n}_{\mathrm{o}}(\omega_{\mathrm{s}} + \omega_{\mathrm{i}})$$

besides energy conservation

$$\omega_{\mathrm{p}} = \omega_{\mathrm{s}} + \omega_{\mathrm{i}} \quad .$$

With a small change in the pump direction the refractive index for the pump $\bar{n}_{\mathrm{p}}$ approaches $\bar{n}_{\mathrm{o}}$ and momentum conservation requires that ω_{s} and ω_{i} are changed. This effect is used for tuning the wavelength of the OPO.

Today so called Raman fiber amplifiers are becoming increasingly interesting for optical data communication. At high light intensity the quartz fibers become nonlinear. One main reason is the Raman effect that normally is responsible for attenuation. But the Raman effect can also be used to amplify small signals when a pump signal generates the optical nonlinearity and transfers its energy to the small signal. This case can be described by (5.8) which has to be written in the more general form (χ_2 does not exist in glass fibers)

$$P\{\omega_{\mathrm{s}}\} = \varepsilon_0 \left(\chi\{\omega_{\mathrm{s}}\}E\{\omega_{\mathrm{s}}\} + \chi_3\{\omega_1, \omega_2, \omega_3\}E\{\omega_1\}E\{\omega_2\}E\{\omega_3\}\right) \tag{7.18}$$

with energy conservation $\omega_s = \omega_1 + \omega_2 + \omega_3$. In case $\omega_1 = -\omega_2 = \omega_p$ and $\omega_3 = \omega_s$ the polarization

$$P\{\omega_s\} = \varepsilon_0 \left(\chi\{\omega_s\} + \chi_3\{\omega_p, -\omega_p, \omega_s\}|E\{\omega_p\}|^2\right) E\{\omega_s\} \tag{7.19}$$

follows due to $E\{\omega\} = E^*\{-\omega\}$. The third order susceptibility χ_3 is of remarkable magnitude in a relatively broad range $\Delta\omega$ around ω_p such that different signals ω_s can be amplified with nearly the same amplification. This makes the Raman amplifier useful for signal recovery in wavelength division multiplexing systems.

7.7.4 Free-Electron Lasers

In a free electron laser the gain is produced by an electron beam that intersects with the resonator axis periodically. The electrons of the beam couple with the electric field of the optical standing wave, enlarging it resonantly. Initial photons are supplied from other laser sources such as a pulsed CO_2 laser for example. The electron beam consists of relativistic electrons (velocity close to speed of light) and is undulated with so called wiggler magnets as sketched in figure 7.19.

The magnets provide a periodically changing magnetic field direction which accelerates the electrons thereby producing the swinging trace with spatial periodicity λ_w. The observed gain peaks at a value of

$$\lambda_p = \frac{\lambda_w}{2\gamma^2} \tag{7.20}$$

where $\gamma = E_{tot}/E_{rest}$ is the ratio of the total electron energy to its rest energy (0.1511 MeV). This kind of gain production is called magnetic radiation. The bandwidth obtained is $\Delta\lambda/\lambda = 1/2N$ where N is the number of wiggler periods. As can be seen the wavelength is easily tuned by changing the electron beam energy provided. This gives the ability to realize a tunable laser with a wide spectrum. When the electron energy exceeds the required resonance energy the electrons transfer the excess energy

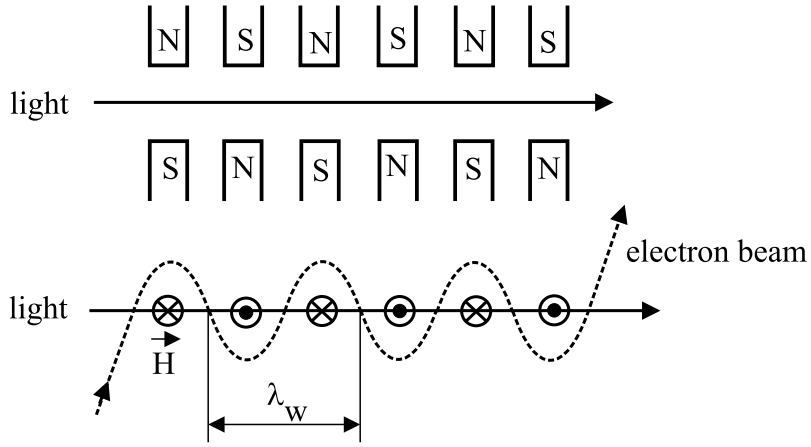

Figure 7.19: Principal setup for a free electron laser. The wiggler magnets provide that the electron beam intersects the light beam periodically transferring its excess energy to the light beam.

to the light beam. The problem is to provide electrons with just the right energy over the whole path. A solution of this problem is tried by several approaches but is not managed currently. The other problem is that the size of free electron lasers is enormous due to the required magnets and the electron beam source which has to produce relativistic electrons.

7.7.5 Doped Insulator Lasers

We use the term doped insulator laser for solid state type lasers whose active medium consist of a host material (amorphous or crystalline) with active ions. The ions typically are rare earth or transition metal elements and are intentionally incorporated (doped) impurities in the host material. For good lasers the host must

- be of good and homogeneous quality,

- show low thermal expansion and low temperature dependence of refractive index,

- be easily doped by addition of impurities to the melt of an amorphous host or be

easily grown with lattice sites that are ready to be substituted by the dopand.

All types of doped insulator lasers must be optically pumped. Typical examples are Nd:YAG, alexandrite, YLF, Ti:sapphire and silicate glass lasers. The first operating solid state laser was a chromium doped ruby laser but is of no importance today. As an example for the operation principle the important Nd:YAG laser is taken. The YAG (yttrium-aluminum-garnet) host does not participate directly in the operation - neither in pumping nor in lasing. But its crystal field modifies the Nd ion energy levels in a way that population inversion becomes more likely. Optical pumping of the Nd is very efficient at 808 nm and the most likely radiating transition takes place at 1060 nm. There are as well some other transitions at longer wavelength but with much lower efficiency. Neodymium can also be incorporated in glass as host material and behaves almost the same way as in YAG. This allows to produce glass fibers with Nd doping which can be used as amplifiers for 1.06 μm light. In optical telecommunications, 1.3 μm and 1.55 μm wavelengths are more preferred and erbium is the rare earth metal of choice for usage in fiber amplifiers for 1.55 μm. Optical pumping of Er is efficient at 980 nm and 1480 nm. Spontaneous and stimulated emission occurs in the spectral region from 1530 nm to 1560 nm. Optical pumping is typically achieved by coupling of 980 nm light into the fiber. The problem is to couple high pumping intensities into the relatively small core. A solution to this problem is a double core fiber as depicted in figure 7.20. The inner core has the same diameter as a standard singlemode fiber and is erbium doped. The outer core is a waveguide (multimode) for the pumping light, providing relatively effective and homogeneous pumping over the whole amplifier length. Erbium doped fiber amplifiers (EDFA's) are widely used in today's long distance transmission systems.

For 1.3 μm transmission systems EDFA's are not suitable. In this case the erbium is substituted by praseodymium which provides amplification around 1300 nm. Also other rare earth metals are possible for usage in fiber lasers and amplifiers as listed in table 7.1.

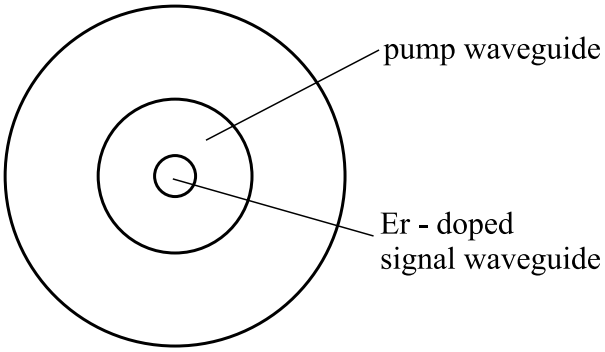

Figure 7.20: Two core erbium doped fiber amplifier. The inner core is the signal waveguide and provides amplification. The outer core guides the pumping light and serves for effective and homogeneous pumping of the amplifier.

Table 7.1: Dopands for fiber lasers and their possible transitions.

Element	λ (μm)	Element	λ (μm)	Element	λ (μm)
Nd	0.38	Tm	0.80	Nd	1.34
Tm	0.455	Er	0.85	Ho	1.38
Tm	0.48	Pr	0.88	Tm	1.47
Pr	0.491	Pr	0.91	Er	1.55
Pr	0.52	Nd	0.94	Er	1.66
Er	0.546	Yb	0.975	Er	1.72
Ho	0.55	Er	0.98	Tm	1.9
Pr	0.61	Yb	1.04	Ho	2.04
Pr	0.635	Nd	1.06	Tm	2.3
Sm	0.651	Pr	1.08	Er	2.75
Pr	0.695	Ho	1.2	Ho	2.9
Pr	0.715	Pr	1.31	Er	3.4

The above mentioned alexandrite and Ti:Sapphire lasers are so called vibronic lasers which provide operation over a wide range of wavelengths resulting into a tunable output or ultra-short pulse output. The reason is that the energy levels of the impurities are split into a band of vibrational sub-levels corresponding to the discrete lattice vibration energies. In the alexandrite ($BeAl_2O_4$) host chromium serves as impurity as is often found in vibronic lasers. It is worth to note that chromium in the case of the ruby (crystalline Al_2O_3) laser has discrete levels in contrary to the alexandrite laser which tells us that the presence of chromium alone does not ensure a tunability.

7.7.6 Semiconductor Lasers

As pointed out before a population inversion zone is necessary for laser operation. This zone is also called active region. In semiconductors such a zone can be produced by a pn-junction with heavy doping of both the p-type and the n-type material (degenerate semiconductors) such that the Fermi levels are separated sufficiently wide. Under forward bias with $qV_\mathrm{a} \geq W_\mathrm{g}$ the depletion zone is filled by carriers from both sides and population inversion is achieved as sketched in figure 7.21.

Light generated in the active region travels to the edges of the semiconductor crystal and is partly reflected there. A good estimation for the reflected power can be found from the Fresnel reflection equation $R = \left(\frac{\eta_2-\eta_1}{\eta_2+\eta_1}\right)^2$ which results into a reflectivity of about 32 % for a GaAs surface ($\eta_2 = 3.6, \eta_1 = 1$). This value is sufficient to form a resonator with the semiconductor edges. Two problems prevent effective operation of such so called homojunction lasers: first the carriers injected into the depletion zone are not encouraged to stay there and may leave it towards the other side where they recombine. This leads to an ineffective use of the current flowing through the active region and is accounted for with an internal current efficiency η_I. The internal current efficiency is the amount of current that contributes to by carrier recombination in the active region. Leakage currents that do not flow through the active region and the above mentioned current that is carried by recombination outside the active region are

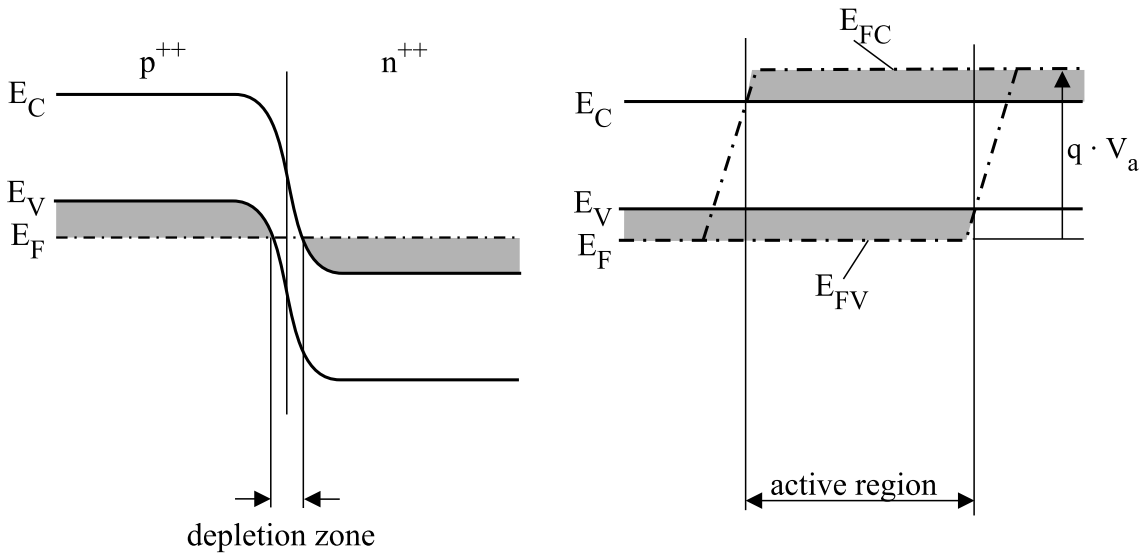

Figure 7.21: Degenerate pn-junction in equilibrium and with applied voltage. Under forward bias carriers are injected into the depletion zone generating population inversion.

incorporated in $\eta_I = \frac{I_{\text{active}}}{I_{\text{total}}}$. Figure 7.22 shows the different current paths inside a laser. The second problem attributed to homojunction lasers is that there is no waveguiding [1] of the lasing mode(s) so that diffraction occurs during propagation in the resonator. Both problems can be fixed by employment of so called double heterojunction lasers.

Due to their wide usage in different applications semiconductor lasers are taken as an example for the description of laser operation by rate equations in the next section. Also discussed there are the different types of semiconductor lasers with some emphasis on the design and special design criteria to achieve efficient operation.

[1] The depletion zone has much less free carriers than the neighboring ones resulting into a slightly higher refractive index with poor confinement.

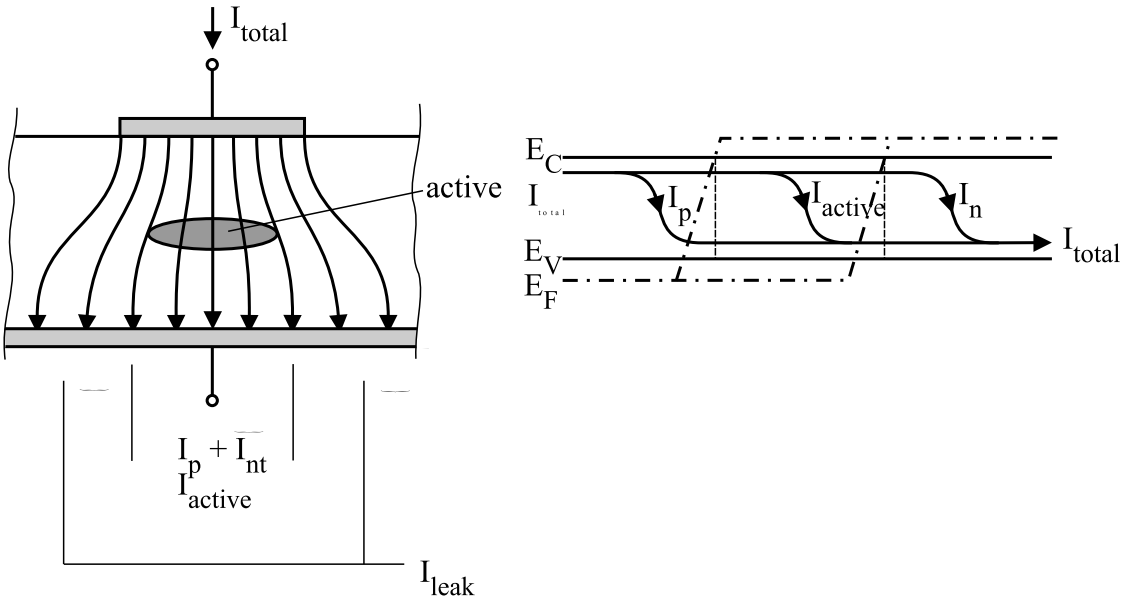

Figure 7.22: Current paths through a laser diode. Besides the current that is carried by recombination in the active zone, recombination currents in the p- and n-doped regions and a current that does not flow through the active region have to be considered. All currents that do not contribute to the recombination in the active region are lost for laser operation and are accounted for in the internal current efficiency η_I

7.8 Heterostructure Semiconductor Lasers

All modern semiconductor lasers employ heterostructures which introduce a region of higher refractive index as well provide an efficient carrier confinement in the active region. Both properties are not met naturally at the same time and in special cases waveguiding for optical confinement must be provided by another heterostructure than the carrier confinement. In some cases only the geometrical dimensions require two different heterostructures for good optical and electronic confinement. In the AlGaAs/GaAs system the refractive index increases with decreasing bandgap making this material system suitable for double heterostructure (DH) lasers with good optical and electronic confinement. We will take this material system for explanation of the double heterostructure laser operation.
Beginning with a discussion on principle laser structures involving quantum effects besides simple DH lasers the description of laser operation by rate equation follows. Modulation properties for signal transmission and noise characteristics are objectives of the next sections followed by a presentation of today's semiconductor laser design and application.

7.8.1 Basic Principles of Laser Structures

For efficient operation three basic requirements have to be met in a laser structure:

1. The current flowing through the laser should flow only through the active region. A current bypass (leakage current) leads to a so called lower current confinement factor Γ_{I}.

2. Carriers once injected into the active region should completely recombine there. When carriers escape from the active region before recombination they do not support any more the stimulated emission and are lost for laser operation. Carrier recombination in the active zone is described by the carrier confinement factor Γ_{carr} .

3. The optical wave should use the available gain effectively. Normally the wave extends over bigger regions than gain is present which is described by the optical confinement factor Γ. The optical confinement factor was the first regarded in semiconductor lasers and is also referred to only as confinement factor.

If the gain is resulting from a quantum structure the free carrier wave functions sometimes extend remarkably over the geometric structure. In that case carrier escape from the quantum structure becomes likely and the attainable gain is lowered. The confinement of the carrier wave function to the quantum structure is described by an electronic confinement factor Γ_{el}. The electronic confinement is not used directly in the mathematical description of lasers by rate equations but is incorporated in gain calculations.
Both the current confinement factor and the carrier confinement factor lead to the definition of a current efficiency

$$\eta_{\mathrm{I}} = \Gamma_{\mathrm{I}}\Gamma_{\mathrm{carr}} \quad .$$

Attention has to be paid to this point because in several text books the current efficiency is named η_{i} which may lead to confusion with the internal quantum efficiency of a LED. The difference is that the conversion of free carriers to photons by spontaneous emission is described in η_{i} thereby regarding non-radiative recombination effects. Additionally no emphasis is made on the position in the device where the recombination takes place (stimulated emission only takes place in the active region, spontaneous emission occurs as well outside the active region). In contrary to this η_{I} describes the amount of current that recombines inside the active region regardless whether the recombination is radiative or non-radiative.

7.8.1.1 Double Heterostructures

A typical double heterostructure consists of GaAs embedded in AlGaAs. The schematical cross section of a double heterostructure laser is shown in figure 7.23

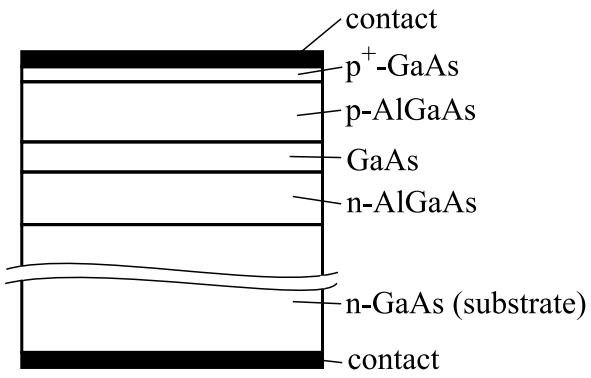

Figure 7.23: Schematic cross section through a double heterostructure AlGaAs/GaAs laser. Even with non degenerately doped semiconductors population inversion is achieved in the GaAs under forward bias. The GaAs substrate and p^+ GaAs contact layer are present only for technological reasons and do (nearly) not participate in the laser operation.

On a GaAs substrate a thick AlGaAs layer is grown epitaxially followed by the active GaAs layer and the second AlGaAs and a highly doped GaAs contact layer. Both the substrate and the contact layer are well separated by the AlGaAs layers from the active region to prevent reabsorption of the emitted light. The AlGaAs exhibits a lower refractive index and a bigger bandgap than the GaAs. Under forward bias a well is formed for both holes and electrons in the GaAs. The band discontinuities prevent injected carriers effectively from escaping the active region. Assuming $10^{18}\,/\mathrm{cm}^3$ carriers in the active region at room temperature as realistic value for laser operation a barrier of 0.3 eV ($\simeq \Delta W_\mathrm{C} \simeq \Delta W_\mathrm{V}$ for AlGaAs/GaAs heterojunctions) results into $4 \cdot 10^{14}\,/\mathrm{cm}^3$ carriers that are able to leave the well. For better light guiding so called buried heterostructure (BH) lasers as sketched in figure 7.24 are made which guide the light in two dimensions by index guiding.

The attainable gain in a double heterostructure is depicted in figure 7.25.

The peak gain g_p grows nearly linearly with injected carrier density

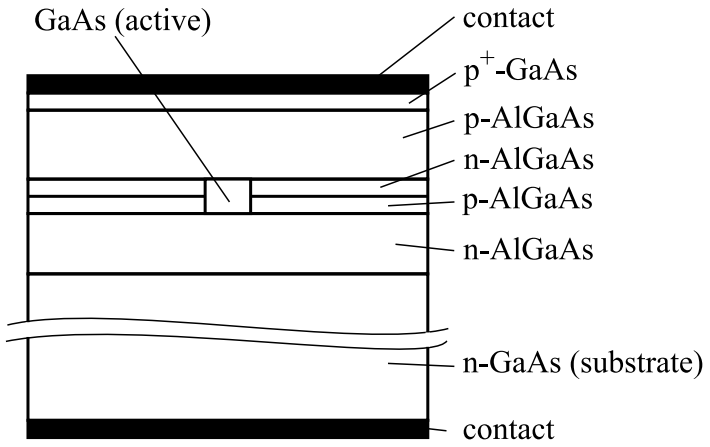

Figure 7.24: Buried heterostructure laser with index guiding. Due to the inverse pn-junction next to the active region leakage current is effectively suppressed. The energy barriers at the double heterostructure effectively prevent the injected carriers from leaving the active region. Both effects result into high internal current efficiencies η_I.

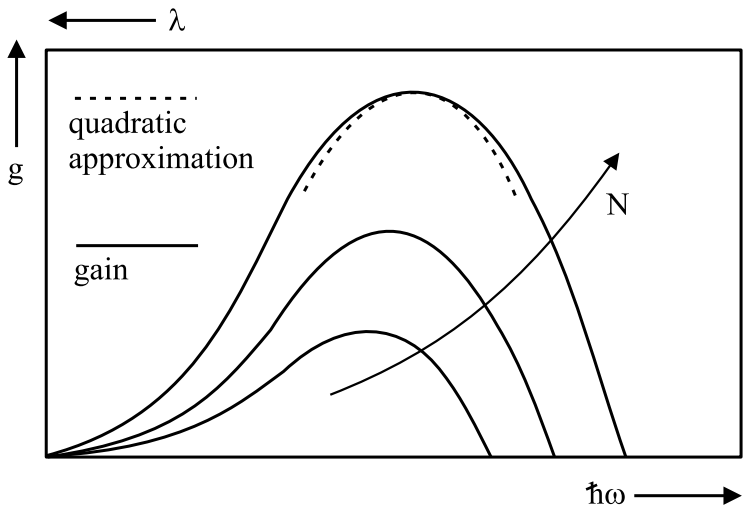

Figure 7.25: Typical gain spectrum in bulk GaAs as function of carrier density and wavelength.

$$g_{\rm p} = g\{\lambda_{\rm p}, n\} = a(n - n_{\rm t})$$

as long as the transparency density $n_{\rm t}$ is compensated. At the peak gain the wavelength dependence can be described by a quadratic approximation

$$g\{\lambda\} = g_{\rm p} - \delta_\lambda^2(\lambda - \lambda_{\rm p})^2$$

with peak gain wavelength $\lambda_{\rm p}$ such that the gain can be described as

$$g\{\lambda, n\} = a(n - n_{\rm t}) - \delta_\lambda^2(\lambda - \lambda_{\rm p})^2 \quad .$$

The differential gain coefficient a is of the order $a \simeq((100/\text{cm})/(10^{18}/\text{cm}^3))$ and the transparency density is around $n_{\rm t} \simeq 10^{18}/\text{cm}^3$.

7.8.1.2 Quantum Wells, Wires and Dots

When the thickness of the small gap layer inside a double heterojunction decreases more and more it approaches the scattering length ($\simeq 50\,\mathrm{nm}$ at room temperature in GaAs) de Broglie wavelength (in thermal equilibrium at room temperature for electrons $\lambda_{\mathrm{el}} \simeq 30\,\mathrm{nm}$ and for holes $\lambda_{\mathrm{h}} \simeq 10\,\mathrm{nm}$).

At that point the description of carriers in the heterostructure must consider quantum effects. The carriers must be taken as particles or waves depending on the property to describe. The energy of a quasi free carrier in a semiconductor consists of the potential energy $W_{\mathrm{pot}} = W_{\mathrm{v}\nu}$ or $W_{\mathrm{c}\nu}$ and the kinetic energy $W_{\mathrm{kin,el}} = \frac{\hbar^2 \cdot k_{\mathrm{el}}^2}{2m_{\mathrm{el}}}$ or $\frac{\hbar^2 \cdot k_{\mathrm{h}}^2}{2m_{\mathrm{h}}}$ for electrons and holes with wavenumber $k = \frac{2\pi}{\lambda}$.

The de Broglie wavelength describes the wavelength that results from the kinetic energy by thermal exitation. In a double heterostructure with very narrow small-gap zone the potential energy of a carrier changes abruptly when it travels at the same total energy from one side to the other side of the heterojunctions. This implies an abrupt change in the kinetic energy or in classical physics an abrupt change in speed which is not possible. A better description makes use of the wave picture in that case. The steady state behaviour can be described by the time independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\Delta\psi\{\vec{r}\} + W_{\mathrm{pot}}\{\vec{r}\} \cdot \psi\{\vec{r}\} = W \cdot \psi\{\vec{r}\}$$

where $\psi\{\vec{r}\}$ denotes the wave function and W is the total energy of the carrier. Only some discrete values of W are allowed in a thin double heterostructure.

Mathematically such discrete values are called eigenvalues of the differential equation and the corresponding ψ are eigenfunctions. The Schrödinger equation has the form of the well known wave equation leading to the name wavefunction for ψ. As in the case of electromagnetic waves the wave function has to satisfy boundary conditions when the potential energy switches at a certain position $\vec{r}_0$ to another value. The boundary conditions are $\psi\{\vec{r}_0\} = \text{const.}$ and $\nabla\psi \circ \vec{n} = \text{const.}$ with $\vec{n}$ normal to the boundary which follows from mathematical considerations when the Schrödinger equation is (for-

mally) integrated across the boundary where the potential finitely changes. Solutions in a so called square potential (W_{pot} is lowered by U over distance $2a$) are similar to solutions for optical waves in a planar film as sketched in figure 7.26. The film wave travels with $\exp\{i\beta z\}$ and must satisfy

$$\frac{\partial^2}{\partial x^2}E - \eta^2\{x\}k^2E = -\beta^2 E \quad .$$

The wavefunction in the well

$$\frac{\partial^2}{\partial x^2}\psi - \frac{2m}{\hbar^2}W_{\text{pot}}\{x\}\cdot\psi = -\frac{2m}{\hbar^2}W\psi$$

has the same form. It is common to depict the wavefunction inside the potential structure with baseline at the eigenvalue W as shown in figure 7.26.

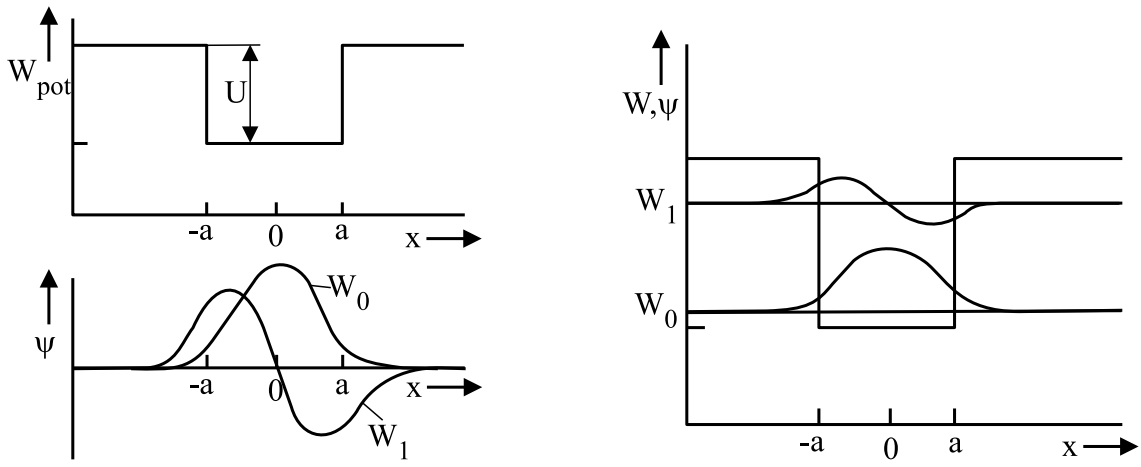

Figure 7.26: Square potential and corresponding wavefunctions. It is common to show the wavefunctions inside the potential structure with baselines at the eigenvalue to separate them.

The formation of eigenvalues says that carriers are only allowed inside the quantum structure when they have a certain energy. The question is how many states are available at the eigenvalues. The density of states in bulk material is $D\{\vec{k}\}\,\mathrm{d}^3\vec{k} = \frac{2}{(2\pi)}\,\mathrm{d}^3k$ in k space. In semiconductors with parabolic band structure

$$W_{\mathrm{B}} = W \pm \frac{\hbar^2}{2m_{\mathrm{c}}} \|\vec{k}\|^2$$

the density of states as function of energy is the well known

$$D_{\mathrm{B}}\{W\}\,\mathrm{d}W = \frac{\sqrt{2m_{\mathrm{B}}^3}}{\pi^2\hbar^3}\sqrt{\pm\frac{W_{\mathrm{B}}-W}{kT}}\,\mathrm{d}\left(\frac{W}{kT}\right) = N_{\mathrm{B}}\cdot\frac{2}{\sqrt{\pi}}\sqrt{\pm\frac{W_{\mathrm{B}}-W}{kT}}\,\mathrm{d}\left(\frac{W}{kT}\right)$$

where the index B has to be replaced by V for valence band and C for the conduction band. The carrier mass m_{B} is (heavy or light) hole and electron mass respectively. The addition is valid for valence and the subtraction for conduction band. For more details refer to appendix F.

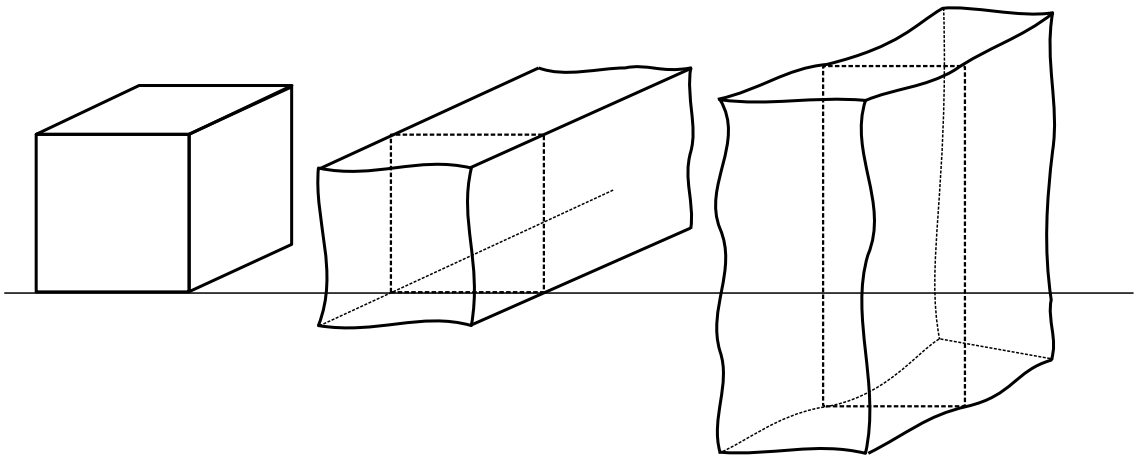

Figure 7.27: Rectangular quantum box, wire and well. The cartesian coordinate system is oriented parallel to the edges of the quantum structures.

In quantum structures the density of states changes dramatically. It is common to distinguish between the three types of quantum structures as sketched in figure 7.27, namely the quantum well, quantum wire and quantum box. The calculation of eigenvalues in cartesian coordinates is very easy when the energy discontinuity U on the boundaries approaches infinity. With dimensions a_{x}, a_{y} and a_{z} in x- , y - and z - direction the allowed k values are $k_{\mathrm{x}m} = m\cdot\frac{2\pi}{a_{\mathrm{x}}}$, $k_{\mathrm{y}n} = n\cdot\frac{2\pi}{a_{\mathrm{y}}}$ and $k_{\mathrm{z}\ell} = \ell\cdot\frac{2\pi}{a_{\mathrm{z}}}$. In the conduction band the density of states results (see appendix F)

$$D_{\mathrm{C}}\,\mathrm{d}W = \begin{cases} \sum\limits_{m} \frac{2}{a_{\mathrm{x}}} \frac{1}{4\pi} \frac{2m_{\mathrm{e}}}{\hbar^2}\,\mathrm{H}\{W - W_{\mathrm{C}m}\}\,\mathrm{d}W \\ \sum\limits_{m}\sum\limits_{n} \frac{2}{a_{\mathrm{x}}a_{\mathrm{y}}} \frac{1}{4\pi} \sqrt{\frac{2m_{\mathrm{e}}}{\hbar^2}}\,(W - W_{\mathrm{C}m,n})^{-1/2}\,\mathrm{H}^2\{W - W_{\mathrm{C}m,n}\}\,\mathrm{d}W \\ \sum\limits_{m}\sum\limits_{n}\sum\limits_{\ell} \frac{2}{a_{\mathrm{x}}a_{\mathrm{y}}a_{\mathrm{z}}} \delta\{W - W_{\mathrm{C}m,n,\ell}\}\,\mathrm{H}^2\{W - W_{\mathrm{C}m,n,\ell}\}\,\mathrm{d}W \end{cases}$$

for quantum wells, wires and dots respectively as sketched in figure 7.28.

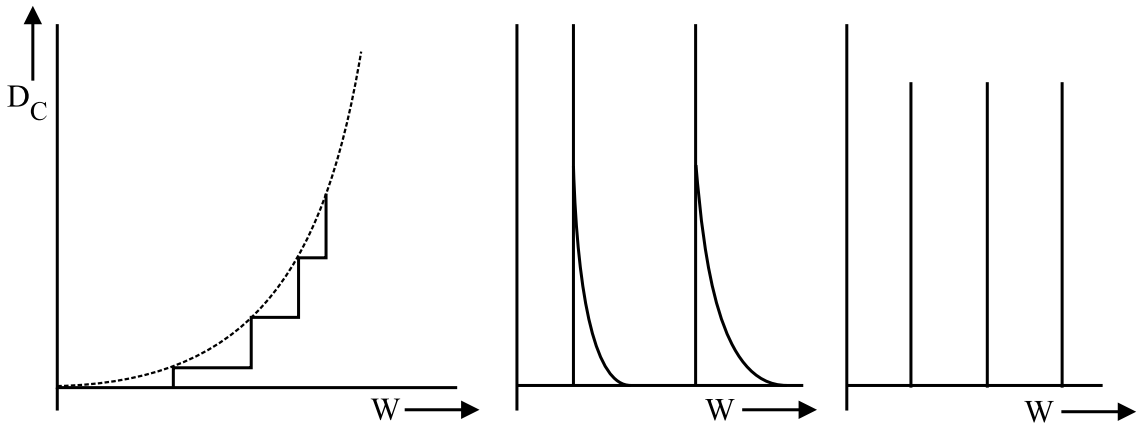

Figure 7.28: Density of states versus Energy in a quantum well, quantum wire and quantum box. The dashed line shows the density of states in the corresponding bulk material.

When the density of states in a quantum well is compared to the corresponding bulk material at energy level

$$W = W_{\mathrm{C}m} = \frac{\hbar^2}{2m_{\mathrm{e}}} \left(\pi \frac{m}{a_{\mathrm{x}}}\right)^2 + W_{\mathrm{C}}$$

the density of states in the quantum well is

$$D_{\mathrm{C}m}\,\mathrm{d}W = \frac{m_{\mathrm{e}}}{\pi\hbar^2} \frac{m}{a_{\mathrm{x}}}\,\mathrm{d}W$$

and in the bulk material

$$D_{\mathrm{C}}\,\mathrm{d}W = \frac{\sqrt{2m_{\mathrm{e}}^3}}{\pi^2\hbar^3}\sqrt{\frac{\hbar^2}{2m_{\mathrm{e}}}\left(\frac{m\pi}{a_{\mathrm{x}}}\right)^2}\,\mathrm{d}W = \frac{m_{\mathrm{e}}}{\pi\hbar^2}\frac{m}{a_{\mathrm{x}}}\,\mathrm{d}W = D_{\mathrm{C}m}\,\mathrm{d}W \quad .$$

This result says that the density of eigenstates in the bulk material and a quantum well are equal at the energy eigenvalues of the quantum well. It seems that no states are lost in the quantum well but they are (energetically) later occupied than in the bulk material. This is true for the density of states but not for the total number because the total number of states follows after spacial integration. As discussed in appendix F for a comparison the barriers of the quantum structures have to be taken into account which lowers the available number of states in the quantum structure compared to the bulk material.

The discontinuous behaviour of the density of states leads to a modified gain description in a quantum well

$$g = g_0 \cdot \ln\{\frac{N}{N_{\mathrm{t}}}\}$$

with differential gain

$$\frac{\partial}{\partial N}g = \frac{g_0}{N} \quad .$$

For comparison the differential gain is taken at transparency density leading to the linear gain coefficient $a = g_0/N_{\mathrm{t}}$ valid in bulk material. Table 7.2 gives some typical values for g_0 and N_{t} in different material systems.

As can be seen clearly the gain is remarkably higher in the quantum wells than in the bulk material for carrier densities near threshold and slightly above corresponding to better high frequency performance in quantum well lasers. Figures 7.29 an 7.30 show calculated gain curves of different quantum structures and the corresponding peak gain in different material systems.

The InGaAs quantum well in table 7.2 is not lattice matched to the GaAs host material leading to a strain in the quantum well. Table 7.3 shows a small part of the periodic system of elements. In the InGaAs quantum well Ga is partly substituted by In which

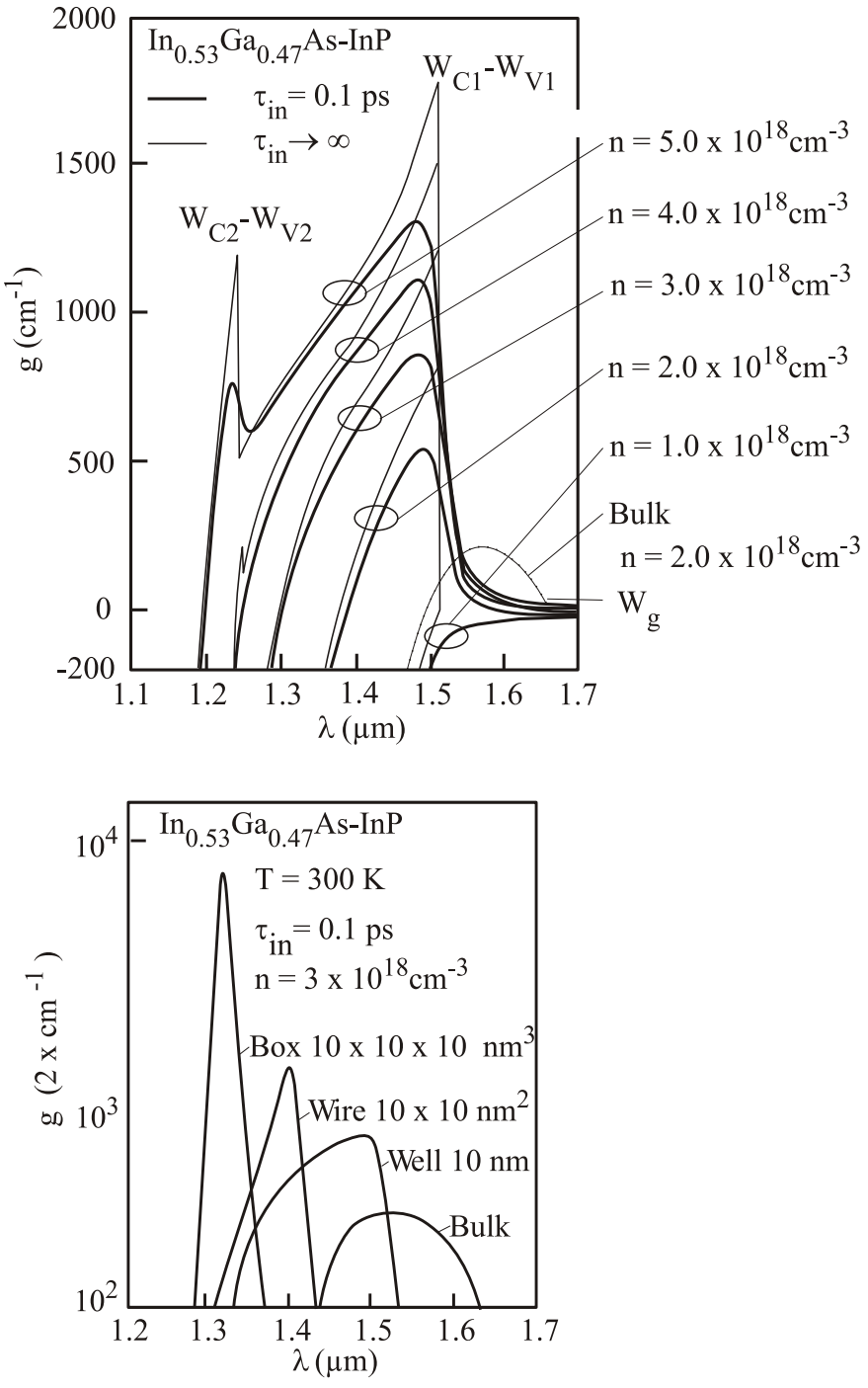

Figure 7.29: Calculated gain curves of quantum wells, wires and boxes. The upper figure shows the gain spectrum of a 10 nm InGaAs-Inp quantum well for different carrier densities. In the lower figure the gain spectra of different quantum structures at constant carrier densities are compared.

Table 7.2: Transparency density and corresponding gain in quantum wells of different material systems. For comparison the corresponding values for bulk GaAs are given.

Material	$a_x 7\,\mathrm{nm}$	$N_t/10^{18}\,\mathrm{cm}^3$	$g_0/\,\mathrm{cm}^{-1}$
GaAs	∞	1.85	1500
$Al_{0.2}Ga_{0.8}As/GaAs$	8	2.6	2400
$In_{0.2}Ga_{0.8}As/GaAs$	8	1.8	2100

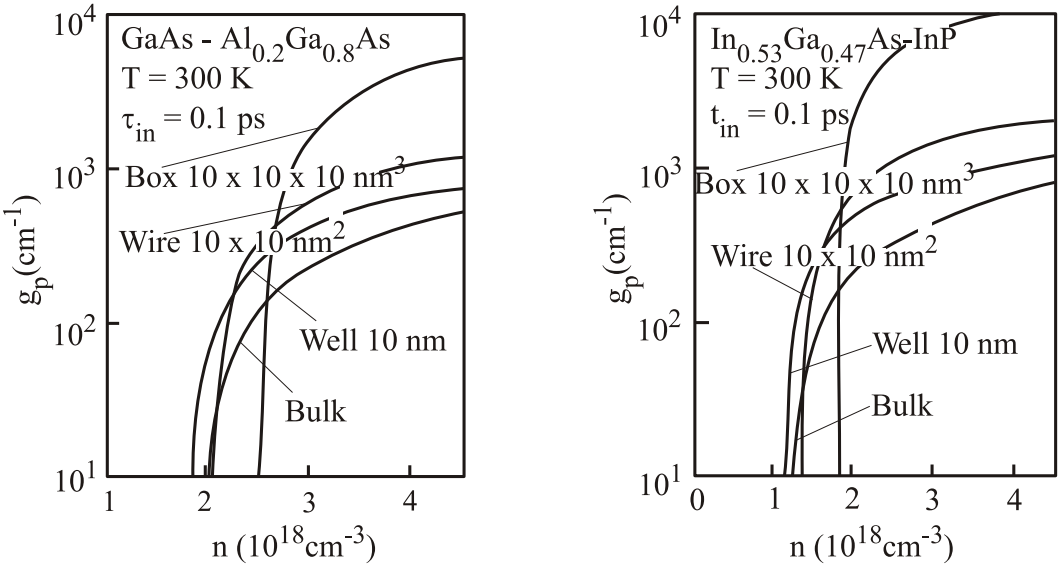

Figure 7.30: Peak gain of quantum wells, wires and boxes in the InGaAs-InP and GaAs-AlGaAs material systems.

is larger. This leads to a bigger lattice constant $\hat{a}_\mathrm{e}$ of the epitaxial layer than that of the GaAs substrate $\hat{a}_\mathrm{s}$. The difference is compensated by elastic compression of the InGaAs parallel to the heterointerfaces which in turn leads to an elastic expansion normal to the heterointerfaces as sketched in figure 7.31.

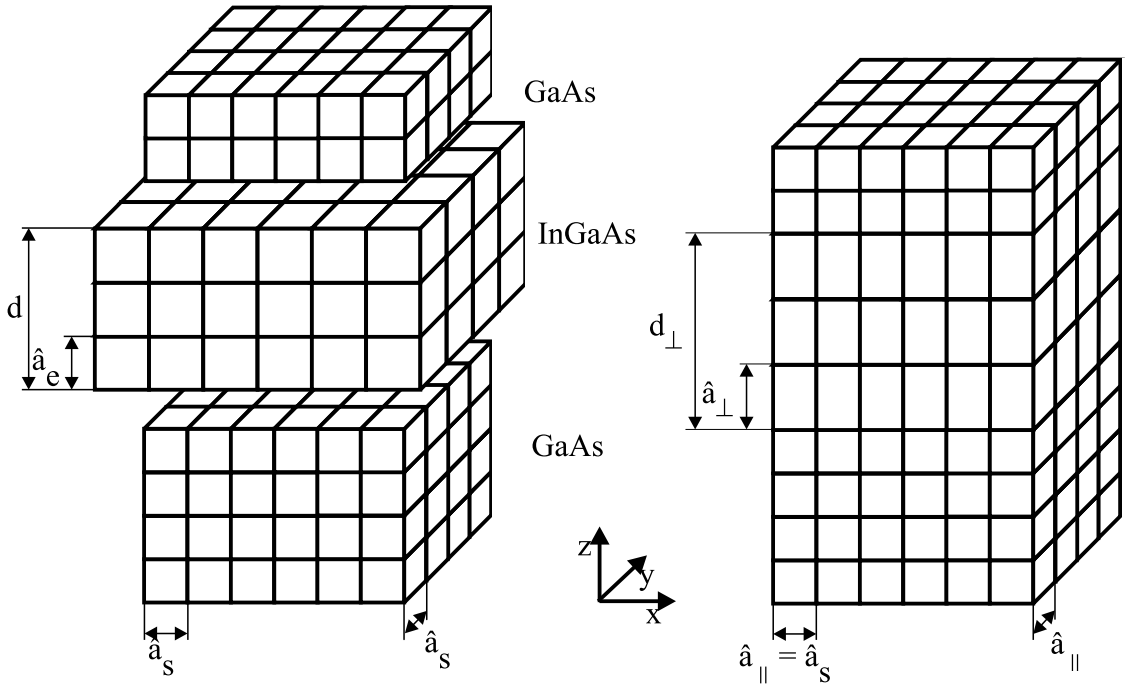

Figure 7.31: Compensation of non lattice matching in an epitaxial structure. The layers are elastically strained until the lattices parallel to the interfaces are matching leading to biaxial strain. The volume of the non matching crystal stays nearly constant which is achieved by additional elastic strain normal to the hetero interface.

The effect of elastic strain is described by the parallel-strain coefficient

$$\epsilon_\parallel = \frac{\hat{\mathrm{a}}_s - \hat{a}_\mathrm{e}}{\hat{a}_\mathrm{e}}$$

which is $\epsilon_\parallel < 0$ for the biaxial compressive strained GaAs-InGaAs-GaAs quantum well. From semiconductor physics research it is known that strain influences the electronic band structure. Especially the light and heavy holes are no longer degenerate in the Γ

Table 7.3: Part of the periodic system of elements displaying the elements today used in III-V semiconductors.

III	IV	V
B	C	N
Al	Si	P
Ga	Ge	As
In	Sn	Sb

point. This leads to two bandgaps $W_{\mathrm{g\,lh}}$ and $W_{\mathrm{g\,hh}}$ depending on the interaction of an electron with a light or heavy hole. The strain induce change in bandgap is expressed by δW_{h} for hydrostatic strain and δW_{s} for shear strain and leads to

$$W_{\mathrm{g\,lh}} = W_{\mathrm{g}} + \delta W_{\mathrm{h}} - \delta W_{\mathrm{s}} W_{\mathrm{g\,hh}} = W_{\mathrm{g}} + \delta W_{\mathrm{h}} + \delta W_{\mathrm{s}} \quad .$$

The hydrostatic strain induced change is proportional to the negative parallel-strain coefficient $\delta W_{\mathrm{h}} \sim -\epsilon_{\parallel}$ whereas the shear strain induced bandgap change is directly proportional to the parallel-strain $\delta W_{\mathrm{s}} \sim \epsilon_{\parallel}$. This leads for the compressively strained InGaAs to an enlarged bandgap for both hole types and even bigger bandgap for light holes than heavy holes. As can be seen in figure 7.32 the light hole valence band in the compressive strained bulk InGaAs lies already lower than in the adjacent GaAs. With additional quantum well modification of the band structure the bandgap is further enlarged.

In the strained quantum well the gain is remarkably changed compared to the unstrained situation. Unfortunately the elastic compensation of strain is only stable for thin layers. The strain leads to misfit dislocation (defects) under epitaxial growth when a certain thickness called critical thickness is reached. This can be explained by the energy stored in the elastic strain and the energy stored in the dislocations. When the strain energy is higher than the dislocation energy the crystal relaxes and forms dislo-

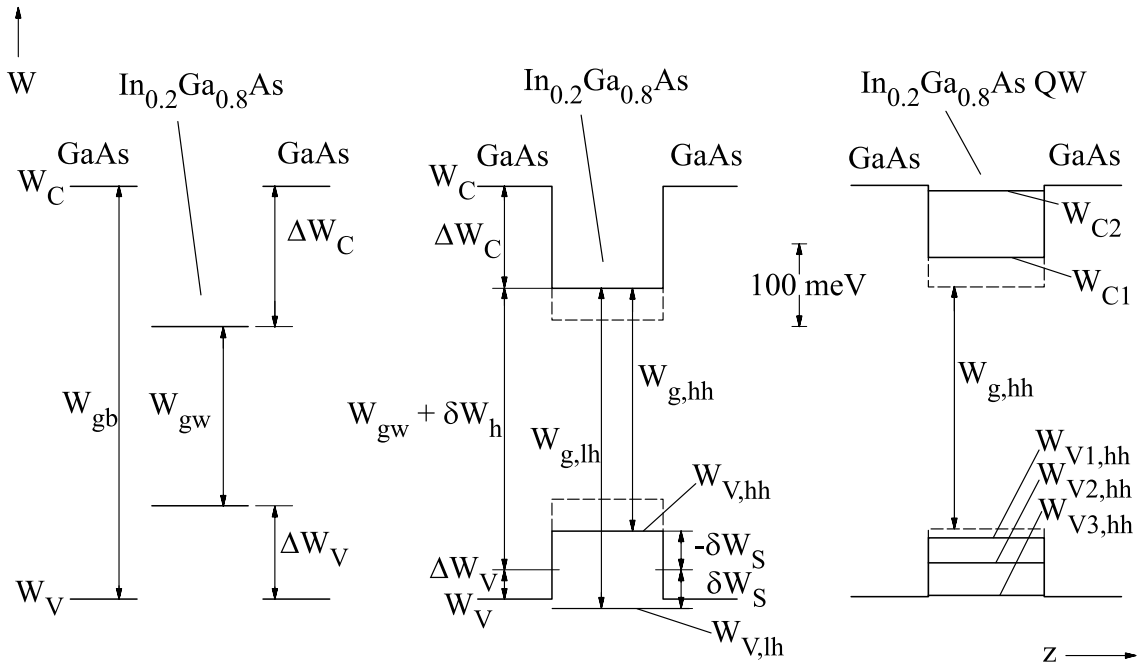

Figure 7.32: Formation of a compressively strained GaAs-InGaAs-GaAs double heterostructure and transition to a quantum well. On the left side the band structure of bulk GaAs and InGaAs without strain is sketched. In the double heterostructure the bands of the InGaAs are modified due to the biaxial compressive strain. Additional modification occurs when the thickness of the InGaAs decreases and a quantum well is formed.

cations. The defects cause a drastic degradation of the optical material properties. The critical thickness d_c for GaAs-In_x Ga_{1-x}As-GaAs heterostructures can be estimated by

$$d_c \cdot \epsilon_{\parallel} \simeq 20\,\mathrm{nm}\%$$

with

$$\epsilon_{\parallel} = \frac{-1}{1 + \frac{14}{x}} \simeq -7\% \cdot x$$

leading to

$$x \cdot d_c \simeq 3\,\mathrm{nm} \quad .$$

Figure 7.33 shows calculated gain spectra with respect to the light polarization relative to the quantum well. The light is called TE (transversal electric) when its electric field is oriented parallel to the boundaries of the quantum well. With electric field normal to the interfaces the relative polarization is called TM (transversal magnetic).
The optical gain in an unstrained layer is higher for the TE polarization and even increased for compressive strain where the TM gain is remarkably reduced. In contrary to this the TM gain increases with tensile strain and the TM gain decreases. This effect can be used in an alternating strain quantum well stack where the total gain of TE and TM polarization are equal and enlarged compared to the bulk material.
The reason for enhanced gain in strained quantum wells is found when the band structure is analyzed in more detail than is shown in figure 7.32. Calculated band structures for comparison of tensile, compressive and unstrained quantum wells are shown in figure 7.34. The InGaAsP-InGaAs-InGaAsP/InP quantum well is lattice matched for $x \simeq 0,47$ in the $In_{1-x}Ga_xAs$. With increasing x the bulk energy gap increases but the lattice constant decreases leading to tensile strain which in turn leads to an effective bandgap that is reduced compared to the bulk material. The strain leads to a decrease in the effective hole masses resulting into a decreased density of hole states. This en-

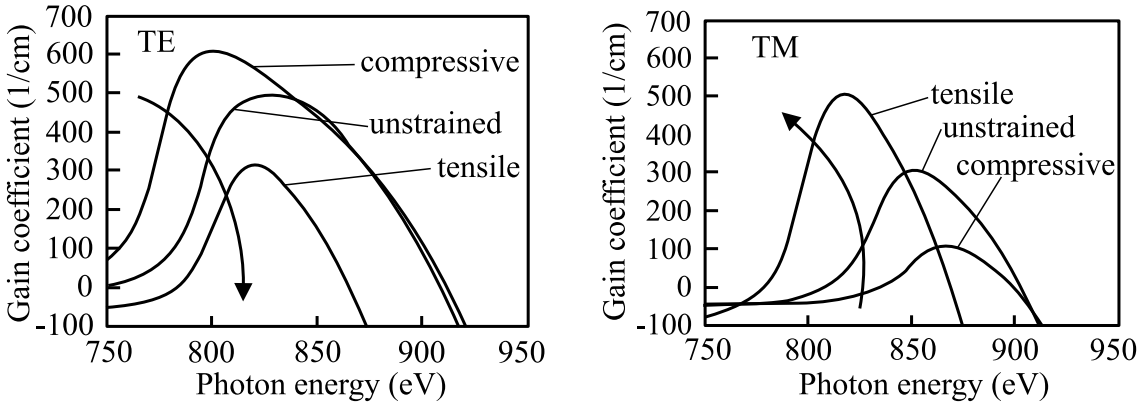

Figure 7.33: Influence of strain in InGaAsP-InGaAs-InGaAsP/InP quantum wells on the gain for TE and TM light polarization. The thickness of the quantum wells is adjusted for nearly equal peak gain wavelength.

ables population inversion with lower carrier densities resulting into lower threshold current densities compared to the unstrained quantum well.

Figure 7.34 shows calculated gain spectra for compressively strained GaAs-InGaAs-GaAs quantum wells with carrier density and temperature as parameter respectively. With increasing carrier density a shoulder forms on the short wavelength side which can be attributed to the second subband transition. The bandfilling shifts the gain spectra to longer wavelength (red shift) with higher pumping. With increasing temperature the carrier mobility inside the bands increases leading to carrier redistribution and decreasing gain. Additionally the bandgap of bulk InGaAs is shifted

$$\frac{\Delta W_{\mathrm{g}}}{\Delta T} \simeq -4 \cdot 10^{-4}\,\mathrm{eV/K}$$

corresponding to

$$\frac{\Delta \lambda_{\mathrm{p}}}{\Delta T} \simeq 0,32\,\mathrm{nm/K} \quad .$$

Figure 7.35 shows a calculated gain spectrum with quantum well thickness as parameter and the corresponding peak gain as a function of carrier density. From the upper

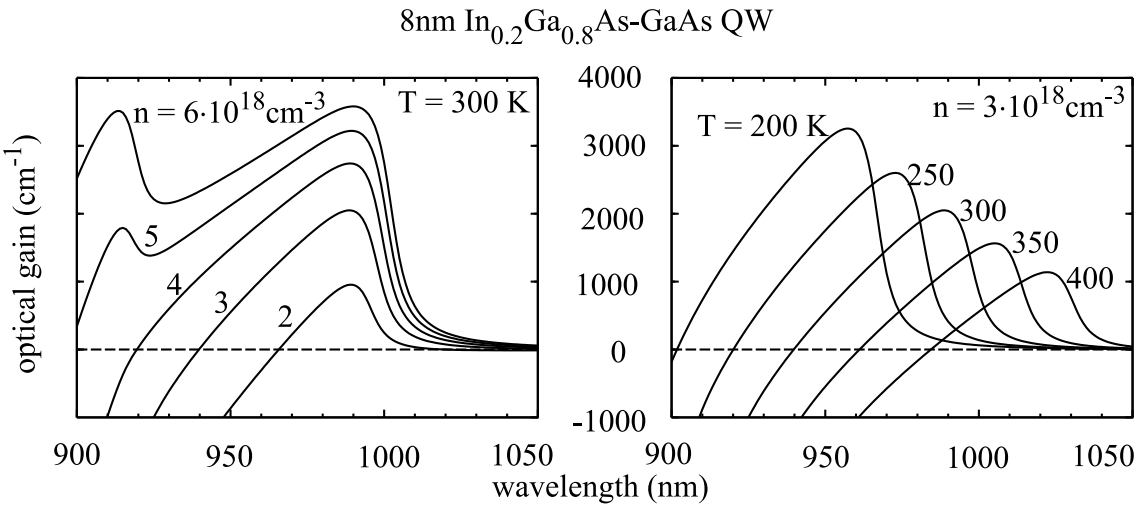

Figure 7.34: Calculated gain spectra of compressively strained InGaAs quantum wells. In the left figure the carrier density is used as parameter and in the right figure the temperature is varied.

figure an optimum quantum well thickness between 6 nm and 8 nm for highest gain can be deduced for the given structure. As stated above the peak gain does not grow linear with increasing carrier density leading to the logarithmic gain law and the gain saturation in the small signal approximation.
As a conclusion the benefits from strained quantum wells are

- New wavelength regimes are accessible(i.e. 980 nm with InGaAs-GaAs. In this specific case the substrate is transparent enabling new designs for lasers, detectors and modulators. The operation wavelength can be optimized for the desired application i.e. pumping of EDFAs at 980 nm.

- A new parameter in band structure engineering is available that enables the adjustment of hole masses.

- The differential gain is increased giving rise to higher small signal resonance frequencies in lasers.

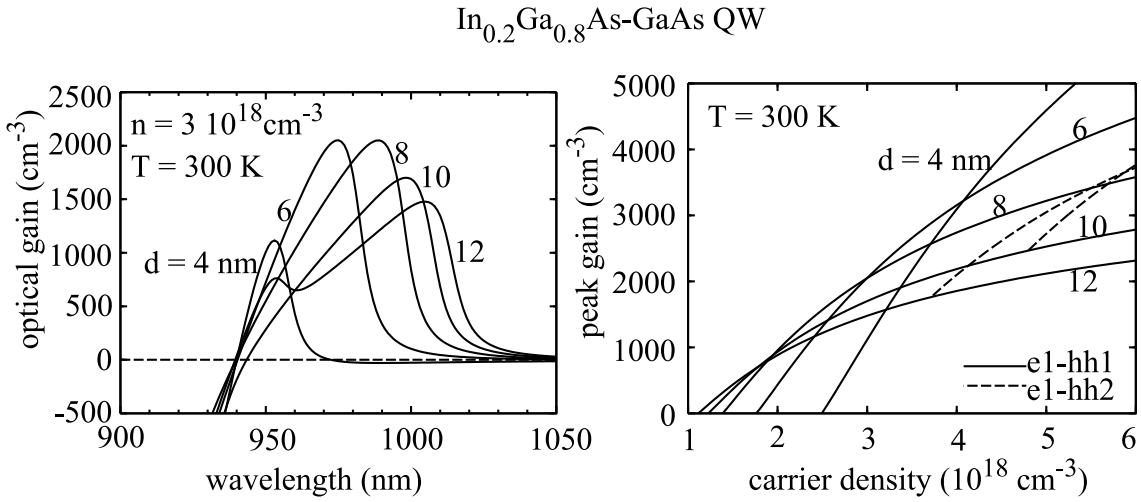

Figure 7.35: Gain spectrum and peak gain as a function of carrier density both with quantum well thickness as parameter.

- The TE or TM polarization is favored enabling single polarization emission.
- With an alternating strained quantum well stack polarization insensitive semiconductor optical amplifiers (SOAs) are achievable (compensation of quantum well and waveguide polarization effects).

7.8.2 Rate Equations

The mathematical description of laser operation makes use of the Einstein equations as deduced in chapter 1. The emission and absorption rates as well as the spectral photon density discussed there are spectral and spatial densities. For calculation of output power and input current we have to go to numbers. As a first step we calculate a (spatial) photon density and the (spatial) absorption and emission rates as an average value over the spectral width $\Delta\nu$ around the emission frequency ν_{m} assuming that there is no big change in the spectral density over $\Delta\nu$. Taking $r_{\mathrm{spon}}, r_{\mathrm{stim}}, r_{\mathrm{abs}}$ and ρ_{phot} from section 1.2.2 we define the absorption and emission rates

$$R_{\mathrm{abs}\,m} = \int_{\hbar\omega_m - h\Delta\nu}^{\hbar\omega_m + h\Delta\nu} r_{\mathrm{abs}} \mathrm{d}(\hbar\omega) \quad , \tag{7.21}$$

$$R_{\mathrm{stim}\,m} = \int_{\hbar\omega_m - h\Delta\nu}^{\hbar\omega_m + h\Delta\nu} r_{\mathrm{stim}} \mathrm{d}(\hbar\omega) \quad , \tag{7.22}$$

$$R_{\mathrm{spon}\,m} = \int_{\hbar\omega_m - h\Delta\nu}^{\hbar\omega_m + h\Delta\nu} r_{\mathrm{spon}} \mathrm{d}(\hbar\omega) \quad , \tag{7.23}$$

and the (spatial) photon density

$$N_{\mathrm{phot}\,m} = \int_{\hbar\omega_m - h\Delta\nu}^{\hbar\omega_m + h\Delta\nu} \rho_{\mathrm{phot}} \mathrm{d}(\hbar\omega) \quad . \tag{7.24}$$

Besides the spontaneous emission there are also non-radiative emission processes like sidewall recombination and Auger recombination that contradict the laser operation and have to be considered. We use R_{non} for non-radiative recombination and define a recombination rate

$$R_{\mathrm{c}} = R_{\mathrm{spon}} + R_{\mathrm{non}} \tag{7.25}$$

that covers the spontaneous and non-radiative recombinations. From section 1.2.2 we already know the spontaneous recombination rate $r_{\text{spon}}\text{d}(\hbar\omega) = A(n+\Delta n)(p+\Delta p)\text{d}(\hbar\omega)$ where Δn and Δp are the amounts of excess carriers that are injected into the active region. Normally $\Delta n = \Delta p$ is assumed and in double heterostructures both Δn and Δp exceed the steady state values n and p in the active region under laser operation giving $r_{\text{spon}}\text{d}(\hbar\omega) = A(\Delta n)^2\text{d}(\hbar\omega)$. It is common to describe the spontaneous emission process by $\frac{\partial}{\partial t}N_{\text{phot}} = R_{\text{spon}} = \Delta n/\tau_{\text{s}}$ which holds only under weak excitation when the steady state carrier densities exceed the excess carrier density. This is clearly not true for laser operation. The non-radiative recombination on sidewalls is directly proportional to the carrier densities $R_{\text{non1}} = A_{\text{non1}}(n+\Delta n+p+\Delta p) \simeq A_{\text{non1}}2\Delta n$ and the Auger recombination originates from photons that transfer their energy to carriers in the form of kinetic energy. Auger recombination is therefore a three carrier process (two for emission and one for kinetic energy transfer) resulting into $R_{\text{non3}} \simeq A_{\text{non3}}\Delta n^3$. As a result the rate equation for the photon density is

$$\frac{\partial}{\partial t}N_{\text{phot}} = R_{\text{spon}} + R_{\text{stim}} - R_{\text{abs}} \tag{7.26}$$

The rate equation for the carrier density under laser operation is the same for electrons and holes $N_{\text{hole}} \simeq N_{\text{electr}} = n + \Delta n \simeq \Delta n = N$ and results into

$$\frac{\partial}{\partial t}N = G_{\text{c}} - R_{\text{c}} - (R_{\text{stim}} - R_{\text{abs}}) \tag{7.27}$$

with carrier generation rate G_{c}.

In a semiconductor laser the light is guided in an optical waveguide as modes and only a fraction of the total spontaneous emission is coupled into mode m. As we discussed in section 7.2 the contribution from spontaneous emission can be calculated by the Langevin force. Here we attribute the fraction that contributes to mode m by the so called spontaneous emission factor $\beta_{\text{spon}m}$. The light inside the resonator is partly lost through the mirrors as output and internal losses. We consider the losses by the photon

lifetime τ. The difference between R_{stim} and R_{abs} is $\sum_m c/\eta_{\mathrm{gr}\,m} g\{\hbar\omega_m, N\} N_{\mathrm{phot}\,m}$ analog to section 1.2.2.

The photon density follows now from (7.26)

$$\frac{\partial}{\partial t} N_{\mathrm{phot}\,m} = \beta_{\mathrm{spon}\,m} R_{\mathrm{spon}}\{N\} + \frac{c}{\eta_{\mathrm{gr}\,m}} g\{\hbar\omega_m, N\} N_{\mathrm{phot}\,m} - \frac{N_{\mathrm{phot}\,m}}{\tau_m} \tag{7.28}$$

and for the electrons we note

$$\frac{\partial}{\partial t} N = G_{\mathrm{c}}\{N\} - R_{\mathrm{c}}\{N\} - \sum_m \frac{c}{\eta_{\mathrm{gr}\,m}} g\{\hbar\omega_m, N\} N_{\mathrm{phot}\,m} \quad . \tag{7.29}$$

The carrier generation follows from current injection over the active region of thickness d with current confinement η_{I} and can be noted as

$$G_{\mathrm{c}} = \eta_{\mathrm{I}} \frac{j}{qd} \tag{7.30}$$

with current density j normal to the active region. The gain in (7.28) is only present in the active region whereas the photons of a waveguide mode extend over the active region and therefore occupy a bigger volume. The optical confinement factor is the ratio of volume where gain and photons are present at the same time to the volume occupied by photons and to the gain volume. The definition of the optical confinement factor makes use of the spatial gain distribution $g_m\{N\} = g\{\hbar\omega_m, N\}$ and the z- component of the time averaged mode Poynting vector $S_{\mathrm{z}\,m}$ and can be written as

$$\Gamma_{\mathrm{m}} = \frac{\frac{1}{A_{\mathrm{active}}} \iint\limits_{A_{\mathrm{active}}} S_{\mathrm{z}\,m} g_m \mathrm{d}^2 r}{\frac{1}{A_{\mathrm{active}}} \iint\limits_{A_{\mathrm{active}}} g_m \mathrm{d}^2 r \, \frac{1}{A_{\mathrm{active}}} \iint\limits_{A_{\mathrm{active}}} S_{\mathrm{z}\,m} \mathrm{d}^2 r} = \frac{\iint\limits_{A_{\mathrm{active}}} S_{\mathrm{z}\,m} g_m \mathrm{d}^2 r}{\frac{1}{A_{\mathrm{active}}} \iint\limits_{A_{\mathrm{active}}} g_m \mathrm{d}^2 r \iint\limits_{A_{\mathrm{active}}} S_{\mathrm{z}\,m} \mathrm{d}^2 r} \tag{7.31}$$

where A_{active} denotes the cross section area (transversal to the waveguide axis) of the active region. This approach makes use of the assumption that the photon and gain distribution does not vary along the laser axis. Often a constant gain over thickness d

is assumed. Moreover the width of the gain region is assumed to extend infinitely. In this case the well known expression

$$\Gamma_{\mathrm{m}} = \frac{\int_{-d/2}^{d/2} |E_m|^2 \mathrm{d}r}{\int_{-\infty}^{\infty} |E_m|^2 \mathrm{d}r} \tag{7.32}$$

results with $S_{\mathrm{z}} \propto |E_m|^2$ for time harmonic fields. The confinement factor differs for different modes m. Photons are only generated in the active region by stimulated and spontaneous emission and therefore (7.28) has to be modified to

$$\frac{\partial}{\partial t} N_{\mathrm{phot}\,m} = \beta_{\mathrm{spon}\,m} \Gamma_m R_{\mathrm{spon}}\{N\} + \frac{c}{\eta_{\mathrm{gr}\,m}} \Gamma_m g_m\{N\} N_{\mathrm{phot}\,m} - \frac{N_{\mathrm{phot}\,m}}{\tau_m} \tag{7.33}$$

for each mode. The generated light in all modes originates from carrier recombination from the same carrier reservoir leading from (7.29) with (7.30) to

$$\frac{\partial}{\partial t} N = \eta_{\mathrm{I}} \frac{j}{qd} - R_{\mathrm{c}}\{N\} - \sum_m \frac{c}{\eta_{\mathrm{gr}\,m}} g_m\{N\} N_{\mathrm{phot}\,m} \quad . \tag{7.34}$$

(7.33) and (7.34) are the so called multimode rate equations for lasers where emission occurs between two levels. If the radiating emission as well takes place between more levels (7.34) is modified because pumping for carrier generation has to be distributed over the lasing levels. The general behaviour remains the same for such many-level lasers and therefore the simpler rate equations are discussed in the following sections. First we try to solve the rate equations for the case that only one mode is present and look at the steady state behaviour followed by the steady state multimode behaviour.

7.8.2.1 Single Mode Rate Equations

For single mode lasers (7.33) and (7.34) are

$$\frac{\partial}{\partial t}N_{\text{phot}} = \beta_{\text{spon}}\Gamma R_{\text{spon}}\{N\} + \frac{c}{\eta_{\text{gr}}}\Gamma g\{N\}N_{\text{phot}} - \frac{N_{\text{phot}}}{\tau} \tag{7.35}$$

and

$$\frac{\partial}{\partial t}N = \eta_{\text{I}}\frac{j}{qd} - R_{\text{c}}\{N\} - \frac{c}{\eta_{\text{gr}}}g\{N\}N_{\text{phot}} \quad . \tag{7.36}$$

Steady state solutions for the photon density follow from $\frac{\partial}{\partial t}N_{\text{phot}} = 0$

$$\left(\frac{c}{\eta_{\text{gr}}}\Gamma g\{N\} - \frac{1}{\tau}\right)N_{\text{phot}} = \beta_{\text{spon}}\Gamma R_{\text{spon}}\{N\} \quad . \tag{7.37}$$

Spontaneous emission into the mode is normally negligible $\beta_{\text{spon}} \leq 10^{-3}$ resulting into

$$\left(\frac{c}{\eta_{\text{gr}}}\Gamma g\{N\} - \frac{1}{\tau}\right)N_{\text{phot}} \approx 0 \quad . \tag{7.38}$$

Two possible solutions arise namely

$$N_{\text{phot}} = 0 \tag{7.39}$$

and

$$\frac{c}{\eta_{\text{gr}}}\Gamma g\{N\} - \frac{1}{\tau} = 0 \quad . \tag{7.40}$$

The first case denotes no laser operation and the second one is found for lasing where arbitrary photon densities are possible when only (7.38) is regarded. From (7.40) follows a constant gain

$$\Gamma g\{N\} = \frac{\eta_{\text{gr}}}{c}\frac{1}{\tau} \quad . \tag{7.41}$$

A comparison with the resonator model in section 2.1 where we found (2.5)

$$\Gamma g = \Gamma g_{\mathrm{th}} = \alpha_{\mathrm{i}} - \frac{1}{2L}\ln\{R_1 R_2\} \tag{7.42}$$

gives

$$\Gamma g_{\mathrm{th}} = \frac{\eta_{\mathrm{gr}}}{c}\frac{1}{\tau} = \Gamma g\{N_{\mathrm{th}}\} \tag{7.43}$$

which in turn leads to a constant carrier density $N = N_{\mathrm{th}}$. This effect is called carrier clamping. The carrier density below threshold ($N_{\mathrm{phot}} = 0$) and the photon density above threshold ($N = N_{\mathrm{th}}$) are resulting from the steady state carrier rate equation

$$\frac{c}{\eta_{\mathrm{gr}}} g\{N\} N_{\mathrm{phot}} = \eta_{\mathrm{I}}\frac{j}{qd} - R_{\mathrm{c}}\{N\} \quad . \tag{7.44}$$

Below threshold

$$j = \frac{qd}{\eta_{\mathrm{I}}} R_{\mathrm{c}}\{N\} \quad . \tag{7.45}$$

is found where $R_{\mathrm{c}}\{N\}$ is a nonlinear function of carrier density as noted above. (7.45) can be solved inversely assuming a carrier density and calculating the resulting current density. Above threshold the carrier density is constant and a constant threshold current density follows from (7.45)

$$j_{\mathrm{th}} = \frac{qd}{\eta_{\mathrm{I}}} R_{\mathrm{c}}\{N_{\mathrm{th}}\} \quad . \tag{7.46}$$

and with (7.44) the photon density

$$N_{\mathrm{phot}} = \frac{\eta_{\mathrm{I}}\frac{j}{qd} - R_{\mathrm{c}}\{N\}}{\frac{c}{\eta_{\mathrm{gr}}} g\{N\}} \quad . \tag{7.47}$$

follows. We can set $N = N_{\rm th}$ and use therefore

$$N_{\rm phot} = \Gamma \frac{\eta_{\rm gr}}{c} \frac{-2L}{\ln\{R_1 R_2\}} \frac{\eta_{\rm I}}{1 - \frac{\alpha_{\rm i} 2L}{\ln\{R_1 R_2\}}} \frac{j - j_{\rm th}}{qd} \quad . \tag{7.48}$$

where we regarded (7.46) and (7.41). The dashed lines in figure 7.36 shows the resulting carrier and photon densities below and above threshold as a function of current density.

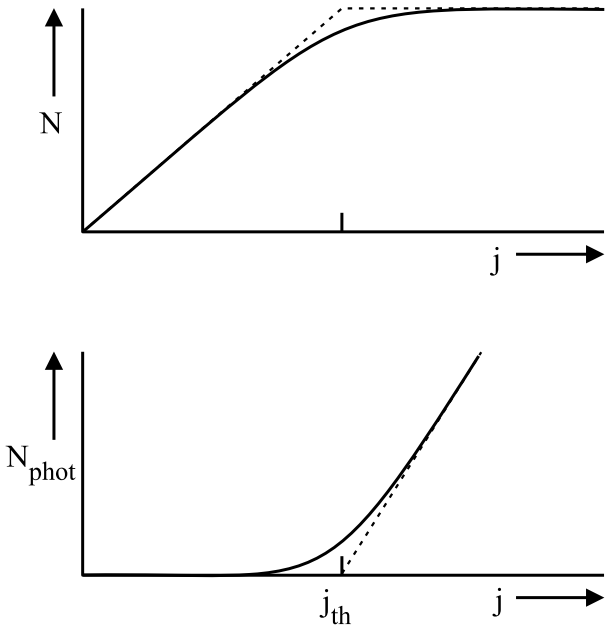

Figure 7.36: Carrier and photon density in a single mode laser as function of the driving current density. Dashed lines denote the case when spontaneous emission is negligible and the solid lines are obtained with spontaneous emission.

When spontaneous emission is no longer negligible (7.37) must be used in the form

$$N_{\rm phot} = \frac{\beta_{\rm spon} \Gamma R_{\rm spon}\{N\}}{\frac{c}{\eta_{\rm gr}} \left(\frac{\eta_{\rm gr}}{c\tau} - \Gamma g\{N\} \right)} \quad . \tag{7.49}$$

Far above threshold the photon density is that high that the spontaneous emission can be neglected and (7.42) must be valid even when spontaneous emission is regarded and therefore (7.42) can be taken for any photon density leading to

$$N_{\text{phot}} = \frac{\beta_{\text{spon}} R_{\text{spon}}\{N\}}{\frac{c}{\eta_{\text{gr}}}\left(g_{\text{th}} - g\{N\}\right)} \quad . \tag{7.50}$$

The steady state current density is

$$\eta_{\text{I}} \frac{j}{qd} = R_{\text{c}}\{N\} + \frac{c}{\eta_{\text{gr}}} g\{N\} N_{\text{phot}} \quad . \tag{7.51}$$

which can be transformed with (7.50) and (7.46) to

$$j = j_{\text{th}} \frac{R_{\text{c}}\{N\}}{R_{\text{c}}\{N_{\text{th}}\}} + \frac{qd}{\eta_{\text{I}}} \frac{g\{N\}}{g_{\text{th}} - g} \beta_{\text{spon}} R_{\text{spon}}\{N\} \quad . \tag{7.52}$$

Both equations (7.50) and (7.52) give values for N_{phot} and j with N as parameter and must be solved numerically.

It must be stated here that the gain g does not reach the threshold gain g_{th} and therefore the carrier density N stays below the threshold density N_{th}. At a first glance this is somewhat surprising and seems to contradict the amplitude condition from the resonator model where we have to satisfy $g = g_{\text{th}}$. The explanation is found in the contribution from spontaneous emission that acts like a small additional gain $g_{\text{spon}}\{N, N_{\text{phot}}\} = \frac{\eta_{\text{gr}}}{c} \beta_{\text{spon}} R_{\text{spon}}\{N\}/N_{\text{phot}}$ such that (7.41) can be rewritten in the form $g\{N\} + g_{\text{spon}}\{N, N_{\text{phot}}\} = g_{\text{th}}$. The gain attributed to spontaneous emission is inversely proportional to the photon density apparent. With increasing photon density g_{spon} decreases and the amplitude condition requires a compensation by higher gain g leading to a higher carrier density.

Far below threshold g is negligible in (7.50) and the photon density increases directly proportional with the spontaneous recombination rate

$$N_{\text{phot}} = \frac{\eta_{\text{gr}} \beta_{\text{spon}}}{c g_{\text{th}}} R_{\text{spon}}\{N\} \quad . \tag{7.53}$$

As we know for weak excitation the spontaneous emission rate is directly proportional to the excess carrier density and can be set $R_{\text{spon}} = N/\tau_{\text{s}}$ resulting into

$$N_{\text{phot}} = \frac{\eta_{\text{gr}}\beta_{\text{spon}}}{c\tau_{\text{s}}g_{\text{th}}}N \quad . \tag{7.54}$$

For weak excitation Auger recombination is negligible and the excess carrier density therefore depends linearly on the current density $N_{\text{phot}} \propto j$. When the carrier density approaches N_{th} the denominator in (7.50) approaches zero and the photon density therefore increases dramatically. Rearranging (7.51)in the form

$$N_{\text{phot}} = \frac{\eta_{\text{gr}}}{cg\{N\}}\frac{\eta_{\text{I}}}{qd}\left(j - \frac{qd}{\eta_{\text{I}}}R_{\text{c}}\{N\}\right) = \frac{\eta_{\text{gr}}}{cg\{N\}}\frac{\eta_{\text{I}}}{qd}\left(j - j_{\text{th}}\frac{R_{\text{c}}\{N\}}{R_{\text{c}}\{N_{\text{th}}\}}\right) \tag{7.55}$$

yields far above threshold where $g \simeq g_{\text{th}}$ and therefore $N \simeq N_{\text{th}}$

$$N_{\text{phot}} \simeq \frac{\eta_{\text{gr}}}{cg_{\text{th}}}\frac{\eta_{\text{I}}}{qd}(j - j_{\text{th}}) = \Gamma\frac{\eta_{\text{gr}}}{c}\frac{-2L}{\ln\{R_1R_2\}}\frac{\eta_{\text{I}}}{1 - \frac{\alpha_{\text{i}}2L}{\ln\{R_1R_2\}}}\frac{j - j_{\text{th}}}{qd} \tag{7.56}$$

which is the same as (7.48).

The optical output power of the laser diode is proportional to the photon density and the energy carried by each photon. The photons travel with group velocity and the volume occupied by the photons can be calculated with the optical confinement factor from $\Gamma V_{\text{phot}} = V_{\text{active}}$ which leads to the total carried energy inside the resonator

$$W_{\text{resonator}} = \hbar\omega N_{\text{phot}}V_{\text{phot}} = \frac{\eta_{\text{gr}}}{c}\frac{-2L}{\ln\{R_1R_2\}}\hbar\omega\frac{\eta_{\text{I}}}{1 - \frac{\alpha_{\text{i}}2L}{\ln\{R_1R_2\}}}\frac{j - j_{\text{th}}}{qd}V_{\text{active}} \quad . \tag{7.57}$$

The photon lifetime is the inverse sum of lifetime τ_{i} due to internal losses α_{i} and lifetime τ_{R} due to losses through the mirrors

$$\frac{1}{\tau} = \frac{1}{\tau_{\text{i}}} + \frac{1}{\tau_{\text{R}}} = \frac{c}{\eta_{\text{gr}}}\left(\alpha_{\text{i}} - \frac{1}{2L}\ln\{R_1R_2\}\right) = \frac{c}{\eta_{\text{gr}}}\alpha_{\text{i}} + \frac{c}{\eta_{\text{gr}}}\frac{-\ln\{R_1R_2\}}{2L} \tag{7.58}$$

The losses through the mirrors are output and with the so called resonator lifetime

$$\tau_{\mathrm{R}} = \frac{\eta_{\mathrm{gr}}}{c} \frac{-2L}{\ln\{R_1 R_2\}} \tag{7.59}$$

the optical output power well above threshold is

$$P_{\text{phot out}} = \frac{W_{\text{resonator}}}{\tau_{\mathrm{R}}} = \hbar\omega \, \frac{\eta_{\mathrm{I}}}{1 - \frac{\alpha_{\mathrm{i}} 2L}{\ln\{R_1 R_2\}}} \, \frac{j - j_{\mathrm{th}}}{qd} \, V_{\text{active}} \quad . \tag{7.60}$$

With thickness d of the active volume $V_{\text{active}} = A_{\text{cur}} d$ the area A_{cur} remains and together with j and j_{th} the currents $I = jA_{\text{cur}}$ and $I_{\mathrm{th}} = j_{\mathrm{th}} A_{\text{cur}}$ are defined resulting into output power

$$P_{\text{phot out}} = \frac{\hbar\omega}{q} \, \frac{\eta_{\mathrm{I}}}{1 - \frac{\alpha_{\mathrm{i}} 2L}{\ln\{R_1 R_2\}}} \, (I - I_{\mathrm{th}}) \quad . \tag{7.61}$$

well above threshold. The differential slope efficiency

$$\eta_{\mathrm{d}}\{I\} = \frac{q}{\hbar\omega} \frac{\partial}{\partial I} P_{\text{phot out}} \bigg|_{I > I_{\mathrm{th}}} \tag{7.62}$$

gives the slope of the output power at current I above threshold. The differential operator can be written as

$$\eta_{\mathrm{d}}\{I\} = \frac{\partial}{\partial(\frac{I}{q})} \frac{P_{\text{phot out}}}{\hbar\omega} = \frac{\partial}{\partial N} N_{\text{phot out}} \simeq \frac{\Delta N_{\text{phot out}}}{\Delta N} \tag{7.63}$$

and can be understood as ratio between additional carriers recombining in the laser and additional photons leaving the resonator. All additional carriers in the active zone are converted into photons and the difference between generated photons and leaving photons is lost inside the resonator (described by α_{i}). When we apply (7.62) to (7.61) we find

$$\eta_{\mathrm{d}}\{I\} = \frac{\eta_{\mathrm{I}}}{1 - \frac{\alpha_{\mathrm{i}} 2L}{\ln\{R_1 R_2\}}} \tag{7.64}$$

where η_I can be understood with the discussion above as ratio between the number of additional carriers recombining inside the active region to the total number of additional carriers in the laser. Non-radiative and most of the spontaneous recombination are losses for laser operation. Well above threshold the carrier density remains close to $N \simeq N_\mathrm{th}$ and therefore the losses are constant. They are included in the threshold current density $j_\mathrm{th} = \frac{q}{d\eta_\mathrm{i}} R_\mathrm{c}\{N_\mathrm{th}\}$.

All equations above make no notice of the temperature which was assumed to be constant. When the temperature rises inside the laser due to absorption and non-radiative recombination processes the recombination rate $R_\mathrm{c}\{N\}$ increases giving rise to increased threshold current density. With increased temperature, leakage currents can increase due to better conductivity and as well more carriers can escape the active region due to their higher thermal energy which makes it easier to jump over the barriers built up by the bandgap discontinuities. Both effects lead to decreasing current efficiency. When leakage currents are dominated by carrier escape it is reasonable to describe the temperature dependence of η_I by

$$\eta_\mathrm{I}\{T\} = \eta_\mathrm{I}\{T_1\} \exp\{-\frac{T}{T_1}\} \quad . \tag{7.65}$$

All recombination rates are strongly dependent on the temperature due to occupation probability of the density of states which is described by a Fermi distribution. This leads to the empirical description

$$j_\mathrm{th}\{T\} = j_\mathrm{th}\{T_0\} \exp\left\{\frac{T}{T_0}\right\} \quad . \tag{7.66}$$

The values of T_0 and T_1 are in today's III-V semiconductor lasers around 100 K and 300 to 400 K, respectively. An increase of the characteristic temperatures improves the high-temperature operation behaviour of the lasers. T_1 can be increased by lowering of carrier escape which can be maintained by higher energy barriers. The value of T_0 strongly depends on the materials used and on the material quality. With bad

material the recombination rate is high and a strong temperature dependence has to be expected.

With changing temperatures the resonator properties are changed as well due temperature dependent refractive indices which leads to other resonance conditions thereby shifting the wavelength and the optical confinement factor.

Also the current distribution can change thereby changing the gain distribution and again the optical confinement. Therefore especially for the measurement of η_{I} care must be taken because of possibly changing geometric length and reflectivities at the output mirrors due to wavelength shift and refractive index change.

7.8.2.2 Multi Mode Rate Equations

In the preceding subsection we concentrated the discussion on single mode laser operation. Here we try to find out how the total power is distributed among the possible modes of a laser. We start from the multi mode rate equations (7.33) and (7.34). For convenience we label the modes with number m centered around the mode which starts to lase with the lowest threshold carrier density and is therefore the dominant mode as we will find later.

When we try to find solutions of the multi mode rate equations without regarding spontaneous emission ($\beta_{\mathrm{spon}} = 0$) the carrier density is clamped to $N = N_{\mathrm{th0}}$ and only the central mode is lasing (all other modes are below threshold). In other words, with negligible spontaneous emission, single mode operation follows.

The solutions of the rate equations that consider spontaneous emission are similar to (7.50). The photon density follows from

$$N_{\mathrm{phot}m} = \frac{\beta_{\mathrm{spon}m} R_{\mathrm{spon}}\{N\}}{\frac{c}{\eta_{\mathrm{gr}m}} \left(g_{\mathrm{th}m} - g_m\{N\}\right)} \tag{7.67}$$

and the current density is calculated as

$$j = j_{\mathrm{th}0}\frac{R_{\mathrm{c}}\{N\}}{R_{\mathrm{c}}\{N_{\mathrm{th}}\}} + \frac{qd}{\eta_{\mathrm{I}}}\sum_{m}\frac{g\{N\}\beta_{\mathrm{spon}m}R_{\mathrm{spon}}}{g_{\mathrm{th}m} - g_m\{N\}} \tag{7.68}$$

where we use the gain and threshold gain of mode m $g_m\{N\} = g\{\hbar\omega_m, N\}$ and $g_{\mathrm{th}m}\{N\} = g_m\{N_{\mathrm{th}m}\}$ and abbreviate the threshold gain of the central mode $m = 0$ with $N_{\mathrm{th}0} = N_{\mathrm{th}}$. (7.67) and (7.68) are again parametric representations of photon and current density with carrier density N as parameter.

As we know from the preceding subsection even with spontaneous emission considered the carrier density asymptotically approaches N_{th}. The maximum photon density in mode m is limited to the case $N = N_{\mathrm{th}}$ and results to

$$N_{\mathrm{phot,sat}m} = \frac{\beta_{\mathrm{spon}m}R_{\mathrm{spon}}\{N_{\mathrm{th}0}\}}{\frac{c}{\eta_{\mathrm{gr}m}}\left(g_m\{N_{\mathrm{th}m}\} - g_m\{N_{\mathrm{th}0}\}\right)} \quad . \tag{7.69}$$

and is called saturation photon density. The photon density as function of carrier density is depicted in figure 7.37.

The saturation photon density is infinite for the central mode. For the sidemodes $|m| > 0$ the maximum power is proportional to the spontaneous emission factor. With decreasing spontaneous emission factor the sidemode intensity decreases and the so called sidemode suppression ratio $10\mathrm{dB}\,\lg\{P_0/P_m\}$ increases. The output spectrum of a multi mode laser diode with different spontaneous emission factors is sketched in figure 7.38.

As can be seen from figure 7.38 for diminishing spontaneous emission single mode operation results. The photon density in this case can not be calculated from (7.67) because numerator and denominator are zero due to $N = N_{\mathrm{th}}$. Instead of (7.67) for diminishing spontaneous emission (7.48) has to be taken to calculate the photon density.

With (7.67) laser operation can be understood as resonantly amplified spontaneous emission. A decreasing spontaneous emission factor is easily compensated by slightly higher carrier density thus lowering the difference between $g_{\mathrm{th}0}$ and g_0 such that the

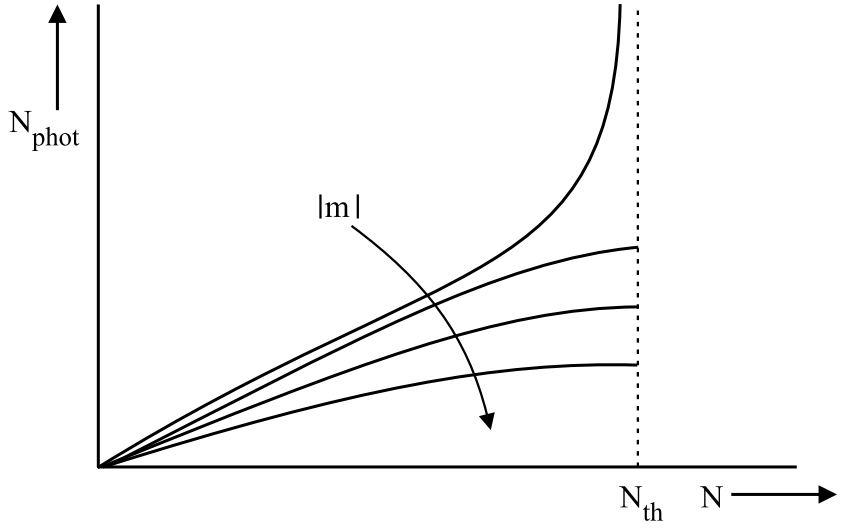

Figure 7.37: Photon density as function of carrier density in a multi mode laser. The photon density saturates for the sidemodes at a certain level that is proportional to the spontaneous emission factor β_{spon}.

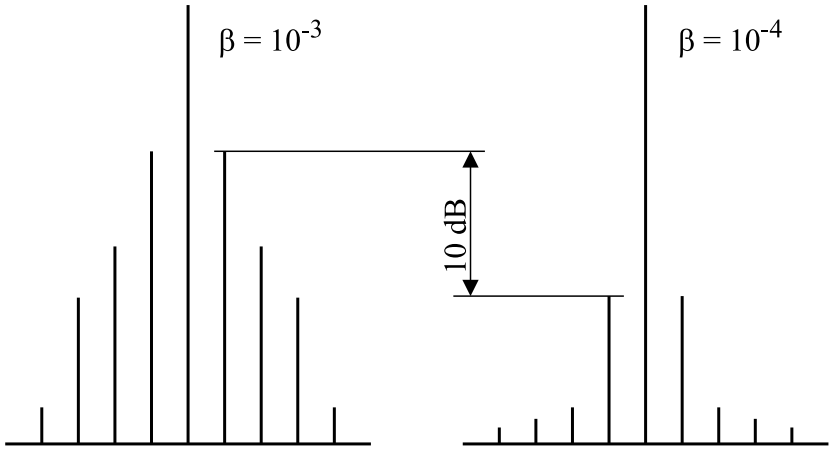

Figure 7.38: Optical output power spectrum of multi mode lasers with different spontaneous emission factors β_{spon}.

same output power is obtained. On the other hand the sidemode suppression ratio grows with decreasing spontaneous emission factor. The question arises if it is possible to influence the spontaneous emission factor thereby obtaining (nearly) single mode operation.

7.8.2.3 Spontaneous Emission Factor

The spontaneous emission factor β_{spon} gives the fraction of spontaneous emission that is coupled into a resonator mode. Spontaneous emission occurs in a spectrum $\Delta\lambda_{\text{spon}}$ around the center wavelength λ_{p}. The spectral width is given by the energetic carrier density distribution and can be assumed to follow from a Fermi distribution in semiconductors. The Fermi distribution decreases to minimal values within an energy band of $\hbar\Delta\omega_{\text{spon}} \simeq 2k_{\text{B}}T$ giving rise to the spectral width of

$$\Delta\lambda_{\text{spon}} \simeq \Delta\omega_{\text{spon}}\lambda_{\text{p}}/\omega_{\text{p}} = \frac{2k_{\text{B}}T\lambda_{\text{p}}^2}{hc} \quad .$$

At room temperature the spectral width is $\Delta\lambda_{\text{spon}} \simeq 40\,\text{nm}$ for a peak wavelength of $\lambda_{\text{p}} = 1\,\mu\text{m}$ and therefore

$$\Delta\lambda_{\text{spon}} \simeq \left(\frac{T}{300\,\text{K}}\right)\left(\frac{\lambda_{\text{p}}}{1\,\mu\text{m}}\right)^2 40\,\text{nm}$$

can be noted. As we know for coherent light only discrete values of the propagation vector $\vec{k}$ are allowed for a laser mode at a given wavelength. Spontaneous emission is arbitrarily directed and the fraction that is coupled into a mode that only accepts discrete directions is very low. The strict law of discrete propagation directions for coherent light follows from boundary conditions that act mathematically like a resonance condition where constructive interference occurs. As we already discussed in section 7.3 for additional incoherent emission processes there is a certain probability that resonance conditions can be satisfied also in a spectral range around the solution for coherent waves. This holds as well for the propagation directions inside a resonator

where not only discrete directions but a spectrum of directions are allowed. The directions are directly related to the propagation constant and are sketched in figure 7.39 first as discrete solutions of the resonance conditions and second as smeared out values due to additional incoherent emission processes.

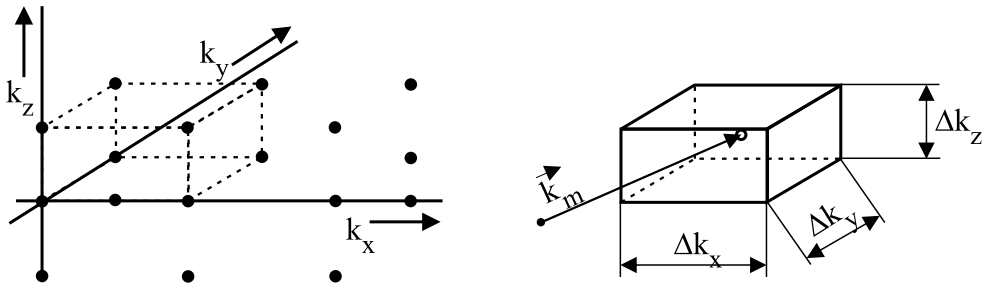

Figure 7.39: Discrete values of the propagation constant for modes of a rectangular resonator. The whole k-space is filled. With incoherent emission processes the discrete values can be smeared out and occupy a volume with dimensions $\Delta k_\mathrm{x}, \Delta k_\mathrm{y}, \Delta k_\mathrm{z}$ centered around the discrete values as sketched on the right side.

The ratio between the volume occupied in k-space by spontaneous emission and by a mode gives an estimation of the spontaneous emission factor for the mode. The spontaneous emission $\vec{k}$-values with peak wavelength λ_p are on the surface of a sphere with radius $k_\mathrm{p} = 2\pi\eta/\lambda_\mathrm{p}$. With spectral width $\Delta\lambda$ a sphere of volume

$$V_\mathrm{k,spon} = 4\pi k_\mathrm{p}^2 \Delta k_\mathrm{spon} = \frac{16\pi^3\eta^2\Delta k_\mathrm{spon}}{\lambda_\mathrm{p}^2} \tag{7.70}$$

is occupied by spontaneous emission in the k-space. Figure 7.40 illustrates the k-space occupied by spontaneous emission. The sphere shell thickness Δk_spon follows directly from the spectral linewidth

$$\Delta k_\mathrm{spon} = \frac{\eta_\mathrm{gr}}{c}\Delta\omega = \eta_\mathrm{gr}\frac{\omega_\mathrm{p}}{c}\frac{\Delta\lambda_\mathrm{spon}}{\lambda_\mathrm{p}} = 2\pi\eta_\mathrm{gr}\frac{\Delta\lambda_\mathrm{spon}}{\lambda_\mathrm{p}^4} \quad .$$

As we already derived in section 1.2.1 the spectral photon density is (1.7)

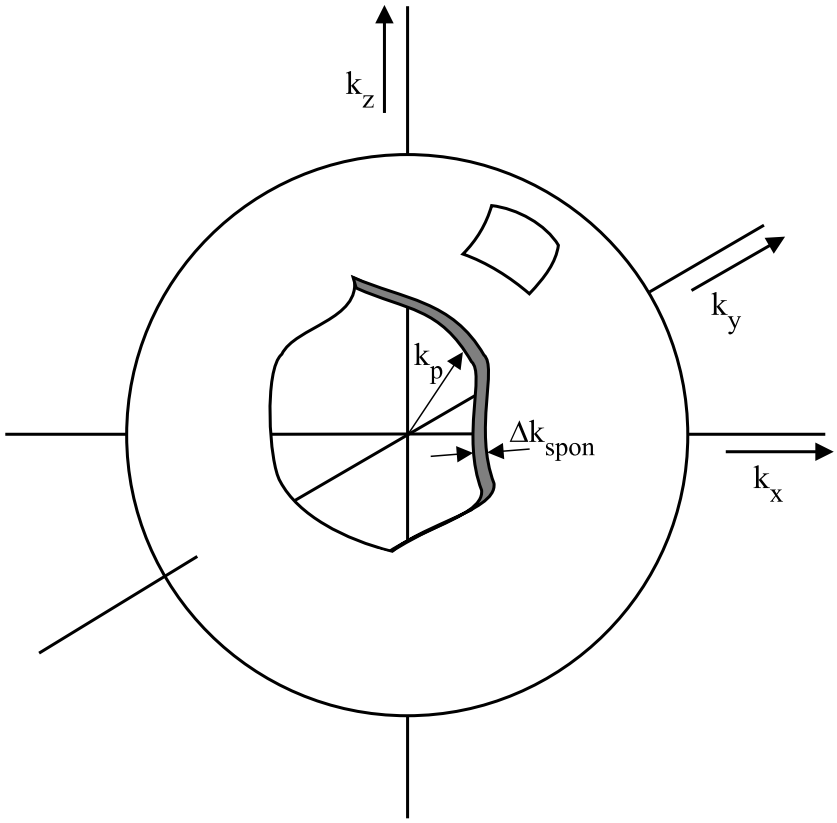

Figure 7.40: The volume occupied by spontaneous emission in k-space is a sphere shell of radius k_{p} and thickness Δk_{spon}.

$$D_{\mathrm{k}}\mathrm{d}^3k = \frac{2}{(2\pi)^3}\mathrm{d}^3k \quad .$$

With the volume covered by spontaneous emission in k-space the mode density is easily derived by integration over the sphere shell (take d^3k in spherical coordinates!). After multiplication with the geometric resonator volume the number of modes excited by spontaneous emission results. The spontaneous emission factor is the fraction of spontaneous emission into just one mode and therefore the inverse of the total number of excited modes.

The minimal extension of the modes of today's lasers exceed 100 nm with the smallest dimension in edge emitting lasers. The biggest extension in k-space is therefore well below Δk_{spon} and the volume occupied by a mode is completely included in the shell of the spontaneous emission volume as depicted in figure 7.41.

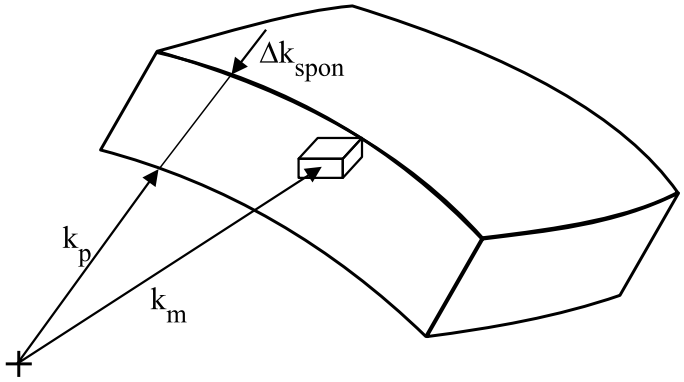

Figure 7.41: Commonly occupied volume of a mode and spontaneous emission in k-space.

The spontaneous emission factor therefore can directly be calculated from the ratio between (1.5)

$$V_{\mathrm{k\ Reson}} = \frac{(2\pi)^3}{2}\frac{1}{V_{\mathrm{Res}}} \quad .$$

and $V_{\text{k spon}}$ (7.70) with $V_{\text{Res}} = V_{\text{active}}/\Gamma$. The spontaneous emission into mode m results to

$$\beta_{\text{spon}m} = \frac{V_{\text{k Reson}}}{V_{\text{k spon}}} \simeq \frac{\lambda_{\text{p}}^4}{8\pi\eta^2\eta_{\text{gr}}\Delta\lambda_{\text{spon}}} \frac{\Gamma_m}{V_{\text{active}}} \quad .$$

The spontaneous emission factor is directly proportional to the fourth power of the spontaneous emission peak wavelength λ_{p} and inversely proportional to the spontaneous emission spectral width $\Delta\lambda_{\text{spon}}$, the active resonator volume V_{active}, the quadratic refractive index η and the group index η_{gr}. For a given operation wavelength the spectral width, the peak wavelength of spontaneous emission and the refractive index and group index are determined by the active material used and can be taken as given values that can not be influenced. The resonator volume can be adjusted and must be made larger to achieve high sidemode suppression ratios. The most important adjustable value is the confinement factor. The active region can be placed that way that it is in the antinode of the main mode and in the node of a sidemode. In that case the confinement factor of the sidemode is smaller than that of the main mode leading to an increase of the sidemode suppression ratio.

When the modes of a resonator are no longer orthogonal as is found in gain guided lasers they occupy a larger volume in k-space and the spontaneous emission factor is increased to

$$\beta_{\text{spon}m} \simeq K_{\text{a}m} \frac{\lambda_{\text{p}}^4}{8\pi\eta^2\eta_{\text{gr}}\Delta\lambda_{\text{spon}}} \frac{\Gamma_m}{V_{\text{active}}} \quad .$$

with the astigmatism factor $K_{\text{a}m}$, which can be estimated from the mode properties discussed in section 3.3 as

$$K_{\text{a}m} \simeq \frac{\left| \iint \left| \vec{E}_m \right|^2 \mathrm{d}^2 r \right|^2}{\left| \iint \left\| \vec{E}_m \right\|^2 \mathrm{d}^2 r \right|^2} \quad .$$

The astigmatism factor is unity for index guided lasers supporting modes with plane phase fronts and can reach $K_{am} \simeq 10$ for gain guided lasers. In typical edge emitting laser diodes the spontaneous emission factor is between 10^{-5} (index guided) and 10^{-4} (gain guided).

7.8.3 Current Modulation

Semiconductor lasers are easily modulated electronically by changing the driving current. The time dependent output follows from the rate equations. Due to the nonlinearity of the rate equations analytical expressions for the modulation behaviour are not easy found. An approximative solution follows from small signal current modulation where the rate equations can be linearized leading to the analytical small signal modulation description. For large signal modulation only numerical descriptions are available with analytical description of some details.

In the rate equations of section 7.8.2 no noise fluctuations are considered. They are represented by Langevin forces and have to be added as a source in both equations. Additionally the gain is assumed to be photon density independent. Measurements show that especially in quantum structures the gain decreases with high photon densities. This effect is called gain compression and is similar to the absorption saturation described in section 5.2. The gain compression can be described by a factor $1/(1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}})$ with gain compression factor

$$\epsilon_{\mathrm{g}} \simeq \frac{\lambda^2 c \tau_{\mathrm{intra}}}{2\pi\eta^3} \frac{\tau_{\mathrm{intra}}}{\tau_{\mathrm{spon}}} \tag{7.71}$$

with intraband relaxation time $\tau_{\mathrm{intra}} \simeq 100\,\mathrm{ps}$ and spontaneous emission lifetime τ_{spon}. The rate equations (7.33) and (7.34) therefore have to be rewritten to

$$\frac{\partial}{\partial t} N_{\mathrm{phot}\,m} = F_{\mathrm{phot}} + \beta_{\mathrm{spon}\,m} \Gamma_m R_{\mathrm{spon}}\{N\} + c_{\mathrm{gr}\,m} \Gamma_m \frac{g_m\{N\}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}\,m}} N_{\mathrm{phot}\,m} - \frac{N_{\mathrm{phot}\,m}}{\tau_m} \tag{7.72}$$

and

$$\frac{\partial}{\partial t} N = F_{\mathrm{N}} + \eta_{\mathrm{I}} \frac{j}{qd} - R_{\mathrm{c}}\{N\} - \sum_m c_{\mathrm{gr}\,m} \frac{g_m\{N\}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}\,m}} N_{\mathrm{phot}\,m} \quad . \tag{7.73}$$

Commonly the Langevin forces for carriers and photons are set $F_{\mathrm{N}} \simeq -F_{\mathrm{phot}}$ because each photon that is generated is caused at least by one recombination. Non radiating recombinations are neglected with that setting.

7.8.3.1 Small Signal Modulation

Small signal current modulation can be described as a small change Δj in driving current compared to the stationary value j_0. This causes a change ΔN in the carrier density which again can be assumed to be small compared to the stationary value N_0. With the carrier modulation also the photon density is modulated around its stationary value $N_{\mathrm{phot}m0}$. The modulation $\Delta N_{\mathrm{phot}m}$ as well is assumed to be small. The stationary values satisfy $\frac{\partial}{\partial t} N_0 = 0$ and $\frac{\partial}{\partial t} N_{\mathrm{phot}m0} = 0$ when the Langevin forces in (7.72) and (7.73) are neglected. An analytic solution for small signal current amplitude modulation follows from a Taylor expansion around the stationary values. We stop the expansion with first order changes. Due to the fact that the refractive indices are carrier density dependent via the free carrier plasma effect (see section 5.7) the group velocity as well is carrier density dependent. Expansion of (7.72) and (7.73) and dropping the stationary values results into

$$\begin{aligned}\frac{\partial}{\partial t}\Delta N_{\mathrm{phot}m} = F_{\mathrm{phot}m} + \Bigg[& \beta_m \Gamma_m \left.\frac{\partial}{\partial N} R_{\mathrm{spon}}\right|_{N_0} \\ & + c_{\mathrm{gr}0}\Gamma_m \frac{g_{m0} N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}} \left(\frac{1}{g_{m0}} \left.\frac{\partial}{\partial N} g_m\right|_{N_0} - \frac{1}{c_{\mathrm{gr}m0}} \left.\frac{\partial}{\partial N} c_{\mathrm{gr}m}\right|_{N_0} \right) \Bigg] \Delta N \\ & + \Bigg[c_{\mathrm{gr}m0}\Gamma_{\mathrm{m}} \frac{g_{m0} N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}} \left(\frac{1}{N_{\mathrm{phot}m0}} - \frac{\epsilon_{\mathrm{g}}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}} \right) \\ & - \frac{1}{\tau_m} \Bigg] \Delta N_{\mathrm{phot}m}\end{aligned} \tag{7.74}$$

$$\frac{\partial}{\partial t}\Delta N = F_{\mathrm{N}} + \frac{\eta_{\mathrm{I}}}{qd}\Delta j - \left[\left.\frac{\partial}{\partial N} R_{\mathrm{c}}\right|_{N_0}\right.$$

$$+\sum_m c_{\mathrm{grm}0}\frac{g_{m0}N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}\left(\frac{1}{g_{m0}}\left.\frac{\partial}{\partial N}g_m\right|_{N_0}-\frac{1}{c_{\mathrm{grm}0}}\left.\frac{\partial}{\partial N}c_{\mathrm{grm}}\right|_{N_0}\right)\Bigg]\Delta N$$
$$-\sum_m c_{\mathrm{grm}0}\frac{g_{m0}N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}$$
$$\left(\frac{1}{N_{\mathrm{phot}m0}}-\frac{\epsilon_{\mathrm{g}}}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}\right)\Delta N_{\mathrm{phot}m} \quad . \tag{7.75}$$

For convenience we assume that the carrier density is that high that the sidemodes are saturated and therefore the sidemode modulation is negligible. Secondly in this case the recombination terms can be assumed to be stable resulting into $\frac{\partial}{\partial N}R_{\mathrm{c}}\big|_{N_0}=0$. The carrier density dependence of the gain is attributed to the differential gain coefficient

$$a_m=\left.\frac{\partial}{\partial N}g_m\right|_{N_0} \tag{7.76}$$

and the plasma effect compared to the differential gain coefficient

$$\frac{g_{m0}}{c_{\mathrm{grm}0}}\frac{\frac{\partial}{\partial N}c_{\mathrm{grm}}\big|_{N_0}}{\frac{\partial}{\partial N}g_m\big|_{N_0}}=\frac{g_{m0}}{a_m c_{\mathrm{grm}0}}\left.\frac{\partial}{\partial N}c_{\mathrm{grm}}\right|_{N_0}\simeq\frac{-3g_{m0}}{a_m}\left.\frac{\partial}{\partial N}\eta\right|_{N_0}\simeq 10^{-3} \tag{7.77}$$

is negligible in the photon and carrier rate equations.

The refractive index change due to the plasma effect not only changes the amplitude resonance conditions as is represented by the photon rate equation but also the phase condition. In contrary to the amplitude condition the contribution to the phase condition is not negligible. This leads to a slight shift $\Delta\nu$ in the resonance frequency. The propagation in z- direction is described by $\exp\{i(k_{\mathrm{z}}z+\arg\{r_1r_2\}-\omega t)+1/2(g-\alpha_{\mathrm{i}})z\}$. The argument of the exp- function is attributed to a (complex) phase

$$\begin{aligned}\phi_{\mathrm{c}m}&=k_{\mathrm{z}m}z-\omega_m t+\arg\{r_1r_2\}-i\frac{1}{2}(g_m-\alpha_{\mathrm{i}})z\\&=\frac{\omega}{c_{\mathrm{phase}m}}z-\omega_m t+\arg\{r_1r_2\}-i\frac{1}{2}(g_m-\alpha_{\mathrm{i}})z\\&=\frac{\omega\eta_{\mathrm{eff}m}}{c_0}z-\omega_m t+\arg\{r_1r_2\}-i\frac{1}{2}(g_m-\alpha_{\mathrm{i}})z \quad .\end{aligned} \tag{7.78}$$

With the plasma effect η_{eff} is time dependent under carrier density modulation and the phase change $\Delta\phi$ can be estimated from a first order Taylor expansion

$$\begin{aligned}\Delta\phi_{cm} &\simeq \left(\frac{\omega_{m0}}{c_0} z \left.\frac{\partial}{\partial N_{\text{total}}}\eta_{\text{eff}m}\right|_{N_0} - i\frac{z}{2}\left.\frac{\partial}{\partial N_{\text{total}}} g_m\right|_{N_0}\right)\Delta N \\ &= \left(\frac{\omega_{m0}}{c_0}\Gamma_m \left.\frac{\partial}{\partial N}\eta_{\text{eff}m}\right|_{N_0} - i\frac{a_m}{2}\right) z\Delta N \\ &= -\left(\frac{\omega_{m0}}{c_0}\Gamma_m \frac{\eta_{\text{eff}m}}{\eta}\left.\frac{\partial}{\partial N}\eta\right|_{N_0} + i\frac{a_m}{2}\right) z\Delta N \end{aligned} \tag{7.79}$$

where we assumed that the mirror reflectivities are not affected by carrier density and resonance frequency change. The time dependent phase change can be transformed into a phase rate equation where we have to take into account the phase Langevin force from spontaneous emission and find

$$\begin{aligned}\frac{\partial}{\partial t}\Delta\phi_{cm} &\simeq F_{\Phi m} - \left(\frac{\omega_{m0}}{c_0}\Gamma_m \frac{\eta_{\text{eff}m0}}{\eta_0}\left.\frac{\partial}{\partial N}\eta\right|_{N_0} + i\frac{a_m}{2}\right)\left(\frac{c_0}{\eta_{\text{eff}m0}}\Delta N + z\frac{\partial}{\partial t}\Delta N\right) \\ &= F_{\Phi m} + \frac{c_0 a_m}{2\eta_{\text{eff}m0}}(\alpha_{\text{H}m} - i)\left(\Delta N + \frac{z\eta_{\text{eff}m0}}{c_0}\frac{\partial}{\partial t}\Delta N\right)\end{aligned} \tag{7.80}$$

with the so called Henry factor

$$\alpha_{\text{H}m} = -\frac{\Gamma_m\omega_{m0}\eta_{\text{eff}m0}}{a_m c_0}\frac{1}{\eta_0}\left.\frac{\partial}{\partial N}\eta\right|_{N_0} \quad . \tag{7.81}$$

The Henry factor is in typical semiconductor lasers around -5. The real part of the phase change $\text{Re}\{\Delta\phi_c\}$ can be directly attributed to a frequency change $\Delta\omega = \frac{\partial}{\partial t}\text{Re}\{\Delta\phi_c\}$. This means that a small carrier density modulation ΔN not only leads to photon density modulation which is an amplitude modulation of the light output. An additional frequency modulation of the light output is also generated and directly proportional to the Henry factor. With this knowledge the Henry factor can directly be measured from a modulation experiment where the amounts of amplitude and frequency modulation are compared.

With the abbreviations and assumptions above the single mode rate equations are

$$\frac{\partial}{\partial t}\Delta N_{\mathrm{phot}m} = F_{\mathrm{phot}m} + c_{\mathrm{gr}m0}\Gamma_m \frac{a_m N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}}\Delta N + \left(c_{\mathrm{gr}m0}\Gamma_m \frac{g_{m0} N_{\mathrm{phot}m0}}{(1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0})^2} - \frac{1}{\tau_m} \right) \Delta N_{\mathrm{phot}m} \tag{7.82}$$

and

$$\frac{\partial}{\partial t}\Delta N = F_{\mathrm{N}} + \frac{\eta_{\mathrm{I}}}{qd}\Delta j - c_{\mathrm{gr}m0}\frac{a_m N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}}\Delta N - c_{\mathrm{gr}m0}\frac{g_{m0} N_{\mathrm{phot}m0}}{(1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0})^2}\Delta N_{\mathrm{phot}m} \quad . \tag{7.83}$$

The small signal modulation behaviour can be studied by a time harmonic current modulation. In this case it is worth to replace the time dependent values $\Delta N, \Delta N_{\mathrm{phot}}, \Delta j, F_{\mathrm{N}}$ and F_{phot} against their Fourier transforms $\Delta\widetilde{N}, \Delta\widetilde{N}_{\mathrm{phot}}, \Delta\widetilde{j}, \widetilde{F}_{\mathrm{N}}$ and $\widetilde{F}_{\mathrm{phot}}$. With Fourier transformation time derivatives can be replaced by simple multiplication with $(-i\omega)$. The Fourier transformed small signal single mode rate equations are

$$-i\omega\Delta\widetilde{N}_{\mathrm{phot}m} = \widetilde{F}_{\mathrm{phot}m} + c_{\mathrm{gr}m0}\Gamma_m \frac{a_m N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}}\Delta\widetilde{N} + \left(c\eta_{\mathrm{gr}m0}\Gamma_m \frac{g_{m0} N_{\mathrm{phot}m0}}{(1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0})^2} - \frac{1}{\tau_m} \right) \Delta\widetilde{N}_{\mathrm{phot}m} \tag{7.84}$$

and

$$-i\omega\Delta\widetilde{N} = \widetilde{F}_{\mathrm{N}} + \frac{\eta_{\mathrm{I}}}{qd}\Delta\widetilde{j} - c_{\mathrm{gr}m0}\frac{a_m N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}}\Delta\widetilde{N} - c_{\mathrm{gr}m0}\frac{g_{m0} N_{\mathrm{phot}m0}}{(1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0})^2}\Delta\widetilde{N}_{\mathrm{phot}m} \quad . \tag{7.85}$$

For better overview we set

$$c_{\mathrm{grm0}}\frac{g_{m0}N_{\mathrm{photm0}}}{(1+\epsilon_{\mathrm{g}}N_{\mathrm{photm0}})^2}=\frac{1}{T_{m1}}$$

and

$$c_{\mathrm{grm0}}\frac{a_{m}N_{\mathrm{photm0}}}{1+\epsilon_{\mathrm{g}}N_{\mathrm{photm0}}}=\frac{1}{T_{m2}}$$

giving

$$\Delta\widetilde{N}_{\mathrm{photm}}\left(-i\omega-\frac{\Gamma_m}{T_{m1}}+\frac{1}{\tau_m}\right)=\widetilde{F}_{\mathrm{photm}}+\Gamma_m\frac{\Delta\widetilde{N}}{T_{m2}} \tag{7.86}$$

and

$$\frac{\Delta\widetilde{N}}{T_{m2}}\left(1-i\omega T_{m2}\right)=\widetilde{F}_{\mathrm{N}}+\frac{\eta_{\mathrm{I}}}{qd}\Delta\widetilde{j}-\frac{\Delta\widetilde{N}_{\mathrm{photm}}}{T_{m1}} \tag{7.87}$$

Replacing $\Delta\widetilde{N}$ in (7.86) with (7.87) and rearranging results into

$$\Delta\widetilde{N}_{\mathrm{photm}}=\frac{1}{T_{m2}}\frac{\widetilde{F}_{\mathrm{photm}}(1-i\omega T_{m2})+\Gamma_m\widetilde{F}_{\mathrm{N}}+\Gamma_m\frac{\eta_{\mathrm{I}}}{qd}\Delta\widetilde{j}}{\frac{1}{\tau_m T_{m2}}-\omega^2-i\omega\left(\frac{1}{T_{m2}}+\frac{1}{T_{m1}}\left(\frac{T_{m1}}{\tau_m}-\Gamma_m\right)\right)} \tag{7.88}$$

which is the transfer function of a second order low pass with resonance frequency $f_{\mathrm{r}m}=1/(2\pi\sqrt{T_{m2}\tau_m})$ and damping factor $\gamma_{\mathrm{r}m}=1/\tau_m-\Gamma_m/T_{m1}+1/T_{m2}$.
At high frequencies $\omega T_{m2}\gg 1$

$$\Delta\widetilde{N}_{\mathrm{photm}}\{\omega T_{m2}\gg 1\}\simeq\frac{\frac{1}{T_{m2}}\Gamma_m\frac{\eta_{\mathrm{I}}}{qd}\Delta\widetilde{j}-i\omega\widetilde{F}_{\mathrm{photm}}}{\omega_{\mathrm{r}m}^2-\omega^2-i\omega\gamma_{\mathrm{r}m}} \tag{7.89}$$

results for small signal current modulation. For low frequencies $\omega T_{m2}\ll 1$ the amplitude modulation is

$$\Delta\widetilde{N}_{\mathrm{photm}}\{\omega T_{m2}\ll 1\}\simeq\frac{\widetilde{F}_{\mathrm{photm}}+\Gamma_m\widetilde{F}_{\mathrm{N}}+\Gamma_m\frac{\eta_{\mathrm{I}}}{qd}\Delta\widetilde{j}}{\omega_{\mathrm{r}m}^2}\quad . \tag{7.90}$$

We define a modulation transfer function

$$H_m = \left| \frac{\Delta \widetilde{N}_{\mathrm{phot}m}\{f\}}{\Delta \widetilde{N}_{\mathrm{phot}m}\{f=0\}} \right|^2_{F_{\mathrm{phot}m}=F_{\mathrm{N}}=0} \quad . \tag{7.91}$$

Figure 7.42 shows the normalized small signal transfer function of a typical laser.

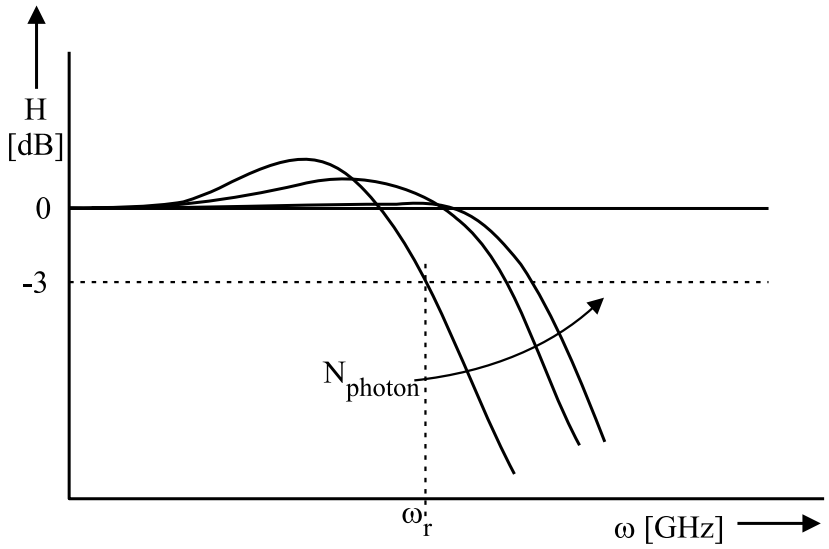

Figure 7.42: Small signal transfer function for a typical semiconductor laser. The resonance frequency and damping grow with increasing steady state currents leading to a higher 3dB frequency with increased output power.

In practice a shift to higher resonance frequencies with increasing pumping is observed. The resonance frequency is

$$f_{\mathrm{r}m} = \frac{1}{2\pi}\sqrt{\frac{1}{\tau T_{m2}}} = \frac{1}{2\pi}\sqrt{\frac{1}{\tau_m} c_{\mathrm{gr}0} \frac{a_m N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}} N_{\mathrm{phot}m0}}} \tag{7.92}$$

Typical resonance frequencies are of the order 10 GHz in semiconductor lasers.
With increasing steady state current j_0 the corresponding photon density $N_{\mathrm{phot}m0}$ increases and therefore the resonance frequency. The resonance frequency depends also on the differential gain coefficient a which is determined by the active material and

structure of the active zone. As we already know in quantum wells the differential gain is higher and therefore a higher resonance frequency can be expected. The problem is that at elevated photon densities the gain saturates and the increase in resonance frequency is partly compensated. The resonance frequency also depends on the photon lifetime. With decreasing photon lifetime the resonance frequency increases but as well the threshold carrier density and therefore the current density increases. When all other parameters are fixed and a linear threshold current-carrier dependence is assumed the threshold current grows quadratic with resonance frequency.
The damping coefficient $\gamma_{\mathrm{r}m}$ in the transfer function is

$$\begin{aligned}\gamma_{\mathrm{r}m} &= c_{\mathrm{gr}0}\frac{1}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}\left[a_m N_{\mathrm{phot}m0} - \Gamma_m \frac{g_{m0}}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}\right] + \frac{1}{\tau_m} \\ &\simeq c_{\mathrm{gr}0}\frac{1}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}\left[a_m N_{\mathrm{phot}m0} - \Gamma_m \frac{g_{\mathrm{th}m}}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}\right] + \frac{1}{\tau} \end{aligned} \tag{7.93}$$

and increases linearly with the steady state photon density thereby limiting the increase in 3bB corner frequency with increasing photon density.

7.8.3.2 Large Signal Modulation

With large signal current modulation the nonlinearities in the rate equations cannot be any longer neglected. There are no analytical solutions but when the current density is switched from $j = j_1 < j_{\mathrm{th}}$ both the begin and the end of the process of photon and carrier density change can be described analytically. When the current switches at time t_0 lasing starts after a delay time τ_{D} at $t_0 + \tau_{\mathrm{D}}$. The carrier density before switching is well below threshold and only spontaneous emission is observed. Lasing starts when the carrier density is near threshold $N \simeq N_{\mathrm{th}}$. The carrier density rate equation without stimulated emission

$$\frac{\partial}{\partial t}N = \frac{j}{qd} - R_{\mathrm{c}}\{N\} \tag{7.94}$$

describes the action before lasing. Due to the weak exitation $R_c\{N\}$ can be replaced by $R_{spon} \simeq n/\tau_s$. The steady state carrier density before switching is $N_1 = \frac{\tau_s}{qd} j_1$. After switching it evolves as

$$N = \frac{\tau_s}{qd}[j_2 - (j_2 - j_1)\exp -\{\frac{t}{\tau_s}\}] \tag{7.95}$$

until $N \simeq N_{th}$ is reached. The delay time can be calculated when t is replaced by τ_D and N by N_{th} resulting into

$$\tau_D = \tau_s \ln\{\frac{j_2 - j_1}{j_2 - j_{th}}\} \tag{7.96}$$

where $j_{th} = N_{th}\frac{\tau_s}{qd}$ has been used. The effect of a delay time can be explained with the picture of a bucket that has holes on its sidewall as sketched in figure 7.43 When Water is poured into the bucket it flows through the holes. Lasing is equivalent to the overflow when the bucket is full. The filling rate is then divided into the amount that flows through the holes (threshold) and the amount that gives overflow. When the filling rate is below the threshold value a steady state amount of wafer can be measured in the bucket. After switching the flow above threshold the wafer level increases until the top is reached and overflow occurs. The time needed from switching to begin of overflow is the delay time. When the bucked has outlets of different depth each outlet represents a mode of lasing.

Long time after switching each mode is at its new steady state value. Switching can be regarded as a exitation for the modes. From the small signal behaviour of a single mode laser we know that the modes can be treated like damped oscillators. When an oscillator is switched from one steady state to another one its transient response is a damped oscillation around the new value.

Figure 7.44 shows the numerical calculated switching behaviour of a laser with three modes considered.

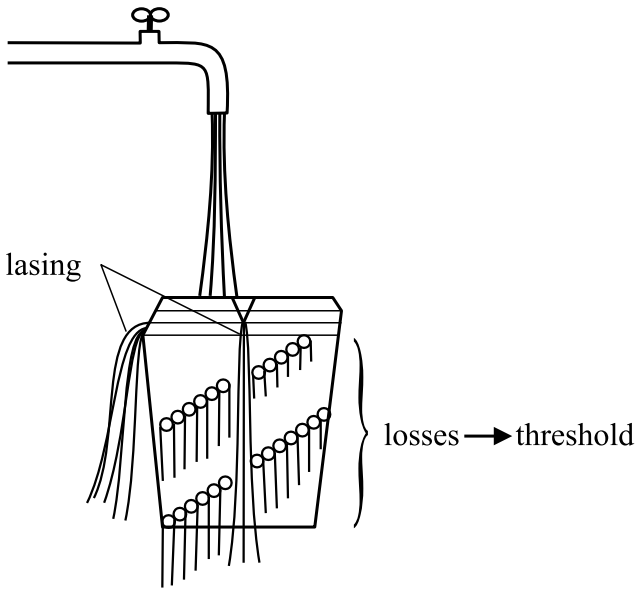

Figure 7.43: Model for switching of a laser. As long as the filling rate is below threshold a steady state level can be observed in the bucket. Above threshold the different modes (outlets on the rim of the bucket) are exited.

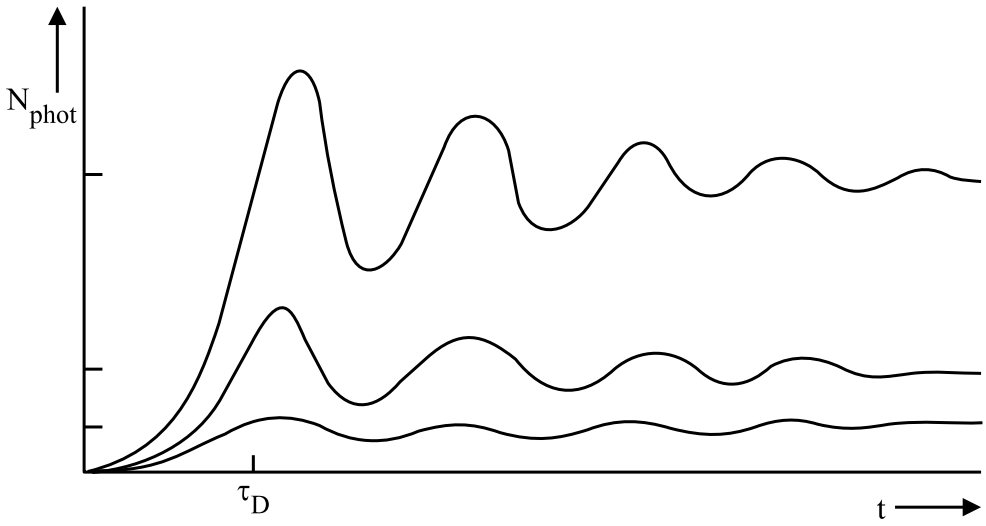

Figure 7.44: Numerical result for switching from $j = 0$ to $j = 1.5 j_{\text{th}}$ in a three mode laser. The damped oscillation occurs with the frequency already calculated in the preceding section.

7.8.4 Noise in Semiconductor Lasers

Noise in semiconductor lasers originates from the spontaneous emission which results from statistical recombination processes. The noise is always present because each carrier can "decide" weather it recombines stimulated or spontaneous. As has been discussed already in section 7.3 the noise can be described by the Langevin forces. Due to the fact that the spontaneous emission is small compared to the stimulated emission the Langevin forces can be considered in the small signal rate equations of the preceding section. The photon and carrier Langevin forces lead mainly to amplitude noise whereas the phase Langevin force is the main contribution to spectral broadening of the laser line.

7.8.4.1 Relative Intensity Noise

From (7.88) we know that the amplitude noise has in principle the same spectral behaviour as the small signal current modulation. The noise to signal ratio is expressed by the relative intensity noise RIN function

$$RIN = \left| \frac{\Delta \widetilde{N}_{\text{phot}}\{f\}B}{N_{\text{phot0}}} \right|^2_{\Delta j=0} \quad . \tag{7.97}$$

The RIN gives the photon density noise at a certain frequency compared to the steady state photon density. The definition of RIN follows from the fact that in the detector the light is converted into a current and the signal and noise power in the detector is proportional to the quadratic current.

For the relative intensity noise $|\widetilde{F}_{\text{phot}}|^2$ and $|\widetilde{F}_{\text{N}}|^2$ must be calculated. We remember the Wiener-Kinchine-theorem (see appendix E) for ergodic statistical processes $a\{t\}$ that connects the power density spectrum

$$|\widetilde{a}\{f\}|^2 = \int_{-\infty}^{\infty} |a\{t\}|^2 \exp\{i2\pi ft\}\mathrm{d}t$$

$$\simeq \int_{-\infty}^{\infty} \langle a\{t+T\}a^*\{t\}\rangle \exp\{i2\pi fT\}\mathrm{d}T \tag{7.98}$$

with Bandwidth $B = 1/T$ to the autocorrelation function

$$\langle a\{t+T\}a^*\{t\}\rangle = \frac{1}{T}\int_0^T a\{t+t'\}a^*\{t\}\mathrm{d}t' \quad . \tag{7.99}$$

For the noise calculation $a\{t\}$ has to be replaced by $a_m\{t\} = E_m\{t\}\exp\{i(\phi_m\{t\} - \omega_m\{t\}t)\}$ where amplitude, frequency and phase noise are considered. The autocorrelation is

$$\langle a_m\{t+T\}a_m^*\{t\}\rangle = \frac{1}{T}\int_0^T \langle E_m\{t+T\}E_m^*\{t\}\rangle \exp\{i(\phi_m\{t+T\} - \phi_m\{t\} - \omega_m T)\}\mathrm{d}t$$
$$. \tag{7.100}$$

Here the amplitude noise is of interest and therefore the frequency and phase noise are neglected as well as the phase shift $\omega_m T$ leading to

$$\langle a_m\{t+T\}a_m^*\{t\}\rangle \simeq \frac{1}{T}\int_0^T E_m\{t+t'\}E_m^*\{t\}\mathrm{d}t \tag{7.101}$$

with power density spectrum

$$|\widetilde{a}_m\{f\}|^2 \simeq \int_{-\infty}^{\infty} \langle E_m\{t+T\}E_m^*\{t\}\rangle \exp\{i2\pi fT\}\mathrm{d}T \tag{7.102}$$

The photon noise is generated by Langevin forces that describe the spontaneous emission contribution to the electric field. The noise spectrum follows when $|\widetilde{a}_m\{f\}|^2$ is replaced by $\Delta\widetilde{N}_{\mathrm{phot}m}\{f\}\Big|_{\Delta j=0}$ and $E_m\{t\}$ is substituted by $F_{\mathrm{phot}m}\{t\}$.

The photon noise rate autocorrelation function follows from (7.6) with (7.99) to

$$\langle F_{\mathrm{phot}m}\{t+T\}F^{*}_{\mathrm{phot}m}\{t\}\rangle = \frac{N_{\mathrm{phot}m0}}{V_{\mathrm{active}}}\frac{4}{T}\left(\sum_{j=1}^{m_{\mathrm{spon}}}\cos\{\theta_j\}\right)^2\delta\{T\} \tag{7.103}$$

and gives the power density spectrum

$$|\widetilde{F}_{\mathrm{phot}m}|^2 = \frac{N_{\mathrm{phot}m0}}{V_{\mathrm{active}}}\frac{4}{T}\left(\sum_{j=1}^{m_{\mathrm{spon}}}\cos\{\theta_j\}\right)^2 = \frac{N_{\mathrm{phot}0}}{V_{\mathrm{active}}}\frac{2m_{\mathrm{spon}}}{T} \tag{7.104}$$

Within time T the generation of m_{spon} spontaneous emitted photons that contribute to mode m in the active region can be described by

$$\frac{m_{\mathrm{spon}}}{TV_{\mathrm{active}}} = \Gamma_m\beta_m R_{\mathrm{spon}}\{N_0\} \tag{7.105}$$

giving

$$|\widetilde{F}_{\mathrm{phot}m}|^2 = 2\Gamma_m\beta_m R_{\mathrm{spon}}\{N_0\}N_{\mathrm{phot}m0} \simeq 2\Gamma_m\beta_m R_{\mathrm{spon}}\{N_{\mathrm{th}}\}N_{\mathrm{phot}m0} \tag{7.106}$$

The power density of amplitude noise grows linearly with increasing photon density and therefore with emitted power but the RIN decreases with emitted power

$$RIN \simeq \frac{1}{N_{\mathrm{phot}m0}}\frac{4\Gamma_m\beta_m R_{\mathrm{spon}}\{N_{\mathrm{th}}\}}{|\omega_{\mathrm{r}m}^2 - \omega^2 + i\gamma_{\mathrm{r}m}\omega|^2}B \quad . \tag{7.107}$$

Typical values for the RIN are around $-130\,\mathrm{dB}$ in modern semiconductor lasers.

7.8.4.2 Phase Noise and Linewidth

For the calculation of phase noise and linewidth we start from (7.105). The linewidth $\delta\nu$ is defined as FWHM of the emission spectrum $|\widetilde{a}_m\{\nu_{m0}\pm\delta\nu/2\}|^2 = |\widetilde{a}_m\{\nu_{m0}\}|^2/2$ where ν_{m0} denotes the center frequency of emission. It can be assumed that the amplitude

noise is frequency independent in the spectral range of interest which translates into $E\{t\} = E\{t+T\}$ and therefore $E^*\{t\}E\{t+T\} \simeq N_{\mathrm{phot}m0}E_0^2$. The averaged emission is therefore proportional to

$$\begin{aligned}\langle a_m\{t+T\}a_m^*\{t\}\rangle &\simeq N_{\mathrm{phot}m}E_0^2\frac{1}{T}\int_0^T \exp\{i(\phi_m\{t+T\}-\phi_m\{t\}-\omega_{m0}T)\}\mathrm{d}t \\ &= N_{\mathrm{phot}m}E_0^2\frac{1}{T}\int_0^T \exp\{i(\Delta\phi_m\{t,T\}-\omega_{m0}T)\}\mathrm{d}t \quad . \end{aligned} \tag{7.108}$$

In principle a simple Fourier transformation leads to the required power density spectrum when T is made sufficient large, that is the bandwidth $B = 1/T$ must be much less than the spectral linewidth $\delta\nu$. The problem is that the phase varies statistically. We can assume an ergodic process here and therefore the time average can be replaced by a state average

$$\frac{1}{T}\int_0^T f\left\{x\left\{t\right\}\right\}\mathrm{d}t = \int_{-\infty}^{\infty} f\{x\}p\{x\}\mathrm{d}x \tag{7.109}$$

when the probability density of the states is known. It is reasonable to assume a gaussian probability density for the phase change after time T

$$p\{\Delta\phi\{t,T\}\} = \frac{1}{\sqrt{2\pi\langle\Delta\phi^2\{T\}\rangle}}\exp\left\{-\frac{\Delta\phi^2\{t,T\}}{2\langle\Delta\phi^2\{T\}\rangle}\right\} \quad . \tag{7.110}$$

The averaged power density follows then to be

$$\begin{aligned}\langle a_m\{t+T\}a_m^*\{t\}\rangle \simeq\ & N_{\mathrm{phot}m}E_0^2\frac{1}{\sqrt{2\pi\langle\Delta\phi_m^2\{T\}\rangle}} \\ & \int_{-\infty}^{\infty}\exp\left\{i(\Delta\phi_m\{t,T\}-\omega_{m0}T)\right\}\exp\left\{-\frac{\Delta\phi^2\{t,T\}}{2\langle\Delta\phi_m^2\{T\}\rangle}\right\}\mathrm{d}\Delta\phi_m \\ =\ & N_{\mathrm{phot}m}E_0^2\exp\left\{-\frac{1}{2}\langle\Delta\phi_m^2\{T\}\rangle - i\omega_{m0}T\right\}\end{aligned} \tag{7.111}$$

with power density spectrum (see appendix E)

$$|\widetilde{a}_m\{\nu\}|^2 = \lim_{T\to\infty}\frac{2}{B}|\widetilde{a}_{\mathrm{T}m}\{\nu\}|^2$$
$$= \lim_{T\to\infty} 2N_{\mathrm{phot}m}E_0^2\left|\int_0^T \exp\left\{-\frac{1}{2}\langle\Delta\phi_m^2\{T'\}\rangle + i(\omega-\omega_{m0})T'\right\}\mathrm{d}T'\right| \quad (7.112)$$

As we know already the phase change is attributed to the phase Langevin force and we can calculate the averaged phase change within time T starting from (7.80)

$$\frac{\partial}{\partial t}\Delta\phi_{\mathrm{c}m} \simeq F_{\Phi m} + \frac{c_0 a_m}{2\eta_{\mathrm{eff}m0}}(\alpha_{\mathrm{H}m}-i)\left(\Delta N + \frac{z\eta_{\mathrm{eff}m0}}{c_0}\frac{\partial}{\partial t}\Delta N\right) \quad (7.113)$$

and assume that the carrier density change ΔN stays time constant ($\frac{\partial}{\partial t}\Delta N \simeq 0$) In single mode lasers this leads under the assumption of negligible light amplitude modulation ($\Delta N_{\mathrm{phot}} \ll \Delta N$) with (7.83) to

$$F_{\mathrm{N}} - c_{\mathrm{gr}m0}\frac{a_m N_{\mathrm{phot}m0}}{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}\Delta N = 0 \quad (7.114)$$

and therefore results

$$\frac{\partial}{\partial t}\Delta\phi_{\mathrm{c}m} \simeq F_{\Phi m} + \frac{c_0 a_m}{2\eta_{\mathrm{eff}m0}}(\alpha_{\mathrm{H}m}-i)\frac{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}{c_{\mathrm{gr}m0}a_m N_{\mathrm{phot}m0}}F_{\mathrm{N}}$$
$$= F_{\Phi m} + \frac{c_{\mathrm{phase}m0}}{c_{\mathrm{gr}m0}}(\alpha_{\mathrm{H}m}-i)\frac{1+\epsilon_{\mathrm{g}}N_{\mathrm{phot}m0}}{2N_{\mathrm{phot}m0}}F_{\mathrm{N}} \quad . \quad (7.115)$$

Only the real part of $\Delta\phi_{rmc}$ has to be calculated. The imaginary part leads to amplitude noise that has been dealt with in the preceding section. The averaged phase change within interval T follows from

$$\Delta\Phi_m\{t,T\} = \int_t^{t+T}\frac{\partial}{\partial t''}\mathrm{Re}\left\{\Delta\phi_{\mathrm{c}m}\right\}\Big|_{t'}\mathrm{d}t' \quad . \quad (7.116)$$

The carrier Langevin force $F_{\rm N}$ is unknown. As a lower limit it can be replaced by the photon Langevin force $F_{\rm phot} \leq F_{\rm N}$ and

$$\Delta\Phi_m\{t,T\} \simeq \frac{1}{\sqrt{N_{{\rm phot}m0}}} \sum_{j=j_1}^{j_2} \sin\{\theta_j\} - \alpha_{{\rm H}m} \frac{c_{{\rm phase}m0}}{c_{{\rm gr}m0}} (1+\epsilon_{\rm g} N_{{\rm phot}m0}) \cos\{\theta_j\} \tag{7.117}$$

results. The number of spontaneous emission acts within T is $j_2 - j_1 = \beta_{\rm spon} R_{\rm spon}\{N_0\} T$ and the quadratic phase change follows from

$$\Delta\Phi_m^2\{t,T\} \simeq \frac{j_2-j_1}{2} \frac{1}{\sqrt{N_{{\rm phot}m0}}} \left[1 + \alpha_{{\rm H}m}^2 \left(\frac{c_{{\rm phase}m0}}{c_{{\rm gr}m0}} \right)^2 (1+\epsilon_{\rm g} N_{{\rm phot}m0})^2 \right] \tag{7.118}$$

which is obtained when equally distributed θ_j are assumed as is reasonable. The averaged quadratic phase change

$$\langle \Delta\Phi_m^2\{t,T\} \rangle \simeq \frac{\beta_{\rm spon} R_{\rm spon}\{N_0\}}{2\sqrt{N_{{\rm phot}m0}}} \left[1 + \alpha_{{\rm H}m}^2 \left(\frac{c_{{\rm phase}m0}}{c_{{\rm gr}m0}} \right)^2 (1+\epsilon_{\rm g} N_{{\rm phot}m0})^2 \right] T \tag{7.119}$$

grows linear with time interval length leading to the emission spectrum

$$|\widetilde{a}_m\{\nu\}|^2 = 2 N_{{\rm phot}m0} E_0^2 \left| \frac{1}{\frac{\beta_{\rm spon} R_{\rm spon}\{N_0\}}{4\sqrt{N_{{\rm phot}m0}}} \left[1 + \alpha_{{\rm H}m}^2 \left(\frac{c_{{\rm phase}m0}}{c_{{\rm gr}m0}} \right)^2 (1+\epsilon_{\rm g} N_{{\rm phot}m0})^2 \right] + i(\omega_{m0} - \omega)} \right| \tag{7.120}$$

The FWHM follows when ω is replaced by $\omega_{m0} \pm \pi\delta\nu$ giving the linewidth

$$\delta\nu = \frac{\beta_{\rm spon} R_{\rm spon}\{N_0\}}{4\pi} \left[1 + \alpha_{{\rm H}m}^2 \left(\frac{c_{{\rm phase}m0}}{c_{{\rm gr}m0}} \right)^2 (1+\epsilon_{\rm g} N_{{\rm phot}m0})^2 \right] \frac{1}{N_{{\rm phot}m0}} \tag{7.121}$$

Above threshold N_{phot} can be replaced by $\frac{\eta_{\text{I}}}{c_{\text{gr}} g_{\text{th}} q d}(j - j_{\text{th}})$. The main tendency for the linewidth can be described therefore by

$$\delta\nu = \frac{1 + \alpha_{\text{H}m}^2}{j - j_{\text{th}}} \tag{7.122}$$

which says that the linewidth decreases with increasing current density. It is quadratic proportional to the Henry factor α_{H} leading to the alternative name linewidth enhancement factor.

7.8.5 Edge-Emitting Lasers

7.8.6 High-Power Edge-Emitting Lasers

7.8.7 Vertical-Cavity Surface-Emitting Lasers

7.8.8 Quantum-Cascade Lasers

Chapter 8

Detectors

8.1 Thermal Detectors

8.1.1 Thermoelectric Detectors

8.1.2 Bolometer

8.1.3 Pneumatic Detectors

8.1.4 Pyroelectric Detectors

8.2 Photon-Effect Detectors

8.2.1 Photoemission Detectors

8.2.1.1 Vakuum Photodiodes

8.2.1.2 Photomultipliers

8.2.1.3 Image Intensifiers

8.2.2 Photoconductive Detectors

8.2.2.1 Vidikons

8.2.2.2 Multi-Quantum-Well Detectors

8.2.3 Junction Detectors

8.2.3.1 pn-Photodetectors

8.2.3.2 pin-Photodetectors

8.2.3.3 Schottky-Photodiodes

8.2.3.4 Avalanche-Photodiodes

8.2.3.5 Multilayer Photodetectors

8.2.3.6 Resonant-Cavity Photodetectors

8.2.3.7 Charge-Coupled-Devices

8.2.3.8 Solar Cells

8.2.3.9 Liquid-Crystal Light-Valves

Chapter 9

Optical Communications Systems

9.1 Modulation Schemes

9.2 Noise Considerations

9.3 Lasers

9.4 Detectors

9.5 Fiber-Optical Communications

9.6 Free-Space Communications

9.7 Integrated Optics

Appendix A

Wave Velocities

The propagation of time harmonic waves is described by the argument of $\exp\{i(\vec{k} \circ \vec{r} - \omega t)\}$. The propagation velocity $\vec{v}$ is the speed at which points with constant phase $\phi = \mathrm{Re}\left\{\vec{k} \circ \vec{r}\right\} - \omega t$ travel and follows from

$$\frac{\partial}{\partial t}\phi = 0 = \mathrm{Re}\left\{\vec{k}\right\} \circ \vec{v} - \omega \quad . \tag{A.1}$$

Normally only the propagation in z- direction is regarded. The phase velocity is the speed that points with constant phase travel in z- direction $c_{\mathrm{phase}} = \vec{v} \circ \vec{e}_{\mathrm{z}}$. With

$$k_{\mathrm{z}}^2 = (\vec{k} \circ \vec{e}_{\mathrm{z}}) = k^2 - k_\rho^2 = \eta_{\mathrm{eff}}^2 k_0^2 \tag{A.2}$$

the phase velocity c_{phase} results similar to (A.1) from $\mathrm{Re}\{k_{\mathrm{z}}\}\, c_{\mathrm{phase}} - \omega = 0$ and is defined as

$$c_{\mathrm{phase}} := \frac{1}{\frac{1}{\omega}\mathrm{Re}\{k_{\mathrm{z}}\}} = \frac{\omega}{\mathrm{Re}\{k_{\mathrm{z}}\}} = \frac{c_0}{\mathrm{Re}\{\eta_{\mathrm{eff}}\}} = \frac{c_0}{\bar{n}_{\mathrm{eff}}} \quad . \tag{A.3}$$

The energy transport in z-direction can be expressed with the $z-$component of the Pointing vector

$$S_z = c_{gr}(w_{el} + w_{magn}) \tag{A.4}$$

where the group velocity is defined as

$$c_{gr} := \frac{1}{\frac{\partial}{\partial\omega}\mathrm{Re}\{k_z\}} \quad . \tag{A.5}$$

With (A.4) the group velocity can be interpreted as energy transport velocity. An expresssion for the group velocity follows, when (A.2) is used in (A.5). In dispersive media $\frac{\partial}{\partial\omega}\eta \neq 0$ follows

$$\begin{aligned}\frac{\partial}{\partial\omega}k_z^2 = 2k_z\frac{\partial}{\partial\omega}k_z = 2k\frac{\partial}{\partial\omega}k - 2k_\rho\frac{\partial}{\partial\omega}k_\rho = 2\frac{\omega\eta}{c_0^2}\frac{\partial}{\partial\omega}(\omega\eta) - 2k_\rho\frac{\partial}{\partial\omega}k_\rho\Big|_{\frac{\partial}{\partial\omega}k_\rho=0} \\ = 2\frac{\omega\eta}{c_0^2}\left(\eta + \omega\frac{\partial}{\partial\omega}\eta\right) \quad .\end{aligned} \tag{A.6}$$

With the group index

$$\eta_{gr} = \eta + \omega\frac{\partial}{\partial\omega}\eta \tag{A.7}$$

the second and the last term in (A.7) give with (A.5)

$$c_{gr} = c_0\mathrm{Re}\left\{\frac{\eta_{eff}}{\eta\eta_{gr}}\right\} \quad . \tag{A.8}$$

In lossless media ($\mathrm{Im}\{\eta\} = 0$)follows with (A.3)

$$c_{gr}c_{phase} = \frac{c_0^2}{\bar{n}\cdot\bar{n}_{gr}} = c^2\frac{\bar{n}}{\bar{n}_{gr}} \tag{A.9}$$

which simplifies in non dispersive media to the well known relation $c_{gr}c_{phase} = c^2$.

It would be nice if we could find an effective group index which relates the group velocity to the vacuum light velocity. From $k_z = \frac{\omega \eta_{\text{eff}}}{c_0}$ we define similar to (A.7)

$$\eta_{\text{gr eff}} = \eta_{\text{eff}} + \omega \frac{\partial}{\partial \omega} \eta_{\text{eff}} \quad . \tag{A.10}$$

With (A.8) we find $\eta_{\text{gr eff}} = \frac{\eta \eta_{\text{gr}}}{\eta_{\text{eff}}}$ and with $\bar{n}_{\text{gr eff}} = \text{Re}\{\eta_{\text{gr eff}}\}$ we can set therefore

$$c_{\text{gr}} = \frac{c_0}{\text{Re}\{\eta_{\text{gr eff}}\}} = \frac{c_0}{\bar{n}_{\text{gr eff}}} \quad . \tag{A.11}$$

It has to be noted here that **the energy transport speed of time harmonic waves in dispersive media is no longer the group velocity** c_{gr} but it is a good measure in the most cases.

Appendix B

Dispersion in Power Law Profile Fibers

For calculation of the dispersion in power law profile fibers the high order mode k_ρ must be calculated. In this case the number of modes that are carrying the signal at a given k_z gives the required measure. All modes from $n = 0$ to $n = n_{\max} = \sqrt{\eta^2 k_0^2 - k_z^2} \cdot \varrho$ are included in a ring area with radius between a_1 and ϱ. There are two polarizations (TE and TM) and two orientations of the transverse field (x- or y- direction) giving for each modenumber m 4 possible waves. In highly multimode fibers the propagation in z- direction follows $k_z \rightarrow k_0 \cdot \eta_2$) in the case $m \gg 1$ and we find the number of modes M by summation of all azimuthal orders with the use of (3.87) $M \simeq \frac{4}{\pi} \sum_{m=0}^{m_{\max}} \int_{a_1}^{a_2} k_{\varrho m} \, \mathrm{d}\varrho$. The azimuthal order again is a big number and we can assume it to be nearly continuous and replace the summation by integration

$$M \simeq \frac{4}{\pi} \int_{a_1}^{a_2} \int_{0}^{n_{\max}} \sqrt{k^2 - k_z^2 - \left(\frac{n}{\varrho}\right)^2} \, \mathrm{d}n \, \mathrm{d}\varrho = \int_{a_1}^{a_2} (\eta^2 k_0^2 - k_z^2) \cdot \varrho \, \mathrm{d}\varrho \quad .$$

Replacing η^2 by (3.86) we find with $a_1 = 0$ and $\left(\frac{a_2}{a}\right)^g = \frac{\eta_1^2 k_0^2 - k_z^2}{(\eta_1^2 - \eta_2^2) k_0^2}$

$$M = k_0^2(\eta_1^2 - \eta_2^2) \cdot \int_0^{a_2} \left(\left(\frac{a_2}{a}\right)^g - \left(\frac{\varrho}{a}\right)^g \right) \cdot \varrho \, \mathrm{d}\varrho = \frac{V^2}{2} \cdot \frac{g}{g+2} \cdot \left(\frac{a_2}{a}\right)^{g+2} \quad .$$

With the use of fiber and phase parameters $V = k_0 \cdot a \cdot \sqrt{\eta_1^2 - \eta_2^2}$ and $B = \frac{k_z^2 - \eta_2^2 k_0^2}{(\eta_1^2 - \eta_2^2)k_0^2}$ the compact representation

$$M = \frac{V^2}{2}\frac{g}{g+2} \cdot \left(\frac{\eta_1^2 k_0^2 - k_z^2}{(\eta_1^2 - \eta_2^2)k_0^2} \right)^{\frac{g+2}{g}} = \frac{V^2}{2}\frac{g}{g+2} \cdot (1-B)^{\frac{g+2}{g}}$$

follows. Under assumption that the number of guided modes does not change remarkably with changing frequency $\partial M/\partial\omega \simeq 0$ the change must happen in k_ρ. Differentiation by ω therefore must give the required value and results into

$$0 = m \left[\frac{1}{V^2} \cdot \frac{\mathrm{d}}{\mathrm{d}\omega} V^2 + \frac{g}{g+2} \cdot \frac{1}{1-B} \cdot \frac{\mathrm{d}}{\mathrm{d}\omega} (1-B) \right]$$

giving

$$\eta_1\eta_1' - \eta_2\eta_2' + \frac{g+2}{g} \cdot \frac{\eta_1\eta_1' - \frac{k_z c_0}{k_0} \cdot \tau}{1-B} - \frac{g+2}{g} \cdot (\eta_1\eta_1' - \eta_2\eta_2') = 0$$

and by rearrangement the dispersion

$$\tau = \frac{1}{c_0}\frac{k_0}{k_z} \left(\eta_1\eta_1' - \frac{2}{g+2}(\eta_1\eta_1' - \eta_2\eta_2')(1-B) \right) \tag{B.1}$$

follows. This holds for $k_z \to \eta_2 \cdot k_0$. At $k_z = \eta_1 \cdot k_0 \sqrt{\frac{\eta_1'}{\eta_1} \cdot \frac{\eta_1^2 - \eta_2^2}{\eta_1\eta_1' - \eta_2\eta_2'} \frac{g+2}{g} - 1}$ the travelling time τ is minimal

$$\tau_{\min} = \frac{2\eta_1}{c_0} \cdot \frac{2}{g+2} \cdot \frac{\eta_1\eta_1' - \eta_2\eta_2'}{\eta_1^2 - \eta_2^2} \sqrt{\frac{\eta_1'}{\eta_1} \cdot \frac{\eta_1^2 - \eta_2^2}{\eta_1\eta_1' - \eta_2\eta_2'} \cdot \frac{g+2}{2} - 1} \quad .$$

At $k_z = \eta_2 k_0$ $(B = 0)$ we have

$$\tau\{B=0\} = \tau_0 = \frac{1}{c_0} \cdot \left(\frac{\eta_1 \eta_1'}{\eta_2} - \frac{2}{g+2} \cdot \frac{\eta_1 \eta_1' - \eta_2 \eta_2'}{\eta_2} \right) \tag{B.2}$$

The difference $\tau_0 - \tau_{\min}$ is minimal at $\frac{g+2}{2} = 2\frac{\eta_1}{\eta_1'} \cdot \frac{\eta_1+\eta_2}{3\eta_1+\eta_2} \cdot \frac{\eta_1\eta_1'-\eta_2\eta_2'}{\eta_1^2-\eta_2^2}$
but then the square root in $\tau_{\min}$ is imaginary and this is no solution of the problem. The right solution is found when the highest order modes travel as fast as the lowest order ones. This is $\tau\{B=0\} = \frac{\eta_1'}{c_0}$ and from this $\frac{2}{g+2} = \frac{\eta_1'(\eta_1-\eta_2)}{\eta_1\eta_1'-\eta_2\eta_2'}$

$$g_{\text{opt}} = 2 \cdot \frac{\eta_2}{\eta_1} \cdot \frac{(\eta_1' - \eta_2')}{(\eta_1 - \eta_2)} \tag{B.3}$$

and

$$\tau_{\min} = \frac{\eta_1'}{c_0} \cdot \frac{2\sqrt{\eta_1\eta_2}}{\eta_1+\eta_2} \simeq \frac{\eta_1'}{c_0} \left[1 - \frac{1}{8} \left(\frac{\eta_1 - \eta_2}{\eta_1} \right)^2 \right] \tag{B.4}$$

give the total dispersion of

$$\Delta\tau \simeq -\frac{1}{8} \left(\frac{\eta_1 - \eta_2}{\eta_1} \right)^2 \frac{\eta_1'}{c_0} \quad .$$

Appendix C

Reciprocity, Orthogonality and Normalization of Optical Modes

Optical modes in waveguides do not need to be pure TE or TM modes. We therefore use a representation for electric and magnetic fields which splits the fields into a $z-$component and a transverse one which is orthogonal to z. This representation obliges that the waveguide extends infinitely into $z-$direction. Furthermore we only look at steady state monochromatic waves using $\exp\{-i\,\omega t\}$ as time dependence. In the Maxwell equations the nabla operator $\nabla = \frac{\partial}{\partial x}\vec{e}_{\mathrm{x}}\frac{\partial}{\partial y}\vec{e}_{\mathrm{y}}\frac{\partial}{\partial z}\vec{e}_{\mathrm{z}}$ is used. With field propagation in z- direction the z- derivative can be substituted by $\pm ik_{\mathrm{z}}$. For simpler notation we define a transversal nabla operator

$$\nabla_{\mathrm{t}} = \frac{\partial}{\partial x}\cdot\vec{e}_{\mathrm{x}} + \frac{\partial}{\partial y}\cdot\vec{e}_{\mathrm{y}}$$

such that ∇ can be replaced by $\nabla_{\mathrm{t}} + \frac{\partial}{\partial z}\vec{e}_{\mathrm{z}} = \nabla_{\mathrm{t}} \pm ik_{\mathrm{z}}$. The electric and magnetic fields are expanded into

$$\begin{aligned}\vec{H} &= (\vec{H}_{\mathrm{t}} + H_{\mathrm{z}}\cdot\vec{e}_{\mathrm{z}})\cdot\exp\{-i\,\omega t\} = \vec{H}_{\mathrm{A}}\{x,y,z\}\cdot\exp\{-i\,\omega t\}\\ \vec{E} &= (\vec{E}_{\mathrm{t}} + E_{\mathrm{z}}\cdot\vec{e}_{\mathrm{z}})\cdot\exp\{-i\,\omega t\} = \vec{E}_{\mathrm{A}}\{x,y,z\}\cdot\exp\{-i\,\omega t\} \quad . \end{aligned} \tag{C.1}$$

The Maxwell equations can now be rewritten in the form

$$\begin{aligned}
\nabla_{\mathrm{t}} \times \vec{E}_{\mathrm{t}} &= i\,\omega\mu\mu_0 H_{\mathrm{z}} \cdot \vec{e}_{\mathrm{z}} \\
\nabla_{\mathrm{t}} \times \vec{H}_{\mathrm{t}} &= -i\,\omega\varepsilon\varepsilon_0 E_{\mathrm{z}} \cdot \vec{e}_{\mathrm{z}} \\
\nabla_{\mathrm{t}} \times E_{\mathrm{z}} \cdot \vec{e}_{\mathrm{z}} + \frac{\partial}{\partial z}\vec{E}_{\mathrm{t}} &= i\,\omega\mu\mu_0 \vec{H}_{\mathrm{t}} \\
\nabla_{\mathrm{t}} \times H_{\mathrm{z}} \cdot \vec{e}_{\mathrm{z}} + \frac{\partial}{\partial z}\vec{H}_{\mathrm{z}} &= -i\,\omega\varepsilon\varepsilon_0 \vec{E}_{\mathrm{t}} \quad . \qquad \text{(C.2)}
\end{aligned}$$

Modes of order m have the propagation constant $k_{\mathrm{z}m}$. We assume only one mode m with the field representations

$$\begin{aligned}
\vec{E} &= (\vec{E}_{\mathrm{t}m} + E_{\mathrm{z}m} \cdot \vec{e}_{\mathrm{z}}) \cdot \exp\{i(k_{\mathrm{z}m}\, z - \omega t)\} \\
&= \vec{E}_m\{x, y\} \cdot \exp\{i(k_{\mathrm{z}m}\, z - \omega t)\} \\
\vec{H} &= (\vec{H}_{\mathrm{t}m} + H_{\mathrm{z}m} \cdot \vec{e}_{\mathrm{z}}) \cdot \exp\{i(k_{\mathrm{z}m}\, z - \omega t)\} \\
&= \vec{H}_m\{x, y\} \cdot \exp\{i(k_{\mathrm{z}m}\, z - \omega t)\} \qquad \text{(C.3)}
\end{aligned}$$

such that we can use $\frac{\mathrm{d}}{\mathrm{d}z}\vec{E} = i\,k_{\mathrm{z}m}\vec{E}$ and $\frac{\mathrm{d}}{\mathrm{d}z}\vec{H} = i\,k_{\mathrm{z}m}\vec{H}$. The phase choice in (C.3) is somewhat arbitrary because all combinations $\pm(\pm k_{\mathrm{z}} z - \omega t)$ satisfy the wave equation. This means that the choice of time as well of propagation direction are arbitrary. We attribute fields $\vec{E}_1$ and $\vec{H}_1$ to the choice of time and propagation direction in (C.3) and compare them to fields with inverse time $t' = -t$

$$\begin{aligned}
\vec{E}_1\{\vec{r}, t\} &= \vec{E}\{\vec{r}, t\} \\
\vec{H}_1\{\vec{r}, t\} &= \vec{H}\{\vec{r}, t\} \\
\vec{E}_2\{\vec{r}, t\} &= \vec{E}\{\vec{r}, t'\} = \vec{E}_1\{\vec{r}, t\} \\
\vec{H}_2\{\vec{r}, t\} &= \vec{H}\{\vec{r}, t'\} = -\vec{H}_1\{\vec{r}, t\} \quad .
\end{aligned}$$

The amplitudes in representation (C.1) are therefore $\vec{E}_{\mathrm{A},2} = \vec{E}^*_{\mathrm{A},1}$ and $\vec{H}_{\mathrm{A},2} = -\vec{H}^*_{\mathrm{A},2}$.

With reversal of the $z-$coordinate $z' = -z$ the Maxwell-equation require for the fields $\vec{E}_3$ and $\vec{H}_3$

$$\begin{aligned}
\vec{E}_{3,\mathrm{t}}\{\vec{r},t\} &= \vec{E}_{\mathrm{t}}\{x,y,-z,t\} = \vec{E}_{1,\mathrm{t}}\{\vec{r},t\} \\
\vec{H}_{3,\mathrm{t}}\{\vec{r},t\} &= \vec{H}_{\mathrm{t}}\{x,y,-z,t\} = -\vec{H}_{1,\mathrm{t}}\{\vec{r},t\} \\
E_{3,\mathrm{z}}\{\vec{r},t\} &= E_{\mathrm{z}}\{x,y,-z,t\} = E_{1,\mathrm{z}}\{\vec{r},t\} \\
H_{3,\mathrm{z}}\{\vec{r},t\} &= H_{\mathrm{z}}\{x,y,-z,t\} = H_{1,\mathrm{z}}\{\vec{r},t\} \quad .
\end{aligned} \tag{C.4}$$

Both time- or $z-$reversal give for guided modes $k_{zm} = \beta_m$ a backward traveling wave and therefore in this case E_2 and E_3 must be identical such that a comparison gives

$$\begin{aligned}
\vec{E}_{\mathrm{t}m} = \vec{E}^*_{\mathrm{t}m} \quad , \quad \vec{H}_{\mathrm{t}m} = \vec{H}^*_{\mathrm{t}m} \\
E_{\mathrm{z}m} = -E^*_{\mathrm{z}m} \ , \ H_{\mathrm{z}m} = -H^*_{\mathrm{z}m}
\end{aligned} \quad . \tag{C.5}$$

C.1 Reciprocity

The reciprocity of two different sets of fields can be shown by basic Maxwell equations. Let us consider two different fields $\vec{E}_1$ and $\vec{E}_2$ and the fitting fields $\vec{H}_1$ and $\vec{H}_2$ which solve the Maxwell equations

$$\nabla \times \vec{E}_1 = i\,\omega\mu\mu_0\vec{H}_1 \tag{C.6}$$

$$\nabla \times \vec{H}_2 = -i\,\omega\varepsilon\varepsilon_0\vec{E}_2 \quad . \tag{C.7}$$

Conjugating (C.7) and multiplying (C.6) with H_2^* and (C.7) with E_1 gives after substraction

$$\nabla \circ (\vec{E}_1 \times \vec{H}_2^*) = -i\,\omega(\varepsilon\varepsilon_0\vec{E}_1 \circ \vec{E}_2^* - \mu\mu_0\vec{H}_1 \circ \vec{H}_2*) \tag{C.8}$$

from which in turn after exchange of indices 1 and 2 and addition to (C.8) follows

$$\nabla \circ (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2^* \times \vec{H}_1) = 0 \quad . \tag{C.9}$$

which is the basis for mode exitation at discontinuities.

To handle coupling of modes at perturbations we assume that there exists a perturbation polarization $\vec{P}$ giving $\vec{D} = \epsilon_0 \epsilon \vec{E} + \vec{P}$. The polarization is assumed to be due to a change in dielectric constants which is the most likely to occur. Even interface perturbations can be regarded as a perturbation of dielectric constants. With the perturbation (C.7) changes to

$$\nabla \times \vec{H}_2 = -i\,\varepsilon\varepsilon_0 \vec{E}_2 - i\,\omega \vec{P}_2 \quad . \tag{C.10}$$

Going the same way that leads to (C.9) we then find

$$\nabla \circ (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2^* \times \vec{H}_1) = i\,\omega(\vec{P}_1 \times \vec{E}_2^* - \vec{P}_2^* \times \vec{E}_1) \quad . \tag{C.11}$$

Taking $\vec{E}_2$ and $\vec{H}_2$ as fields of the non-perturbed waveguide we have $\vec{P}_2 = 0$ and find

$$\nabla \circ (\vec{E}_1 \times \vec{H}_2^* - \vec{E}_2^* \times \vec{H}_1) = i\,\omega(\vec{P}_1 \circ \vec{E}_2^*) \quad . \tag{C.12}$$

C.2 Orthogonality, Normalization

Taking (C.9) for two different modes

$$\vec{E}_1 = \vec{E}_m \cdot \exp\{i(k_{zm}\, z - \omega t)\} \tag{C.13}$$

$$\vec{E}_2 = \vec{E}_p \cdot \exp\{i(k_{zp}\, z - \omega t)\} \tag{C.14}$$

we find

$$\nabla_\mathrm{t} \circ (\vec{E}_m^* \times \vec{H}_p - \vec{E}_p \times \vec{H}_m^*) + i(k_{\mathrm{z}p} - k_{\mathrm{z}m})(\vec{E}_{\mathrm{t}p} \times \vec{H}_{\mathrm{t}m}^* - \vec{E}_{\mathrm{t}m}^* \times \vec{H}_{\mathrm{t}p}) \circ \vec{e}_\mathrm{z} = 0 \tag{C.15}$$

Integrating (C.9) over the whole transverse area S_t and using for the first term

$$\iint\limits_{S_\mathrm{t}} \nabla_\mathrm{t} \circ \vec{A}\ \mathrm{d}^2 r = \oint_{C_{S_\mathrm{t}}} \vec{A} \circ\ \mathrm{d}\vec{\ell} \tag{C.16}$$

we find for transversely vanishing modes (this is true for guided modes with real propagation constants) that the first term is zero. For non-degenerate modes $k_{\mathrm{z}p} \neq k_{\mathrm{z}m}$ we have to meet then

$$\iint\limits_{S_\mathrm{t}} (\vec{E}_{\mathrm{t}p} \times \vec{H}_{\mathrm{t}m}^* - \vec{E}_{\mathrm{t}m}^* \times \vec{H}_{\mathrm{t}p}) \circ \vec{e}_\mathrm{z}\ \mathrm{d}^2 r = 0 \tag{C.17}$$

In the case $m' = -m$ follows $k_{\mathrm{z},m'} = -k_{\mathrm{z}m}$ which is the same as a transition to reverse $z-$direction. Using (C.5) follows

$$-\iint\limits_{S_\mathrm{t}} \left(\vec{E}_{\mathrm{t}p} \times \vec{H}_{\mathrm{t}m}^* + \vec{E}_{\mathrm{t}m}^* \times \vec{H}_{\mathrm{t}p}\right) \circ \vec{e}_\mathrm{z}\ \mathrm{d}^2 r = 0 \tag{C.18}$$

and comparison with (C.17) gives

$$\iint\limits_{S_\mathrm{t}} (\vec{E}_{\mathrm{t}p} \times \vec{H}_{\mathrm{t}m}^*)\ \mathrm{d}^2 r = 0 \tag{C.19}$$

which means that electric and magnetic fields of two different non-degenerate ($k_{\mathrm{z}m} \neq k_{\mathrm{z}p}$) modes are orthogonal.

For normalization we take (C.19) with $m = p$ which is similar to the z-component of the (over transverse area) integrated complex pointing vector of a mode with order m

$$\langle S_{zm} \rangle = \iint\limits_{S_t} (\vec{E}_{tm} \times \vec{H}^*_{tm}) \circ \vec{e}_z \, \mathrm{d}^2 r \quad . \tag{C.20}$$

Normalized mode fields are obtained with $\vec{\bar{E}}_m = \frac{\vec{E}_m}{\sqrt{\langle S_{z,m} \rangle}}$ and $\vec{\bar{H}}_m = \frac{\vec{H}_m}{\sqrt{\langle S_{z,m} \rangle}}$. Regarding (C.5) (C.20) can be simplified to

$$\langle S_{zm} \rangle = \iint\limits_{S_t} (\vec{E}_{tm} \times \vec{H}_{tm}) \circ \vec{e}_z \, \mathrm{d}^2 r \quad . \tag{C.21}$$

such that the orthogonal relation

$$\iint\limits_{S_t} (\vec{\bar{E}}_{tm} \times \vec{\bar{H}}^*_{tp}) \circ \vec{e}_z \, \mathrm{d}^2 r = \delta_{m,p} \quad . \tag{C.22}$$

follows. The quided modes are normally orthogonal which has to be proven for the structure under consideration.

Even when the guided modes are orthogonal they do not give a complete set of orthogonal functions preventing expansion into modes of arbitrary fields. In most practical cases transversal fields have to be expanded and then it is sufficient to deal only with guided modes because they are a full set for such fields.

For completion to a full set we have to deal with radiating modes. It can be stated that the radiating modes have a continuous $\vec{k}$ spectrum. Normalization can be done as for the guided modes when we observe the fact that for the fields of a guided mode the transverse propagation vector $\vec{k}_{t,m} = (\vec{e}_z \times \vec{k}_m) \times \vec{e}_z$ is equal for both fields $\vec{E}_m$ and $\vec{H}_m$. This leads to the more general formulation

$$\langle S\{\vec{k}_t\} \rangle = \oiint\limits_{S_V} (\vec{E}_t\{\vec{k}_t\} \times \vec{H}^*_t\{\vec{k}_t\}) \circ \mathrm{d}^2 \vec{S} \tag{C.23}$$

which holds as well for the guided modes. Note that the whole surface of a given volume has to be taken here. The normalized transverse fields are

$$\begin{aligned}\vec{\bar{E}}_{\mathrm{t}}\{\vec{k}_{\mathrm{t}}\} &= \frac{\vec{E}_{\mathrm{t}}\{\vec{k}_{\mathrm{t}}\}}{\sqrt{\langle S\{\vec{k}_{\mathrm{t}}\}\rangle}} \\ \vec{\bar{H}}_{\mathrm{t}}\{\vec{k}_{\mathrm{t}}\} &= \frac{\vec{H}_{\mathrm{t}}\{\vec{k}_{\mathrm{t}}\}}{\sqrt{\langle S\{\vec{k}_{\mathrm{t}}\}\rangle}} \quad .\end{aligned} \tag{C.24}$$

The modified representation of the orthogonal relation (C.22)

$$\oiint_{S_{\mathrm{V}}} (\vec{E}_{\mathrm{t}}\{\vec{k}_{\mathrm{t}}\} \times \vec{H}^{*}_{\mathrm{t}}\{\vec{k}'_{\mathrm{t}}\}) \circ \mathrm{d}^2\vec{S} = \delta^{(2)}\{\vec{k}_{\mathrm{t}} - \vec{k}'_{\mathrm{t}}\}$$

must be satisfied by all modes, guided and radiative ones, on the volume V otherwise they are not orthogonal and field expansions are critical with such a set of modes.

Appendix D

Kramers Kronig Relation

The dielectric constant ε can be written as a complex number where the imaginary part accounts for convection current and the real part for displacement current. In non magnetic dielectric constant as $\eta = \sqrt{\varepsilon}$ the refractive index as well is a complex number. Both the dielectric constant ε as well the refractive index can be taken as analytical function of frequency f respectively angular frequency ω or wavelength λ. Usually the refractive index is written as $\eta = n + i\kappa$ where κ and the power absorption coefficient are related by $\kappa = \alpha\lambda/(4\pi) = \alpha/2 \cdot c_0/\omega$. The Kramers Kronig relation can be deduced via Cauchy's integral relation

$$\oint \frac{g\{z\}}{z-a} dz = i2\pi g\{a\} \cdot w\{a\} \tag{D.1}$$

for any arbitrary regular complex function that diminishes at infinity

$$\lim_{|z-a|\to\infty} g\{z\} = 0. \tag{D.2}$$

In case that for infinite arguments $g\{z\}$ is a nonvanishing constant

$$\lim_{|z-a|\to\infty} g\{z\} = k_{\mathrm{g}}$$

a new function $h\{z\} = g\{z\} - k_{\mathrm{g}}$ is defined that satisfies (D.2). The integration over a constant $g\{z\} = k_{\mathrm{g}}$ in (D.1) results to be zero.

The integral trace winds $w\{a\}$ times around a. This means that w is a number that is zero in case that a is not included in the trace. The integral trace used here is sketched in figure D.1. It does not contain the point a leading to $w\{a\} = 0$.

Figure D.1: Integral trace for the Cauchy integral in (D.1).

oe2fap41

For further investigation the trace is splitted into three parts. One part is the half circle with infinite radius where the integral gives zero. The second part is the real axis (x) without the surrounding of a and the third part is the half circle aroung a with radius $|z - a| = r_\varepsilon$. The integral (D.1) can be rewritten to

$$-\int_{-r_\varepsilon}^{r_\varepsilon} \frac{g\{z\}}{z-a} dz = \int_{-\infty}^{-r_\varepsilon} \frac{g\{x\}}{x-a}\,\mathrm{d}x + \int_{r_\varepsilon}^{\infty} \frac{g\{x\}}{x-a}\,\mathrm{d}x$$

and for diminishing radius r_ε the left hand side is

$$\lim_{r_\varepsilon \to 0} \int_{-r_\varepsilon}^{r_\varepsilon} \frac{g\{z\}}{z-a} dz = -i\pi g\{a\}$$

because the circle winds mathematically negativ half around a. The right hand side is the so called main value of the integral at a that is defined with $b < a < c$

$$P\int_b^c f\{z\}\,\mathrm{d}z = \lim_{r_\varepsilon \to 0} \int_b^{a-r_\varepsilon} f\{z\}\,\mathrm{d}z + \int_{a+r_\varepsilon}^c f\{z\}\,\mathrm{d}z$$

resulting into

$$i \cdot \pi \cdot (g\{a\} - k_\mathrm{g}) = P\int_{-\infty}^{\infty} \frac{g\{x\}}{x-a}\,\mathrm{d}x. \tag{D.3}$$

Remembering that g is a complex function $g = \mathrm{Re}\{g\} + i\mathrm{Im}\{g\}$ (D.3) can be splitted into two integral equations of real numbers

$$\begin{aligned}
\pi\left(\mathrm{Re}\{g\{a\}\} - \mathrm{Re}\{k_\mathrm{g}\}\right) &= P\int_{-\infty}^{\infty} \frac{\mathrm{Im}\{g\{x\}\}}{x-a}\,\mathrm{d}x \\
-\pi\left(\mathrm{Im}\{g\{a\}\} - \mathrm{Im}\{k_\mathrm{g}\}\right) &= P\int_{-\infty}^{\infty} \frac{\mathrm{Re}\{g\{x\}\}}{x-a}\,\mathrm{d}x.
\end{aligned} \tag{D.4}$$

When the dielectric constant or the refractive index are regarded x is replaced by ω. Both ε and η exhibit a nonvanishing real part for infinite frequencies namely $\lim_{\omega\to\infty} \mathrm{Re}\{\varepsilon\} = 1$ respectively $\lim_{\omega\to\infty} \mathrm{Re}\{\eta\} = 1$. Therefore in (D.4) $\mathrm{Re}\{k_\mathrm{g}\} = 1$ and $\mathrm{Im}\{k_\mathrm{g}\} = 0$ have to be taken.

The dielectric constant as well as the refractive index follow

$$g\{z\} = g^*\{-z\}$$

leading to

$$\begin{aligned}
\mathrm{Re}\{g\{a\}\} - \mathrm{Re}\{k_\mathrm{g}\} &= \frac{2}{\pi} P\int_0^{\infty} \frac{x \cdot \mathrm{Im}\{g\{x\}\}}{x^2 - a^2}\,\mathrm{d}x \\
\mathrm{Im}\{g\{a\}\} - \mathrm{Im}\{k_\mathrm{g}\} &= -\frac{2a}{\pi} P\int_0^{\infty} \frac{\mathrm{Re}\{g\{x\}\}}{x^2 - a^2}\,\mathrm{d}x.
\end{aligned} \tag{D.5}$$

For the refractive index it is comfortable to switch from angular frequency to wavelength $\omega = \frac{c_0}{\lambda}$ and replace the extinction coefficient κ by the absorption coefficient

$$\alpha = \kappa \cdot \frac{4\pi}{\lambda}$$

resulting into

$$\begin{aligned} n\{\lambda_1\} - 1 &= \frac{\lambda_1^2}{2\pi^2} P \int_0^\infty \frac{\alpha\{\lambda\}}{\lambda_1^2 - \lambda^2} \, \mathrm{d}\lambda \\ \alpha\{\lambda_1\} &= -8P \int_0^\infty \frac{n\{\lambda\}}{\lambda_1^2 - \lambda^2} \, \mathrm{d}\lambda \quad . \end{aligned} \tag{D.6}$$

For the calculation of n or α the spectral behaviour of the corresponding α or n must be known for the whole spectrum from zero to infinity which is normally not true. In several cases n or α are changed by some effects and only the corresponding variation of the other part is required. Due to the linearity of (D.6) a variation $\Delta\alpha$ or Δn is related to

$$\begin{aligned} \Delta n\lambda_1 &= \frac{\lambda_1^2}{2\pi} P \int_0^\infty \frac{\Delta\alpha\{\lambda\}}{\lambda_1^2 - \lambda^2} \, \mathrm{d}\lambda \\ \Delta\alpha\lambda_1 &= -8P \int_0^\infty \frac{\Delta n\{\lambda\}}{\lambda_1^2 - \lambda^2} \, \mathrm{d}\lambda. \end{aligned} \tag{D.7}$$

It is worth to note that normally the variations $\Delta\alpha$ and Δn are limited to a small spectral region $\Delta\lambda$ and are well known from measurements such that the integration can be carried out very easy.

Appendix E

The Wiener-Kintchine Theorem

Equation (7.98)

$$|\widetilde{a}\{f\}|^2 = \int_{-\infty}^{\infty} |a\{t\}|^2 \exp\{i2\pi ft\}\mathrm{d}t \simeq \int_{-\infty}^{\infty} \langle a\{t+T\}a^*\{t\}\rangle \exp\{i2\pi fT\}\mathrm{d}T \quad \text{(E.1)}$$

can be shown with the definition of a Fourier transformation for time signals $b\{t\}$ of length T

$$\mathcal{F}_{\mathrm{T}}\{b\{t\}\} = \widetilde{b}_{\mathrm{T}}\{f\} = \frac{1}{T}\int_0^T b\{t\}\exp\{i2\pi ft\}\mathrm{d}t \quad \text{(E.2)}$$

in contrast to the usual Fourier transformation

$$\mathcal{F}\{b\{t\}\} = \widetilde{b}\{f\} = \int_{-\infty}^{\infty} b\{t\}\exp\{i2\pi ft\}\mathrm{d}t \quad . \quad \text{(E.3)}$$

For sufficient long times T the Fourier transforms are related by

$$\widetilde{b}\{f\} \simeq 2T\widetilde{b}_{\mathrm{T}}\{f\} = 2/B\widetilde{b}_{\mathrm{T}}\{f\} \quad \text{(E.4)}$$

with bandwidth $B = 1/T$. Back transformation

$$b\{t\} = \mathcal{F}_{\mathrm{T}}^{-1}\{\widetilde{b}_{\mathrm{T}}\{f\}\} = \frac{1}{B}\int_{-\infty}^{\infty} \widetilde{b}_{\mathrm{T}}\{f\}\exp\{-i2\pi ft\}\mathrm{d}t \tag{E.5}$$

$$= \mathcal{F}^{-1}\{\widetilde{b}\{f\}\} = \int_{-\infty}^{\infty} \widetilde{b}\{f\}\exp\{-i2\pi ft\}\mathrm{d}t \tag{E.6}$$

looks similar in both cases. We substitute $b\{t\} = \langle a\{t+T\}a^*\{t\}\rangle$ and use $\mathcal{F}_{\mathrm{T}}\{b^*\} = \mathcal{F}_{\mathrm{T}}^*\{b\}$. With Fourier back transformation the autocorrelation function can be rewritten as

$$\begin{aligned}\langle a\{t+T\}a^*\{t\}\rangle &= \frac{1}{T}\int_0^T a^*\{t\}a\{t+T\}\mathrm{d}t \\ &= \frac{1}{BT}\int_{-\infty}^{\infty}\widetilde{a}_{\mathrm{T}}^*\int_0^T a\{t+T\}\exp\{i2\pi ft\}\mathrm{d}t\mathrm{d}f \quad .\end{aligned} \tag{E.7}$$

A time translation is transformed into a phase translation

$$\mathcal{F}_{\mathrm{T}}\{a\{t+T\}\} = \mathcal{F}_{\mathrm{T}}\{a\{t\}\}\exp\{-i2\pi fT\} \tag{E.8}$$

and therefore

$$\begin{aligned}\langle a\{t+T\}a^*\{t\}\rangle &= \frac{1}{B}\int_{-\infty}^{\infty}\langle\widetilde{a}_{\mathrm{T}}\widetilde{a}_{\mathrm{T}}^*\rangle\exp\{-i2\pi fT\}\mathrm{d}f \\ &= \frac{1}{B}\int_{-\infty}^{\infty}\langle|\widetilde{a}_{\mathrm{T}}|^2\rangle\exp\{-i2\pi fT\}\mathrm{d}f \quad .\end{aligned} \tag{E.9}$$

With (E.9) and Fourier transformation the power density spectrum results to

$$\begin{aligned}\langle|\widetilde{a}_{\mathrm{T}}|^2\rangle &= B\int_0^\infty \langle a^*\{t\}a\{t+T\}\rangle \exp\{i2\pi fT\}\mathrm{d}T \\ &= \frac{B}{2}\int_{-\infty}^\infty \langle a^*\{t\}a\{t+T\}\rangle \exp\{i2\pi fT\}\mathrm{d}T \simeq \frac{B}{2}|\widetilde{a}\{f\}|^2 \end{aligned} \tag{E.10}$$

q.e.d.

Appendix F

Density of States in Quantum Structures

The density of states is essential for gain calculations in quantum structures. In k-space it is given by $D_\mathrm{k}\mathrm{d}^3k$. In the gain calculation the density $D\mathrm{d}W$ is required. For the materials considered here a quadratic energy - wavenumber relation

$$W_\mathrm{B} = W \pm \frac{\hbar^2}{2m_\mathrm{B}}\|\vec{k}_\mathrm{W}\|^2 = W \pm \frac{\hbar^2}{2m_\mathrm{B}}\sqrt{k_\mathrm{x}^2 + k_\mathrm{y}^2 + k_\mathrm{z}^2} \tag{F.1}$$

is assumed. W_B denotes the band energy level of valence or conduction band at the Γ point W_V or W_C and m_B is the corresponding carrier mass m_lh, m_hh or m_e respectively. In the valence band the '+' sign has to be used and in the valence band the '−' sign is valid.

The density $D\,\mathrm{d}W = D\{W\}\,\mathrm{d}W$ denotes the density of states per unit energy at a given constant energy W. It is found by integration of D_k over all values $\vec{k} = \vec{k}_\mathrm{W}$ that correspond to the energy W

$$D_\mathrm{B}\,\mathrm{d}W = \iint\limits_{k_\mathrm{W}} D_\mathrm{k}\mathrm{d}^2k\mathrm{d}k \tag{F.2}$$

and with

$$\mathrm{d}W = (\nabla_{\mathrm{k}} W) \circ \mathrm{d}\vec{k} = \pm(-2)\frac{\hbar^2}{2m_{\mathrm{B}}}(\vec{k} \circ \mathrm{d}\vec{k}) = \pm(-2)\frac{\hbar^2}{2m_{\mathrm{B}}}\|\vec{k}\|\mathrm{d}k$$

from (F.1) $\mathrm{d}k$ can be replaced giving

$$\begin{aligned} D_{\mathrm{B}}\,\mathrm{d}W &= \pm\left(-\frac{1}{2}\right)\frac{2m_{\mathrm{B}}}{\hbar^2}\iint\limits_{k_{\mathrm{W}}}\frac{1}{\|\vec{k}\|}D_{\mathrm{k}}\,\mathrm{d}^2k\,\mathrm{d}W \\ &= \pm\left(-\frac{1}{2}\right)\sqrt{\frac{2m_{\mathrm{B}}}{\hbar^2}}\iint\limits_{k_{\mathrm{W}}}\frac{D_{\mathrm{k}}}{\sqrt{\pm(W_{\mathrm{B}}-W)}}\,\mathrm{d}^2k\,\mathrm{d}W \quad . \end{aligned} \tag{F.3}$$

For constant energy the wavenumber is

$$\|\vec{k}_{\mathrm{W}}\|^2 = \pm\frac{2m_{\mathrm{B}}}{\hbar^2}(W_{\mathrm{B}}-W) = k_{\mathrm{x}}^2 + k_{\mathrm{y}}^2 + k_{\mathrm{z}}^2 \quad .$$

All values of $\vec{k}_{\mathrm{W}}$ are ending on the surface of a sphere with constant radius $\|\vec{k}_{\mathrm{W}}\|$. Therefore it can be convenient to write the integration in spherical coordinates $\mathrm{d}^2k \rightarrow \|\vec{k}_{\mathrm{W}}\|^2\sin\{\theta\}\,\mathrm{d}\theta\,\mathrm{d}\phi$ resulting into

$$D_{\mathrm{B}}\,\mathrm{d}W = \pm\left(-\frac{1}{2}\right)\frac{2m_{\mathrm{B}}}{\hbar^2}\|\vec{k}_{\mathrm{W}}\|\int_0^{2\pi}\int_0^{\pi} D_{\mathrm{k}}\sin\{\theta\}\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}W \quad .$$

The k- space density of states in bulk material is

$$D_{\mathrm{k}}\mathrm{d}^3k = \frac{2}{(2\pi)^3}\mathrm{d}^3k \quad .$$

In rectangular quantum structures with extensions $a_{\mathrm{x}}, a_{\mathrm{y}}$ and a_{z} in x, y and z- direction respectively it is modified due to the discrete solutions $k_{\mathrm{x}} = k_{\mathrm{m}}, k_{\mathrm{y}} = k_{\mathrm{n}}$ and $k_{\mathrm{z}} = k_{\ell}$. In the case of infinite barrier height $k_{\mathrm{m}} = m \cdot 2\pi/a_{\mathrm{x}}$, $k_{\mathrm{n}} = n \cdot 2\pi/a_{\mathrm{y}}$ and $k_{\ell} = \ell \cdot 2\pi/a_{\mathrm{z}}$

result and the densities are given

in a quantum well

$$D_{\mathrm{k}}\mathrm{d}^3k = \sum_{m=1}^{m_{\max}} \frac{2}{(2\pi)^3}\frac{2\pi}{a_{\mathrm{x}}}\delta\{k_{\mathrm{x}} - k_{\mathrm{m}}\}\,\mathrm{d}^3k$$

in a quantum wire

$$D_{\mathrm{k}}\mathrm{d}^3k = \sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}} \frac{2}{(2\pi)^3}\frac{(2\pi)^2}{a_{\mathrm{x}}a_{\mathrm{y}}}\delta\{k_{\mathrm{x}} - k_{\mathrm{m}}\}\delta\{k_{\mathrm{y}} - k_{\mathrm{n}}\}\,\mathrm{d}^3k$$

and in a quantum box, also called quantum dot,

$$\begin{aligned} D_{\mathrm{k}}\mathrm{d}^3k &= \sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\sum_{\ell=1}^{\ell_{\max}} \frac{2}{(2\pi)^3}\frac{(2\pi)^3}{a_{\mathrm{x}}a_{\mathrm{y}}a_{\mathrm{z}}}\delta\{k_{\mathrm{x}} - k_{\mathrm{m}}\}\delta\{k_{\mathrm{y}} - k_{\mathrm{n}}\}\delta\{k_{\mathrm{z}} - k_{\ell}\}\,\mathrm{d}^3k \\ &= \sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\sum_{\ell=1}^{\ell_{\max}} \frac{2}{a_{\mathrm{x}}a_{\mathrm{y}}a_{\mathrm{z}}}\delta^3\{\vec{k} - \vec{k}_{m,n,\ell}\}\,\mathrm{d}^3k \quad . \end{aligned}$$

The parameters $m_{\max}$, $n_{\max} = n_{\max}\{m\}$, and $\ell_{\max} = \ell_{\max}\{m,n\}$ are given by

$$\begin{aligned} m_{\max}\frac{2\pi}{a_{\mathrm{x}}} &\leq \quad \pm\tfrac{2m_{\mathrm{B}}}{\hbar^2}(W_{\mathrm{B}} - W) \quad &&\leq (m_{\max}+1)\frac{2\pi}{a_{\mathrm{x}}} \\ n_{\max}\frac{2\pi}{a_{\mathrm{y}}} &\leq \pm\tfrac{2m_{\mathrm{B}}}{\hbar^2}(W_{\mathrm{B}} - W_{\mathrm{B}m}) &&\leq (n_{\max}+1)\frac{2\pi}{a_{\mathrm{y}}} \\ \ell_{\max}\frac{2\pi}{a_{\mathrm{z}}} &\leq \pm\tfrac{2m_{\mathrm{B}}}{\hbar^2}(W_{\mathrm{B}} - W_{\mathrm{B}m,n}) &&\leq (\ell_{\max}+1)\frac{2\pi}{a_{\mathrm{z}}} \end{aligned}$$

when

$$\begin{aligned} W_{\mathrm{B}m} &= W_{\mathrm{B}} - \left(\pm\frac{\hbar^2}{2m_{\mathrm{B}}}\left(k_{\mathrm{m}}^2 + k_{\mathrm{y}}^2 + k_{\mathrm{z}}^2\right)\right) \\ W_{\mathrm{B}m,n} &= W_{\mathrm{B}} - \left(\pm\frac{\hbar^2}{2m_{\mathrm{B}}}\left(k_{\mathrm{m}}^2 + k_{\mathrm{n}}^2 + k_{\mathrm{z}}^2\right)\right) \\ W_{\mathrm{B}m,n,\ell} &= W_{\mathrm{B}} - \left(\pm\frac{\hbar^2}{2m_{\mathrm{B}}}\left(k_{\mathrm{m}}^2 + k_{\mathrm{n}}^2 + k_{\ell}^2\right)\right) \end{aligned}$$

are used. The density of states energy spectrum follows for bulk material

$$\begin{aligned} D_{\mathrm{B}}\,\mathrm{d}W &= \pm\left(-\frac{1}{2}\right)\left(\frac{2m_{\mathrm{B}}}{\hbar^2}\right)\|\vec{k}_{\mathrm{W}}\|\int_0^{2\pi}\int_0^{\pi}\frac{2}{(2\pi)^3}\sin\{\theta\}\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}W \\ &= \pm\left(-\frac{1}{2}\right)\left(\frac{2m_{\mathrm{B}}}{\hbar^2}\right)\frac{2}{(2\pi)^3}4\pi\|\vec{k}_{\mathrm{W}}\|\,\mathrm{d}W \\ &= \pm\left(-\frac{4\pi}{(2\pi)^3}\right)\left(\frac{2m_{\mathrm{B}}}{\hbar^2}\right)^{3/2}\sqrt{\pm(W_{\mathrm{B}}-W)}\,\mathrm{d}W \\ &= \pm 2\frac{-2}{\sqrt{\pi}}\left(\frac{2\pi m_{\mathrm{B}}}{h^2}k_{\mathrm{B}}T\right)^{3/2}\sqrt{\pm\frac{W_{\mathrm{B}}-W}{k_{\mathrm{B}}T}}\,\mathrm{d}\left(\frac{W}{k_{\mathrm{B}}T}\right) \end{aligned}$$

and with the so called effective density of states

$$N_{\mathrm{B}} = 2\left(\frac{2\pi m_{\mathrm{B}}}{h^2}k_{\mathrm{B}}T\right)^{3/2}$$

the well known expression

$$D_{\mathrm{B}}\mathrm{d}W = \pm\frac{-2}{\sqrt{\pi}}N_{\mathrm{B}}\sqrt{\pm\frac{W_{\mathrm{B}}-W}{k_{\mathrm{B}}T}}\mathrm{d}\left(\frac{W}{k_{\mathrm{B}}T}\right) \tag{F.4}$$

results. The minus sign in (F.4) is somewhat confusing in comparison to the literature where typically only the absolute value is given but it tells just that the density of states in the valence band decreases with increasing energy and increases with increasing energy in the conduction band.

In a quantum well of thickness a_{x}

$$D_{\mathrm{B}}\mathrm{d}W = \pm\left(-\frac{1}{2}\right)\frac{2m_{\mathrm{B}}}{\hbar^2}\|\vec{k}_{\mathrm{W}}\|\int_0^{2\pi}\int_0^{\pi}\sum_{m=1}^{m_{\max}}\frac{2}{(2\pi)^3}\frac{2\pi}{a_{\mathrm{x}}}\delta\{k_{\mathrm{x}}-k_{\mathrm{m}}\}\sin\{\theta\}\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}W$$

has to be calculated. Replacing $k_{\mathrm{x}} = \|\vec{k}_{\mathrm{W}}\|\cos\{\theta\}$ and $k_{\mathrm{m}} = \|\vec{k}_{\mathrm{W}}\|\cos\{\theta_{\mathrm{m}}\}$ the δ-function can be substituted with the use of

$$\delta\left\{f\left\{x\right\}-f\left\{x_0\right\}\right\}=\frac{1}{|\frac{\partial}{\partial x}f|}\delta\{x-x_0\}$$

by

$$\delta\{k_{\mathrm{x}}-k_{\mathrm{m}}\}\rightarrow\frac{\delta\{\theta-\theta_{\mathrm{m}}\}}{\|\vec{k}_{\mathrm{W}}\|\sin\{\theta\}}$$

giving

$$D_{\mathrm{B}}\,\mathrm{d}W=\pm\left(-\frac{1}{2}\right)\frac{2m_{\mathrm{B}}}{\hbar^2}\frac{2}{(2\pi)^3}\frac{2\pi}{a_{\mathrm{x}}}\sum_{m=1}^{m_{\max}}\int_0^{2\pi}\int_0^{\pi}\delta\{\theta-\theta_{\mathrm{m}}\}\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}W\quad.$$

The integration $\int_{x_0-\epsilon}^{x_0+\epsilon}\delta\{x-x_0\}\mathrm{d}x=\mathrm{H}\{x-x_0\}$ results into the Heaviside function

$$\mathrm{H}\{x\}=\begin{cases}0\text{ when }x<0\\1\text{ when }x\geq 0\end{cases}\quad.$$

The density of states energy spectrum therefore can be written as

$$\begin{aligned}D_{\mathrm{B}}\,\mathrm{d}W&=\pm\left(-\frac{1}{2}\right)\frac{2m_{\mathrm{B}}}{\hbar^2}\frac{2}{(2\pi)^3}\frac{(2\pi)^2}{a_{\mathrm{x}}}\sum_{m=1}^{m_{\max}}\mathrm{H}\{\theta-\theta_{\mathrm{m}}\}\,\mathrm{d}W\\&=\pm\frac{-2}{a_{\mathrm{x}}}\frac{1}{4\pi}\frac{2m_{\mathrm{B}}}{\hbar^2}\sum_{m=1}^{m_{\max}}\mathrm{H}\{k_{\mathrm{x}}-k_{\mathrm{m}}\}\,\mathrm{d}W=\pm\sum_{m=1}^{m_{\max}}(-2D_{\mathrm{B}m})\,\mathrm{H}\{W_{\mathrm{B}}-W_{\mathrm{B}m}\}\,\mathrm{d}W\end{aligned}$$

where

$$D_{\mathrm{B}m}=\frac{2}{a_{\mathrm{x}}}\frac{1}{4\pi}\frac{2m_{\mathrm{B}}}{\hbar^2}$$

gives the carrier density per unit energy at level m and is independent of m. Therefore the total density of occupied states results from the summation over Heaviside functions $\sum_{m=1}^{m_{\max}}\mathrm{H}\{W_{\mathrm{B}}-W_{\mathrm{B}m}\}=m_{\max}$ to be

$$D_{\mathrm{B}}\,\mathrm{d}W = \pm(-m_{\max}D_{\mathrm{B}m})\,\mathrm{d}W = \pm\left(-m_{\max}\frac{2\pi}{a_{\mathrm{x}}}\frac{2m_{\mathrm{B}}}{h^2}\right)\,\mathrm{d}W \quad .$$

For the quantum wire we use (F.3) and find

$$\begin{aligned}
D\mathrm{d}W &= \pm\left(-\frac{1}{2}\right)\frac{2m_{\mathrm{B}}}{\hbar^2}\frac{1}{\|k_{\mathrm{W}}\|}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\int_{k_{\mathrm{x}}}\int_{k_{\mathrm{y}}}\frac{(2\pi)^2}{a_{\mathrm{x}}a_{\mathrm{y}}}\frac{2}{(2\pi)^3}\delta\{k_{\mathrm{x}}-k_m\}\delta\{k_{\mathrm{y}}-k_n\}\,\mathrm{d}k_{\mathrm{x}}\,\mathrm{d}k_{\mathrm{y}}\,\mathrm{d}W \\
&= \pm\left(-\frac{1}{2\pi}\right)\frac{2m_{\mathrm{B}}}{\hbar^2}\frac{1}{a_{\mathrm{x}}a_{\mathrm{y}}}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\frac{1}{\sqrt{k_m^2+k_n^2+k_{\mathrm{z}}^2}}\,\mathrm{H}\{k_{\mathrm{x}}-k_m\}\,\mathrm{H}\{k_{\mathrm{y}}-k_n\}\mathrm{d}W \\
&= \pm\frac{-2}{a_{\mathrm{x}}a_{\mathrm{y}}}\frac{1}{4\pi}\sqrt{\frac{2m_{\mathrm{B}}}{\hbar^2}}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\frac{1}{\sqrt{\pm(W_{\mathrm{B}}-W_{\mathrm{B}m,n})}}\,\mathrm{H}^2\{\pm(W_{\mathrm{B}}-W_{\mathrm{B}m,n})\}\mathrm{d}W \\
&= \pm\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}(-D_{\mathrm{B}m,n})\,\mathrm{H}^2\{\pm(W_{\mathrm{B}}-W_{\mathrm{B}m,n})\}\mathrm{d}W
\end{aligned}$$

with

$$D_{\mathrm{B}m,n} = \frac{2}{a_{\mathrm{x}}a_{\mathrm{y}}}\frac{1}{4\pi}\sqrt{\frac{2m_{\mathrm{B}}}{\hbar^2}}\frac{1}{\sqrt{\pm(W_{\mathrm{B}}-W_{\mathrm{B}m,n})}} \quad .$$

Unfortunately the density of states cannot be further simplified like for a quantum well because $D_{\mathrm{B}m,n}$ depends on m and n.

For a quantum box (quantum dot) it is convenient to start directly from (F.2).

$$\begin{aligned}
D_{\mathrm{B}}\,\mathrm{d}W &= \frac{2}{a_{\mathrm{x}}a_{\mathrm{y}}a_{\mathrm{z}}}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\sum_{\ell=1}^{\ell_{\max}}\iint_{k_{\mathrm{W}}}\delta^3\{\vec{k}-\vec{k}_{m,n,\ell}\}\,\mathrm{d}^3k \\
&= \frac{2}{a_{\mathrm{x}}a_{\mathrm{y}}a_{\mathrm{z}}}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\sum_{\ell=1}^{\ell_{\max}}\mathrm{H}^2\{\vec{k}-\vec{k}_{m,n,\ell}\}\delta\{\vec{k}-\vec{k}_{m,n,\ell}\}\,\mathrm{d}k \\
&= \frac{2}{a_{\mathrm{x}}a_{\mathrm{y}}a_{\mathrm{z}}}\sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\sum_{\ell=1}^{\ell_{\max}}\mathrm{H}^2\{\pm(W_{\mathrm{B}}-W_{\mathrm{B}m,n,\ell})\}\delta\{\pm(W_{\mathrm{B}}-W_{\mathrm{B}m,n,\ell})\}\,\mathrm{d}W \\
&= \sum_{m=1}^{m_{\max}}\sum_{n=1}^{n_{\max}}\sum_{\ell=1}^{\ell_{\max}}D_{\mathrm{B}m,n,\ell}\,\mathrm{H}^2\{\pm(W_{\mathrm{B}}-W_{\mathrm{B}m,n,\ell})\}\,\mathrm{d}W
\end{aligned}$$

with equal density of states

$$D_{\mathrm{B}m,n,\ell} = \frac{2}{a_{\mathrm{x}} a_{\mathrm{y}} a_{\mathrm{z}}} \delta\{\pm(W_{\mathrm{B}} - W_{\mathrm{B}m,n,\ell})\}$$

at each allowed energy level. The total number of states in a quantum structure follows after integration over the structure volume. When only the active volume $V_{\mathrm{small\,gap}}$ with dimensions $a_{\mathrm{x}}, a_{\mathrm{y}}, a_{\mathrm{z}}$ is regarded, no states are lost. This impression is somewhat misleading because the barrier material has to be incorporated as well for a comparison between the structures. The total volume occupied by a quantum structure is $V_{\mathrm{total}} = V_{\mathrm{small\,gap}} + V_{\mathrm{barrier}}$ with volume of barrier material V_{barrier}. For this reason a correction factor $V_{\mathrm{small\,gap}}/V_{\mathrm{total}}$ that accounts for the barrier material has to be introduced in the calculation of state density for comparison with bulk material of the same total volume V_{total}. This is the reason why quantum dot material exhibits not that high gain expected from the simple density of state calculation where the barriers are neglected.