6.097(UG) Fundamentals of Photonics
6.974 (G) Quantum Electronics

Franz X. Kaertner

Spring Term 2006

Contents

Chapter 1

Introduction

The word Photonics has been coined over the last two decades with the goal to describe the field of Optics, Optoelectronic phenomena and its applications by one word similar to the very successful field of Electronics. As the word indicates, Photonics uses primarily photons to carry out its mission in contrast to Electronics which uses primarily electrons or other charged particles. But no strict separation of both fields is possible, at least not on an elementary level. Electrons in acceleration generate electromagnetic waves. Sometimes these waves are guided and over short distances the wave aspect can be neglected and one can talk about electronics only. Similar, pure photonics, i.e. the free electromagnetic field in vaccum, without matter is not of very much interest and use either. Quantum Mechanics teaches us that there are elementary excitations (in energy) of these waves, photons with particle like properties. So far, we are used to thinking of electrons as classical particles but Quantum Mechanics equally assigns to them wave properties, matter waves, like it assigns to the electromagnetic waves particle like properties. Electrons are elementary excitations of matter waves just as photons are elementary excitations of electromagnetic waves. We will talk later in more depth about this wave particle duality, once we have developed the mathematical tools to analyze its meaning quantitatively.

So far we have been educated using the language of classical physics. In the classical limit, it turns out that electrons (particles with spin 1/2) are particles and photons (particles with integral spin) behave like classical waves and the particle nature is of vanishing importance. This is the reason why Photonics eventually is experienced as more abstract than Electronics, especially if you went through three years of education primarily focused on

Electronics and its applications. The particle properties of electromagnetic waves become of importance when the energy of the photons considered, hf, where h is Planck's constant and f is the frequency, is larger than the thermal energy stored in an electromagnetic mode, kT, where k is Boltzmann's constant and T is temperature. At room temperature this is the case for frequencies greater than 6THz. For lower temperatures this transition frequency from classical to quantum behaviour may already occur at GHz frequencies. Certainly, at room temperature, the particle properties are important in the near infrared (IR) and visible spectrum where currently the bulk of the photonic activities are carried out, see Figure 1.1.

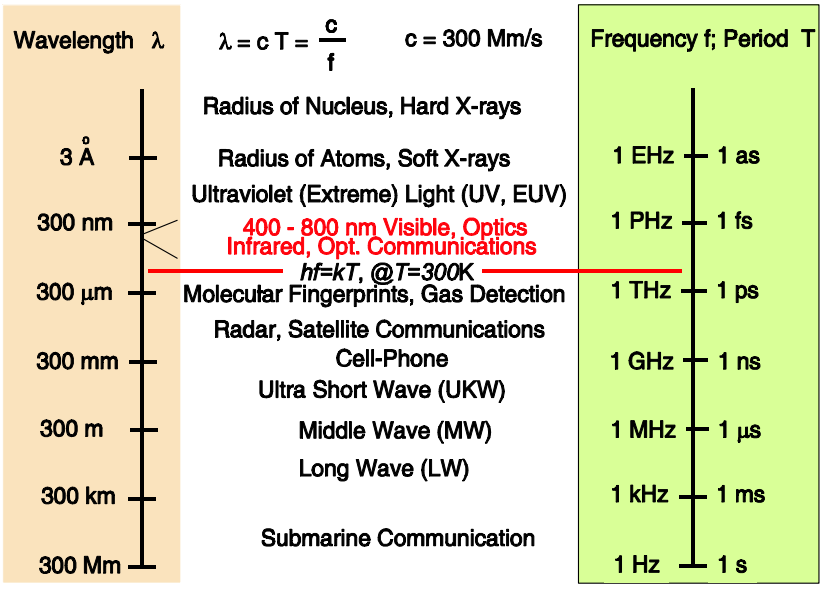

Figure 1.1: Wavelength and frequency ranges of electromagnetic radiation and its use.

An important task of Photonics is the development of coherent sources of radiation, which are in the optical range called LASER's (Light Amplification by Stimulated Emission of Radiation). The first amplifier making explicit use of the quantum properties of matter was the MASER (Microwave Amplification by Stimulated Emission of Radiation) invented by J. P. Gordon, C. H. Townes and Zeiger in 1954. The extension of the MASER principle to

the optical wavelength range was proposed by Schawlow and Townes in 1958 and the first laser was demonstrated in 1960 by Maiman [5].

As in the case of Electronics, Communications is a very important area in Photonics, and is usually called Optical Communications, which is carried out primarily in the infrared to visible wavelength range, see Figure 1.1. The enormous bandwidth available at optical frequencies, roughly four orders of magnitude higher than typical microwave frequencies (10GHz), enables the corresponding advance in information transmission capacity when compared to what electronics can do over a single long distance transmission link at microwave frequencies. Our current society could not be sustained without the worldwide deployed optical cable network, see Figure 1.2, which can be considered the largest connected machinery on earth. Understanding of the basic working principles of the components used in such lightwave networks, like the optical fiber itself, waveguide couplers, modulators, light sources is at the heart of this course.

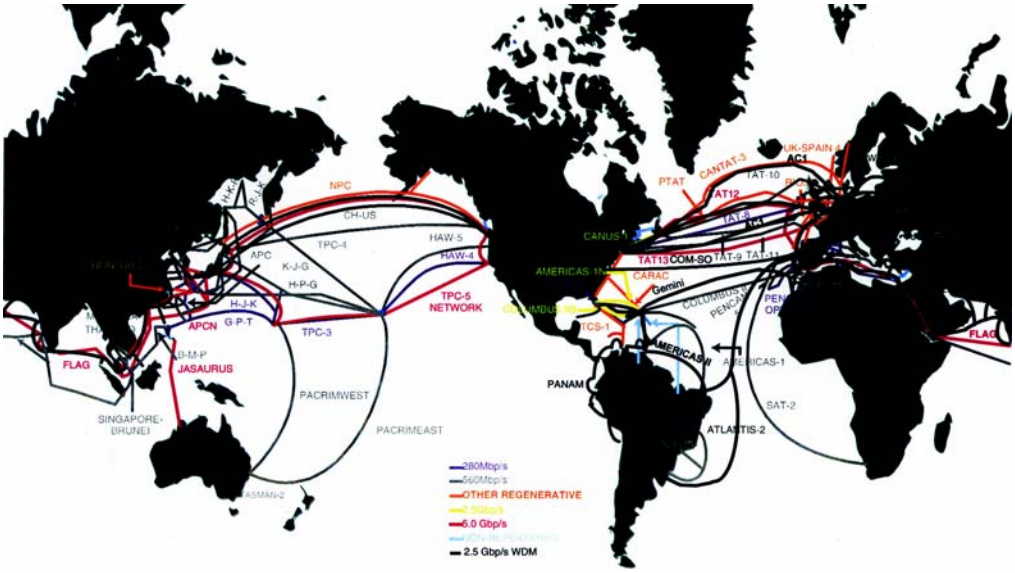

Figure 1.2: Worldwide underwater optical cable systems

The big advantage of Photonics versus Electronics in the area of information transmission is that photons do not interact with each other directly. Maxwell's equations in the classical vacuum are linear. Whereas electrons,

which are charged particles, show very strong interaction via their Coulomb fields. This is one reason why Photonics is very powerful in information transmission; essentially no interference with other signals; whereas electronics is very strong in information processing which is based on nonlinear operations.

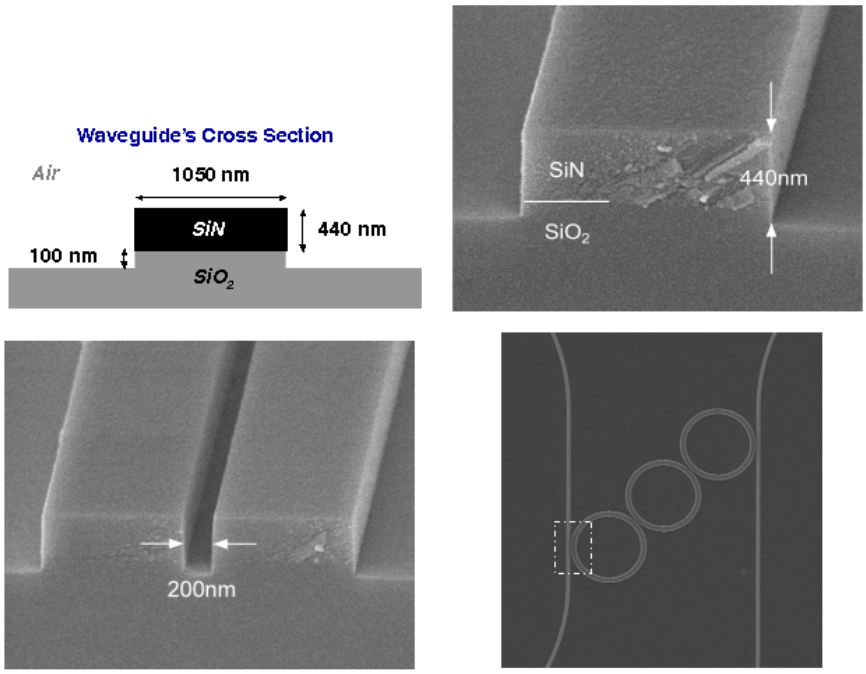

Figure 1.3: High-Index Contrast SiN/SiO_2/Air Filter using Scanning Electron Beam Lithograph. Courtesy of H. Smith MIT-Nanostructures Laboratory

As in the case of Electronics, miniaturization of many components and integration on a common technology platform is one of the major challenges in photonics (Integrated Optics). Over the last few years a new field called Silicon-Photonics came to life. Modern nanofabrication techniques such as Scanning Electron Beam Lithography enable the fabrication of optical components on the scale of the optical wavelength with a relative precision in the few nanometer range in the same material system like electronics, i.e. silicon/silicon nitride/silicon dioxide, see Figure 1.3. Fig. 1.4 shows a measured

filter characteristic useful for optical communications applications. Systems that integrate both optical and electronic subsystems on a single chip called Electronic-Photonic Integrated Circuits are on the horizon.

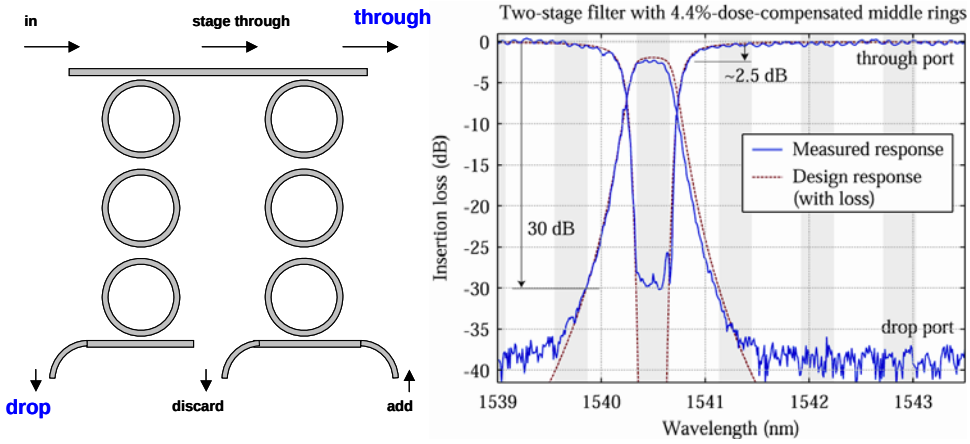

Figure 1.4: Schematic and measured two stage third order ring filter. Courtesy MIT-Optics and Quantum Electronics Group

As with Electronics, Photonics has not only left its mark on communications but is used in a wide variety of sensing applications specific to its wavelength range as well as optical imaging, test and measurement instrumentation, manufacturing and materials processing. High power lasers with average powers of tens of kilowatts can cut many centimeter thick steel plates at high speed. Laser systems reaching petawatts (10^{15}Watt) of peak power were built at Lawrence Livermore Laboratory for the first time in 1996, (see Fig. 1.5) and much more compact versions are under construction at several laboratories around the world. The enormous peak power (the average power consumption of the earth is on the order of a few Terawatt (10^{12}Watt)), even though only available over a fraction of a picosecond, will enable us to investigate new physics and fundamental interactions at extreme intensities, such as the scattering of photons with each other invoking vacuum nonlinearities (the Quantum Electrodynamic Vacuum), see Figure 1.6

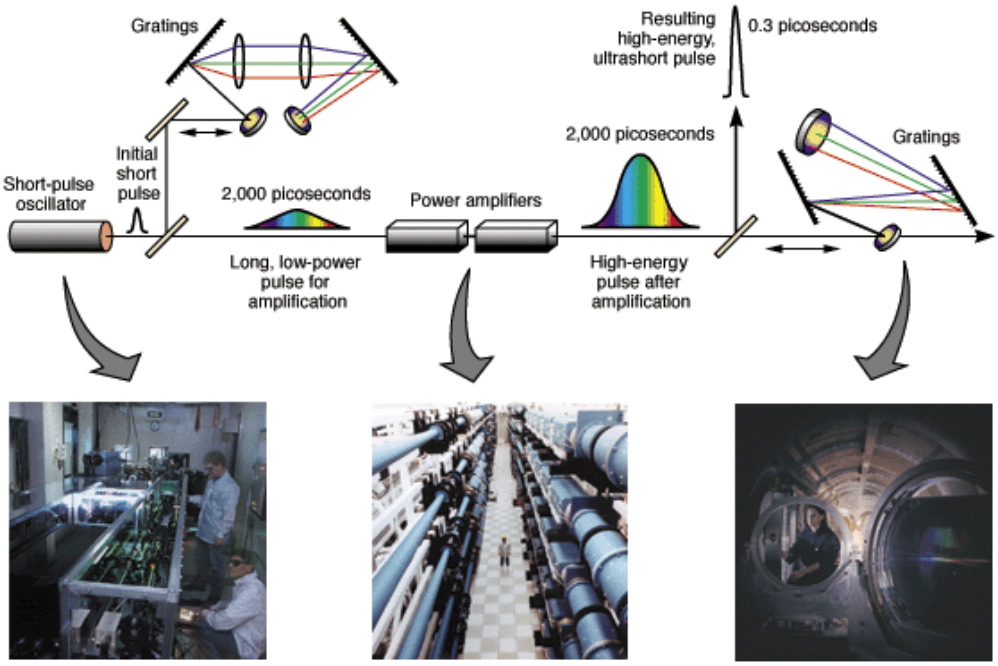

Figure 1.5: First petawatt laser system installed at LLNL using chirped pulse amplification and the NOVA amplifier chain, see http://www.llnl.gov/str/MPerry.html.

Because of the higher operating frequencies, optical sources can provide much higher time resolution than microwave sources. Therefore, it is not surprising that the shortest controlled events that currently can be made are optical pulses that comprise only a few cycles of light. These events are so short that one can only use the pulses themselves to reveal its pulse width, for example by performing an autocorrelation of the pulse with a copy of itself using a nonlinear element, see Fig. 1.7.

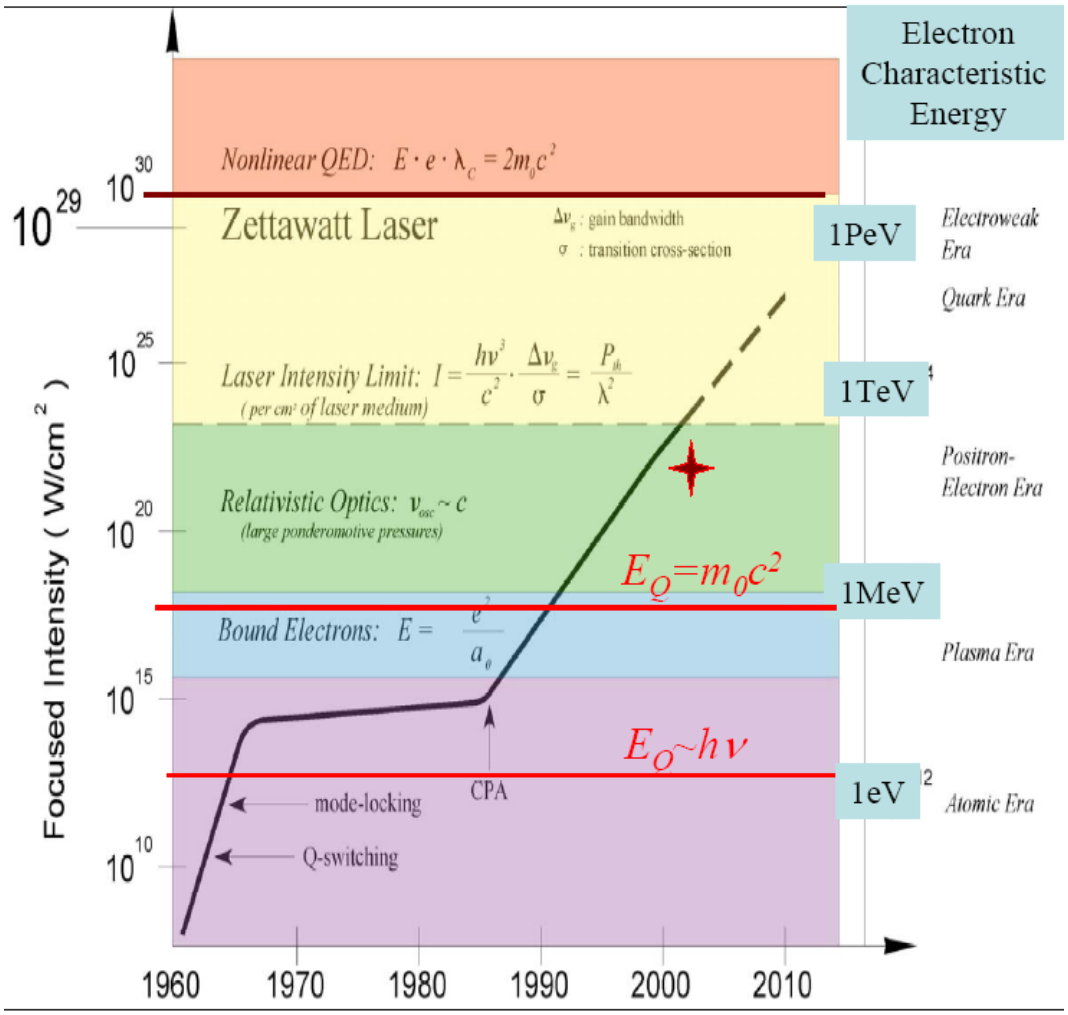

Figure 1.6: Progress in peak intensity generation from lasers over the last 45 years. Courtesy Gerard Mourou.

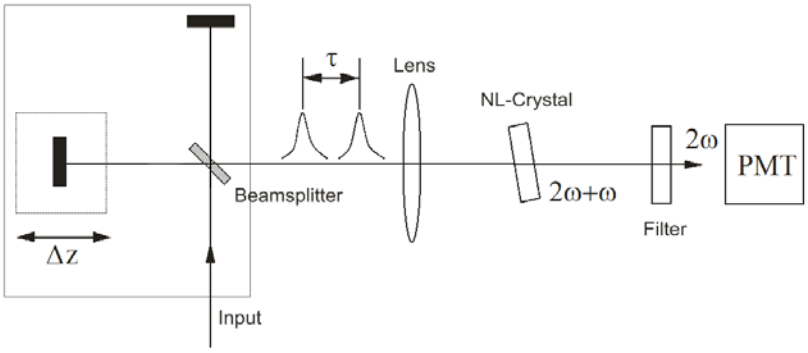

Figure 1.7: Experimental setup for measureing an interferometric autocorrelation. A copy of the pulse is generated and interferometrically delayed with respect to the original pulse. The superposition of the two pulses undergo second harmonic generation in a crystal and the second harmonic light is detected.

Figure 1.8 shows a measured interferometric second harmonic autocorrelation trace of a 5fs (fs=femtosecond, $10^{-15}s$) short pulse at a center frequency of 800nm [19].

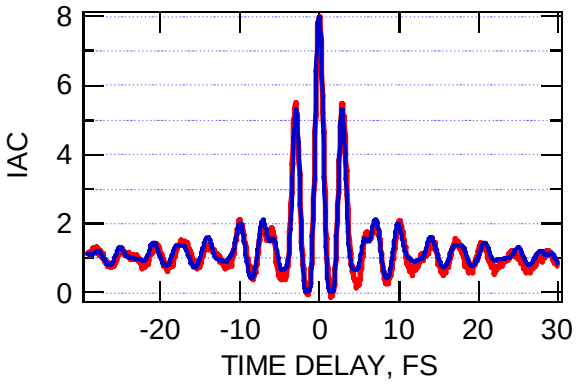

Figure 1.8: Measured and fitted interferometric autocorrelation function of a 5fs pulse [19].

Femtosecond pulses have been used to understand the carrier dynamics in semiconductors, that limit the speed of electronic devices, the dynamics

of chemical reactions such as photosynthesis or the early stages of vision. As in any discipline the development does not stop and most recently pulses as short as 250 attoseconds ($10^{-18}s$)have been generated in the soft-x-ray regime [4]. Laser like sources potentially generating radiation up into the hard x-ray regime are being planned to be built over the next years [5]. The hope is that such sources enable the spatially and temporally resolved direct imaging of molecules and atoms undergoing reactions, which would advance medical diagnostics and drug discovery. As these examples show, Photonics is a booming area, despite the downturn of the Telecommunications industry. The basis for its success is fundamental and, therefore, it will continue to grow in importance with impact in our everyday lifes as has been the case with (Micro-) Electronics, just think about your high-speed internet connection or the CD-player.

The course is organized as follows: In Chapter 2 we will review some concepts of Classical Electromagnetism from 6.013, such as plane electromagnetic waves in isotropic media, the related energy flow, and their properties upon reflections at interfaces. This will enable us to study various important optical devices such as prisms, interferometers, Fabry-Perot resonators and thin film interference coatings. The plane electromagnetic wave is an idealization similar to the perfect monochromatic sin or cosine in signal theory. We will use the plane waves to understand beams of finite size and discuss them as solutions to the paraxial wave equation, the Gaussian beams which in the limit of large beam size become optical rays. The concept of optical rays will enable us to discuss imaging properties and the design of optical systems. Then we will look into guided optical beams such as in optical fibers and waveguides from the point of view of ray optics and from the view point of full solutions to Maxwell's equations. The concept of optical modes and resonances will be discussed. To understand optics in crystalline materials and related phenomena such as birefringence and its multiple use in optical components for polarization manipulation and switching we will briefly discuss wave propagation in anisotropic media.

This basic training in wave physics puts us in an excellent position to understand Quantum Mechanics. We will revisit in chapter III the expermimental facts appearing at the turn of the 19th century that lead to the discovery of Quantum Mechanics. In chapter IV, we will find the Schroedinger Equation, a wave equation for matter waves, and introduce the formalism of Quantum Mechanics at the example of the harmonic oscillator and the hydrogen atom. Both systems are model systems for understanding pho-

tons and energy eigenspectra of quantum systems that can be fully solved analytically.

Chapter V is devoted to the basic understanding of the interaction of light and matter and sets the basis for the study of the laser and its fundamental properties in chapter VI. In chapter VII the modulation of light using electro and acousto-optic effects is discussed. If time permits an introduction to the quantum theory of light fields and photo detection is also treated.

Bibliography

[1] T. H. Maimann, "Stimulated optical radiation in ruby", Nature **187**, 493-494, (1960).

[2] M. D. Perry et al., "Petawatt Laser Pulses," Optics Letters, **24,** 160-163 (1999).

[3] R. Ell, U. Morgner, F.X. Kärtner, J.G. Fujimoto, E.P. Ippen, V. Scheuer, G. Angelow, T. Tschudi: Generation of 5-fs pulses and octave-spanning spectra directly from a Ti:Sappire laser, Opt. Lett. **26**, 373-375 (2001)

[4] H. Hentschel, R. Kienberger, Ch. Spielmann, G. A. Reider, N. Milosevic, T. Brabec, P. Corkum, U. Heinzmann, M. Drescher, F. Krausz: "Attosecond Metrology," Nature **414**, 509-513 (2001).

[5] John Arthur, "Status of the LCLS x-ray FEL program," Review of Scientific Instruments, **73**, 1393-1395 (2002).

Chapter 2

Classical Electromagnetism and Optics

The classical electromagnetic phenomena are completely described by Maxwell's Equations. The simplest case we may consider is that of electrodynamics of isotropic media

2.1 Maxwell's Equations of Isotropic Media

Maxwell's Equations are

$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{J}, \tag{2.1a}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{2.1b}$$

$$\nabla \cdot \vec{D} = \rho, \tag{2.1c}$$

$$\nabla \cdot \vec{B} = 0. \tag{2.1d}$$

The material equations accompanying Maxwell's equations are:

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P}, \tag{2.2a}$$

$$\vec{B} = \mu_0 \vec{H} + \vec{M}. \tag{2.2b}$$

Here, $\vec{E}$ and $\vec{H}$ are the electric and magnetic field, $\vec{D}$ the dielectric flux, $\vec{B}$ the magnetic flux, $\vec{J}$ the current density of free chareges, ρ is the free charge density, $\vec{P}$ is the polarization, and $\vec{M}$ the magnetization.

Note, it is Eqs.(2.2a) and (2.2b) which make electromagnetism an interesting and always a hot topic with never ending possibilities. All advances in engineering of artifical materials or finding of new material properties, such as superconductivity, bring new life, meaning and possibilities into this field.

By taking the curl of Eq. (2.1b) and considering

$$\nabla\times\left(\nabla\times\vec{E}\right)=\nabla\left(\nabla\cdot\vec{E}\right)-\Delta\vec{E},$$

where ∇ is the Nabla operator and Δ the Laplace operator, we obtain

$$\Delta\vec{E}-\mu_0\frac{\partial}{\partial t}\left(\vec{j}+\epsilon_0\frac{\partial\vec{E}}{\partial t}+\frac{\partial\vec{P}}{\partial t}\right)=\frac{\partial}{\partial t}\nabla\times\vec{M}+\nabla\left(\nabla\cdot\vec{E}\right) \tag{2.3}$$

and hence

$$\left(\Delta-\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right)\vec{E}=\mu_0\left(\frac{\partial\vec{j}}{\partial t}+\frac{\partial^2}{\partial t^2}\vec{P}\right)+\frac{\partial}{\partial t}\nabla\times\vec{M}+\nabla\left(\nabla\cdot\vec{E}\right). \tag{2.4}$$

with the vacuum velocity of light

$$c_0=\sqrt{\frac{1}{\mu_0\epsilon_0}}. \tag{2.5}$$

For dielectric non magnetic media, which we often encounter in optics, with no free charges and currents due to free charges, there is $\vec{M}=\vec{0}$, $\vec{J}=\vec{0}$, $\rho=0$, which greatly simplifies the wave equation to

$$\left(\Delta-\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right)\vec{E}=\mu_0\frac{\partial^2}{\partial t^2}\vec{P}+\nabla\left(\nabla\cdot\vec{E}\right). \tag{2.6}$$

2.1.1 Helmholtz Equation

In general, the polarization in dielectric media may have a nonlinear and non local dependence on the field. For linear media the polarizability of the medium is described by a dielectric susceptibility $\chi\left(\vec{r},t\right)$

$$\vec{P}(r,t)=\epsilon_0\int\int d\vec{r}'dt'\;\chi\left(\vec{r}-\vec{r}',t-t'\right)\vec{E}\left(\vec{r}',t'\right). \tag{2.7}$$

The polarization in media with a local dielectric suszeptibility can be described by

$$\vec{P}(r,t) = \epsilon_0 \int dt' \; \chi\left(\vec{r}, t-t'\right) \vec{E}\left(\vec{r}, t'\right). \tag{2.8}$$

This relationship further simplifies for homogeneous media, where the susceptibility does not depend on location

$$\vec{P}(\vec{r},t) = \epsilon_0 \int dt' \; \chi\left(t-t'\right) \vec{E}\left(\vec{r}, t'\right). \tag{2.9}$$

which leads to a dielectric response function or permittivity

$$\epsilon\left(t\right) = \epsilon_0(\delta(t) + \chi\left(t\right)) \tag{2.10}$$

and with it to

$$\vec{D}(r,t) = \int dt' \; \epsilon\left(t-t'\right) \vec{E}\left(\vec{r}, t'\right). \tag{2.11}$$

In such a linear homogeneous medium follows from eq.(2.1c) for the case of no free charges

$$\int dt' \; \epsilon\left(t-t'\right) \left(\nabla \cdot \vec{E}\left(\vec{r}, t'\right)\right) = 0. \tag{2.12}$$

This is certainly fulfilled for $\vec{\nabla} \cdot \vec{E} = 0$, which simplifies the wave equation (2.4) further

$$\left(\Delta - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right) \vec{E} = \mu_0 \frac{\partial^2}{\partial t^2} \vec{P}. \tag{2.13}$$

This is the wave equation driven by the polarization of the medium. If the medium is linear and has only an induced polarization, completely described in the time domain $\chi\left(t\right)$ or in the frequency domain by its Fourier transform, the complex susceptibility $\tilde{\chi}(\omega) = \tilde{\epsilon}_r(\omega) - 1$ with the relative permittivity $\tilde{\epsilon}_r(\omega) = \tilde{\epsilon}(\omega)/\epsilon_0$, we obtain in the frequency domain with the Fourier transform relationship

$$\widetilde{\vec{E}}(z,\omega) = \int_{-\infty}^{+\infty} \vec{E}(z,t) e^{-j\omega t} dt, \tag{2.14}$$

$$\widetilde{\vec{P}}(\omega) = \epsilon_0 \tilde{\chi}(\omega) \widetilde{\vec{E}}(\omega), \tag{2.15}$$

where, the tildes denote the Fourier transforms in the following. Substituted into (2.13)

$$\left(\Delta + \frac{\omega^2}{c_0^2}\right) \widetilde{\vec{E}}(\omega) = -\omega^2 \mu_0 \epsilon_0 \tilde{\chi}(\omega) \widetilde{\vec{E}}(\omega), \tag{2.16}$$

we obtain

$$\left(\Delta + \frac{\omega^2}{c_0^2}(1 + \tilde{\chi}(\omega)\right) \widetilde{\vec{E}}(\omega) = 0, \tag{2.17}$$

with the refractive index $n(\omega)$ and $1 + \tilde{\chi}(\omega) = n(\omega)^2$ results in the Helmholtz equation

$$\left(\Delta + \frac{\omega^2}{c^2}\right) \widetilde{\vec{E}}(\omega) = 0, \tag{2.18}$$

where $c(\omega) = c_0/n(\omega)$ is the velocity of light in the medium. This equation is the starting point for finding monochromatic wave solutions to Maxwell's equations in linear media, as we will study for different cases in the following. Also, so far we have treated the susceptibility $\tilde{\chi}(\omega)$ as a real quantity, which may not always be the case as we will see later in detail.

2.1.2 Plane-Wave Solutions (TEM-Waves) and Complex Notation

The real wave equation (2.13) for a linear medium has real monochromatic plane wave solutions $\vec{E}_{\vec{k}}(\vec{r}, t)$, which can be be written most efficiently in terms of the complex plane-wave solutions $\underline{\vec{E}}_{\vec{k}}(\vec{r}, t)$ according to

$$\vec{E}_{\vec{k}}(\vec{r}, t) = \frac{1}{2}\left[\underline{\vec{E}}_{\vec{k}}(\vec{r}, t) + \underline{\vec{E}}_{\vec{k}}(\vec{r}, t)^*\right] = \Re e\left\{\underline{\vec{E}}_{\vec{k}}(\vec{r}, t)\right\}, \tag{2.19}$$

with

$$\underline{\vec{E}}_{\vec{k}}(\vec{r}, t) = \underline{E}_{\vec{k}}\ e^{\mathrm{j}(\omega t - \vec{k}\cdot\vec{r})}\ \vec{e}(\vec{k}). \tag{2.20}$$

Note, we explicitly underlined the complex wave to indicate that this is a complex quantity. Here, $\vec{e}(\vec{k})$ is a unit vector indicating the direction of the electric field which is also called the polarization of the wave, and $\underline{E}_{\vec{k}}$ is the complex field amplitude of the wave with wave vector $\vec{k}$. Substitution of eq.(2.19) into the wave equation results in the dispersion relation, i.e. a

relationship between wave vector and frequency necessary to satisfy the wave equation

$$|\vec{k}|^2 = \frac{\omega^2}{c(\omega)^2} = k(\omega)^2. \tag{2.21}$$

Thus, the dispersion relation is given by

$$k(\omega) = \pm\frac{\omega}{c_0} n(\omega). \tag{2.22}$$

with the wavenumber

$$k = 2\pi/\lambda, \tag{2.23}$$

where λ is the wavelength of the wave in the medium with refractive index n, ω the angular frequency, $\vec{k}$ the wave vector. Note, the natural frequency $f = \omega/2\pi$. From $\nabla \cdot \vec{E} = 0$, for all time, we see that $\vec{k} \perp \vec{e}$. Substitution of the electric field 2.19 into Maxwell's Eqs. (2.1b) results in the magnetic field

$$\vec{H}_{\vec{k}}(\vec{r}, t) = \frac{1}{2}\left[\underline{\vec{H}}_{\vec{k}}(\vec{r}, t) + \underline{\vec{H}}_{\vec{k}}(\vec{r}, t)^*\right] \tag{2.24}$$

with

$$\underline{\vec{H}}_{\vec{k}}(\vec{r}, t) = \underline{H}_{\vec{k}}\ e^{\mathrm{j}(\omega t - \vec{k}\cdot\vec{r})}\ \vec{h}(\vec{k}). \tag{2.25}$$

This complex component of the magnetic field can be determined from the corresponding complex electric field component using Faraday's law

$$-\mathrm{j}\vec{k} \times \left(\underline{E}_{\vec{k}}\ e^{\mathrm{j}(\omega t - \vec{k}\cdot\vec{r})}\ \vec{e}(\vec{k})\right) = -\mathrm{j}\mu_0\omega\underline{\vec{H}}_{\vec{k}}(\vec{r}, t), \tag{2.26}$$

or

$$\underline{\vec{H}}_{\vec{k}}(\vec{r}, t) = \frac{\underline{E}_{\vec{k}}}{\mu_0\omega} e^{\mathrm{j}(\omega t - \vec{k}\cdot\vec{r})}\vec{k} \times \vec{e} = \underline{H}_{\vec{k}} e^{\mathrm{j}(\omega t - \vec{k}\cdot\vec{r})}\vec{h} \tag{2.27}$$

with

$$\vec{h}(\vec{k}) = \frac{\vec{k}}{|k|} \times \vec{e}(\vec{k}) \tag{2.28}$$

and

$$\underline{H}_{\vec{k}} = \frac{|k|}{\mu_0\omega}\underline{E}_{\vec{k}} = \frac{1}{Z_F}\underline{E}_{\vec{k}}. \tag{2.29}$$

The characteristic impedance of the TEM-wave is the ratio between electric and magnetic field strength

$$Z_F = \mu_0 c = \sqrt{\frac{\mu_0}{\epsilon_0\epsilon_r}} = \frac{1}{n} Z_{F_0} \tag{2.30}$$

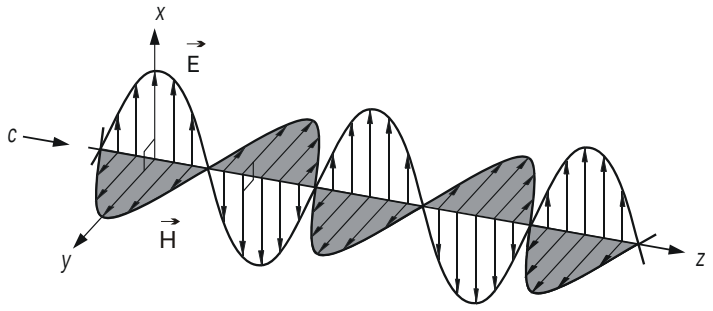

Figure 2.1: Transverse electromagnetic wave (TEM) [6]

with the refractive index $n = \sqrt{\epsilon_r}$ and the free space impedance

$$Z_{F_0} = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 377\,\Omega. \tag{2.31}$$

Note that the vectors $\vec{e}$, $\vec{h}$ and $\vec{k}$ form an orthogonal trihedral,

$$\vec{e} \perp \vec{h}, \quad \vec{k} \perp \vec{e}, \quad \vec{k} \perp \vec{h}. \tag{2.32}$$

That is why we call these waves transverse electromagnetic (TEM) waves. We consider the electric field of a monochromatic electromagnetic wave with frequency ω and electric field amplitude E_0, which propagates in vacuum along the z-axis, and is polarized along the x-axis, (Fig. 2.1), i.e. $\frac{\vec{k}}{|k|} = \vec{e}_z$, and $\vec{e}(\vec{k}) = \vec{e}_x$. Then we obtain from Eqs.(2.19) and (2.20)

$$\vec{E}(\vec{r}, t) = E_0 \cos(\omega t - kz)\ \vec{e}_x, \tag{2.33}$$

and similiar for the magnetic field

$$\vec{H}(\vec{r}, t) = \frac{E_0}{Z_{F_0}} \cos(\omega t - kz)\ \vec{e}_y, \tag{2.34}$$

see Figure 2.1.Note, that for a backward propagating wave with $\underline{\vec{E}}(\vec{r}, t) = \underline{E}\ e^{\mathrm{j}\omega t + \mathrm{j}\vec{k}\cdot\vec{r}}\ \vec{e}_x$, and $\underline{\vec{H}}(\vec{r}, t) = \underline{H}\ e^{\mathrm{j}(\omega t + \vec{k}\vec{r})}\ \vec{e}_y$, there is a sign change for the magnetic field

$$\underline{H} = -\frac{|k|}{\mu_0 \omega} \underline{E}, \tag{2.35}$$

so that the $(\vec{k}, \vec{E}, \vec{H})$ always form a right handed orthogonal system.

2.1.3 Poynting Vectors, Energy Density and Intensity

The table below summarizes the instantaneous and time averaged energy content and energy transport related to an electromagnetic field

Quantity	Real fields	Complex fields
Electric and magnetic energy density	$w_e = \frac{1}{2}\vec{E}\cdot\vec{D} = \frac{1}{2}\epsilon_0\epsilon_r\vec{E}^2$ $w_m = \frac{1}{2}\vec{H}\cdot\vec{B} = \frac{1}{2}\mu_0\mu_r\vec{H}^2$ $w = w_e + w_m$	$\bar{w}_e = \frac{1}{4}\epsilon_0\epsilon_r\left\lvert\underline{\vec{E}}\right\rvert^2$ $\bar{w}_m = \frac{1}{4}\mu_0\mu_r\left\lvert\underline{\vec{H}}\right\rvert^2$ $\bar{w} = \bar{w}_e + \bar{w}_m$
Poynting vector	$\vec{S} = \vec{E}\times\vec{H}$	$\vec{T} = \frac{1}{2}\underline{\vec{E}}\times\underline{\vec{H}}^*$
Poynting theorem	$\mathrm{div}\vec{S} + \vec{E}\cdot\vec{j} + \frac{\partial w}{\partial t} = 0$	$\mathrm{div}\vec{T} + \frac{1}{2}\underline{\vec{E}}\cdot\underline{\vec{j}}^* +$ $+2j\omega(\bar{w}_m - \bar{w}_e) = 0$
Intensity	$I = \left\lvert\vec{S}\right\rvert = cw$	$I = \mathrm{Re}\{\vec{T}\} = c\bar{w}$

Table 2.1: Poynting vector and energy density in EM-fields

For a plane wave with an electric field $\underline{\vec{E}}(\vec{r},t) = \underline{E}e^{\mathrm{j}(\omega t - kz)}\ \vec{e}_x$ we obtain for the energy density in units of [J/m^3]

$$w = \frac{1}{2}\epsilon_r\epsilon_0|\underline{E}|^2, \tag{2.36}$$

the complex Poynting vector

$$\vec{T} = \frac{1}{2Z_F}|\underline{E}|^2\ \vec{e}_z, \tag{2.37}$$

and the intensity in units of [W/m^2]

$$I = \frac{1}{2Z_F}|\underline{E}|^2 = \frac{1}{2}Z_F|\underline{H}|^2. \tag{2.38}$$

2.1.4 Classical Permittivity

In this section we want to get insight into propagation of an electromagnetic wavepacket in an isotropic and homogeneous medium, such as a glass optical fiber due to the interaction of radiation with the medium. The electromagnetic properties of a dielectric medium is largely determined by the electric polarization induced by an electric field in the medium. The polarization is

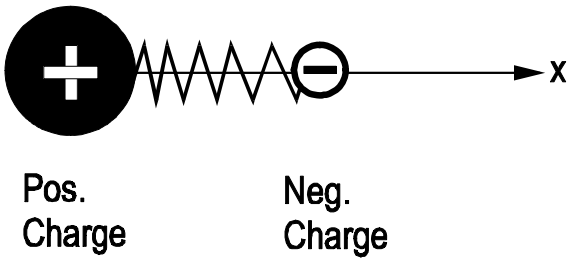

Figure 2.2: Classical harmonic oscillator model for radiation matter interaction

defined as the total induced dipole moment per unit volume. We formulate this directly in the frequency domain

$$\underline{\widetilde{\vec{P}}}(\omega) = \frac{\text{dipole moment}}{\text{volume}} = N \cdot \langle \underline{\widetilde{\vec{p}}}(\omega) \rangle = \epsilon_0 \underline{\widetilde{\chi}}(\omega) \underline{\widetilde{\vec{E}}}(\omega), \tag{2.39}$$

where N is density of elementary units and $\langle \underline{\vec{p}} \rangle$ is the average dipole moment of the unit (atom, molecule, ...). In an isotropic and homogeneous medium the induced polarization is proportional to the electric field and the proportionality constant, $\underline{\widetilde{\chi}}(\omega)$, is called the susceptibility of the medium.

As it turns out (justification later), an electron elastically bound to a positively charged rest atom is not a bad model for understanding the interaction of light with matter at very low electric fields, i.e. the fields do not change the electron distribution in the atom considerably or even ionize the atom, see Figure 2.2. This model is called Lorentz model after the famous physicist A. H. Lorentz (Dutchman) studying electromagnetic phenomena at the turn of the 19th century. He also found the Lorentz Transformation and Invariance of Maxwell's Equations with respect to these transformation, which showed the path to Special Relativity.

The equation of motion for such a unit is the damped harmonic oscillator driven by an electric field in one dimension, x. At optical frequencies, the distance of elongation, x, is much smaller than an optical wavelength (atoms have dimensions on the order of a tenth of a nanometer, whereas optical fields have wavelength on the order of microns) and therefore, we can neglect the spatial variation of the electric field during the motion of the charges within an atom (dipole approximation, i.e. $\vec{E}(\vec{r}, t) = \vec{E}(\vec{r}_A, t) = E(t)\vec{e}_x$).

The equation of motion is

$$m\frac{d^2x}{dt^2} + 2\frac{\Omega_0}{Q}m\frac{dx}{dt} + m\Omega_0^2 x = e_0 E(t), \tag{2.40}$$

where $\underline{E}(t) = \tilde{\underline{E}}e^{\mathrm{j}\omega t}$. Here, m is the mass of the electron assuming the that the rest atom has infinite mass, e_0 the charge of the electron, Ω_0 is the resonance frequency of the undamped oscillator and Q the quality factor of the resonance, which determines the damping of the oscillator. By using the trial solution $x\,(t) = \tilde{x}e^{\mathrm{j}\omega t}$, we obtain for the complex amplitude of the dipole moment $\tilde{p}$ with the time dependent response $p(t) = e_0 x(t) = \underline{\tilde{p}}e^{\mathrm{j}\omega t}$

$$\underline{\tilde{p}} = \frac{\frac{e_0^2}{m}}{(\Omega_0^2 - \omega^2) + 2\mathrm{j}\frac{\Omega_0}{Q}\omega}\underline{\tilde{E}}. \tag{2.41}$$

Note, that we included ad hoc a damping term in the harmonic oscillator equation. At this point it is not clear what the physical origin of this damping term is and we will discuss this at length later in chapter 4. For the moment, we can view this term simply as a consequence of irreversible interactions of the atom with its environment. We then obtain from (2.39) for the susceptibility

$$\underline{\chi}(\omega) = \frac{N\frac{e_0^2}{m}\frac{1}{\epsilon_0}}{(\Omega_0^2 - \omega^2) + 2\mathrm{j}\omega\frac{\Omega_0}{Q}} \tag{2.42}$$

or

$$\underline{\widetilde{\chi}}(\omega) = \frac{\omega_p^2}{(\Omega_0^2 - \omega^2) + 2\mathrm{j}\omega\frac{\Omega_0}{Q}}, \tag{2.43}$$

with ω_p called the plasma frequency, which is defined as $\omega_p^2 = Ne_0^2/m\epsilon_0$. Figure 2.3 shows the normalized real and imaginary part, $\underline{\widetilde{\chi}}(\omega) = \widetilde{\chi}_r(\omega) + \mathrm{j}\widetilde{\chi}_i(\omega)$ of the classical susceptibility (2.43). Note, that there is a small resonance shift (almost invisible) due to the loss. Off resonance, the imaginary part approaches zero very quickly. Not so the real part, which approaches a constant value ω_p^2/Ω_0^2 below resonance for $\omega \to 0$, and approaches zero far above resonance, but much slower than the imaginary part. As we will see later, this is the reason why there are low loss, i.e. transparent, media with refractive index very much different from 1.

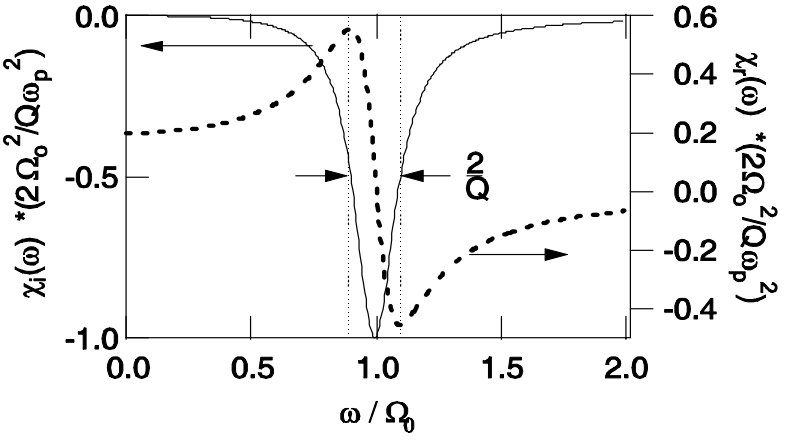

Figure 2.3: Real part (dashed line) and imaginary part (solid line) of the susceptibility of the classical oscillator model for the dielectric polarizability.

2.1.5 Optical Pulses

Optical pulses are wave packets constructed by a continuous superposition of monochromatic plane waves. Consider a TEM-wavepacket, i.e. a superposition of waves with different frequencies, polarized along the x-axis and propagating along the z-axis

$$\underline{\vec{E}}(\vec{r},t) = \int_0^\infty \frac{d\Omega}{2\pi} \underline{\widetilde{E}}(\Omega) e^{\mathrm{j}(\Omega t - K(\Omega)z)} \ \vec{e}_x. \tag{2.44}$$

Correspondingly, the magnetic field is given by

$$\underline{\vec{H}}(\vec{r},t) = \int_0^\infty \frac{d\Omega}{2\pi Z_F(\Omega)} \underline{\widetilde{E}}(\Omega) e^{\mathrm{j}(\Omega t - K(\Omega)z)} \ \vec{e}_y \tag{2.45}$$

Again, the physical electric and magnetic fields are real and related to the complex fields by

$$\vec{E}(\vec{r},t) = \frac{1}{2}\left(\underline{\vec{E}}(\vec{r},t) + \underline{\vec{E}}(\vec{r},t)^*\right) \tag{2.46}$$

$$\vec{H}(\vec{r},t) = \frac{1}{2}\left(\underline{\vec{H}}(\vec{r},t) + \underline{\vec{H}}(\vec{r},t)^*\right). \tag{2.47}$$

Here, $|\underline{\tilde{E}}(\Omega)|e^{\mathrm{j}\varphi(\Omega)}$ is the complex wave amplitude of the electromagnetic wave at frequency Ω and $K(\Omega) = \Omega/c(\Omega) = n(\Omega)\Omega/c_0$ the wavenumber, where,

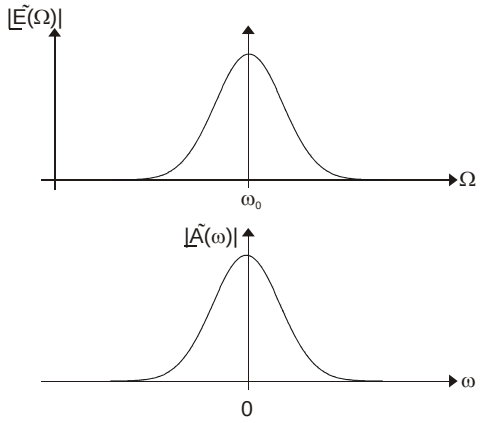

Figure 2.4: Spectrum of an optical wave packet described in absolute and relative frequencies

$n(\Omega)$ is again the refractive index of the medium

$$n^2(\Omega) = 1 + \chi(\Omega), \tag{2.48}$$

c and c_0 are the velocity of light in the medium and in vacuum, respectively. The planes of constant phase propagate with the phase velocity c of the wave.

The wavepacket consists of a superposition of many frequencies with the spectrum shown in Fig. 2.4.

At a given point in space, for simplicity $z = 0$, the complex field of a pulse is given by (Fig. 2.4)

$$\underline{E}(z = 0, t) = \frac{1}{2\pi} \int_0^\infty \underline{\tilde{E}}(\Omega) e^{\mathrm{j}\Omega t} d\Omega. \tag{2.49}$$

Optical pulses often have relatively small spectral width compared to the center frequency of the pulse ω_0, as it is illustrated in the upper part of Figure 2.4. For example typical pulses used in optical communication systems for 10Gb/s transmission speed are on the order of 20ps long and have a center wavelength of $\lambda = 1550$nm. Thus the spectral with is only on the order of 50GHz, whereas the center frequency of the pulse is 200THz, i.e. the bandwidth is 4000 smaller than the center frequency. In such cases it is useful to separate the complex electric field in Eq. (2.49) into a carrier frequency ω_0 and an envelope $\underline{A}(t)$ and represent the absolute frequency as

$\Omega = \omega_0 + \omega$. We can then rewrite Eq.(2.49) as

$$\begin{aligned} \underline{E}(z &= 0, t) = \frac{1}{2\pi}\int_{-\omega_0}^{\infty} \underline{\tilde{E}}(\omega_0+\omega)e^{\mathrm{j}(\omega_0+\omega)t}d\omega \qquad (2.50)\\ &= A(t)e^{\mathrm{j}\omega_0 t}. \end{aligned}$$

The envelope, see Figure 2.8, is given by

$$\begin{aligned} \underline{A}(t) &= \frac{1}{2\pi}\int_{-\omega_0\to-\infty}^{\infty} \underline{\tilde{A}}(\omega)e^{\mathrm{j}\omega t}d\omega \qquad (2.51)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \underline{\tilde{A}}(\omega)e^{\mathrm{j}\omega t}d\omega, \qquad (2.52) \end{aligned}$$

where $\tilde{A}(\omega)$ is the spectrum of the envelope with, $\tilde{A}(\omega) = 0$ for $\omega \leq -\omega_0$. To be physically meaningful, the spectral amplitude $\tilde{A}(\omega)$ must be zero for negative frequencies less than or equal to the carrier frequency, see Figure 2.8. Note, that waves with zero frequency can not propagate, since the corresponding wave vector is zero. The pulse and its envelope are shown in Figure 2.5.

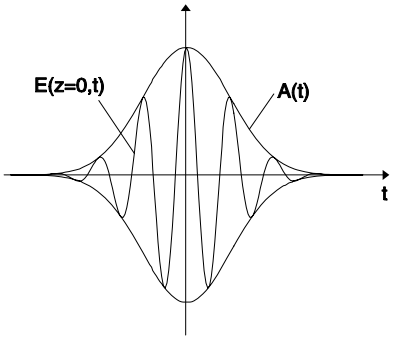

Figure 2.5: Electric field and envelope of an optical pulse.

Table 2.2 shows pulse shape and spectra of some often used pulses as well as the pulse width and time bandwidth products. The pulse width and bandwidth are usually specified as the Full Width at Half Maximum (FWHM) of the intensity in the time domain, $|\underline{A}(t)|^2$, and the spectral density $\left|\underline{\tilde{A}}(\omega)\right|^2$ in the frequency domain, respectively. The pulse shapes and corresponding spectra to the pulses listed in Table 2.2 are shown in Figs 2.6 and 2.7.

Pulse Shape	Fourier Transform	Pulse Width	Time-Band-width Product
$\underline{A}(t)$	$\underline{\tilde{A}}(\omega) = \int_{-\infty}^{\infty} a(t)e^{-\mathrm{j}\omega t}dt$	Δt	$\Delta t \cdot \Delta f$
Gaussian: $\mathrm{e}^{-\frac{t^2}{2\tau^2}}$	$\sqrt{2\pi}\tau\mathrm{e}^{-\frac{1}{2}\tau^2\omega^2}$	$2\sqrt{\ln 2}\tau$	0.441
Hyperbolic Secant: $\mathrm{sech}(\frac{t}{\tau})$	$\frac{\tau}{2}\ \mathrm{sech}\big(\frac{\pi}{2}\tau\omega\big)$	1.7627 τ	0.315
Rect-function: $= \begin{cases} 1, & \lvert t\rvert \le \tau/2 \\ 0, & \lvert t\rvert > \tau/2 \end{cases}$	$\tau\frac{\sin(\tau\omega/2)}{\tau\omega/2}$	τ	0.886
Lorentzian: $\frac{1}{1+(t/\tau)^2}$	$2\pi\tau\mathrm{e}^{-\lvert\tau\omega\rvert}$	1.287 τ	0.142
Double-Exp.: $\mathrm{e}^{-\lvert\frac{t}{\tau}\rvert}$	$\frac{\tau}{1+(\omega\tau)^2}$	ln2 τ	0.142

Table 2.2: Pulse shapes, corresponding spectra and time bandwidth products.

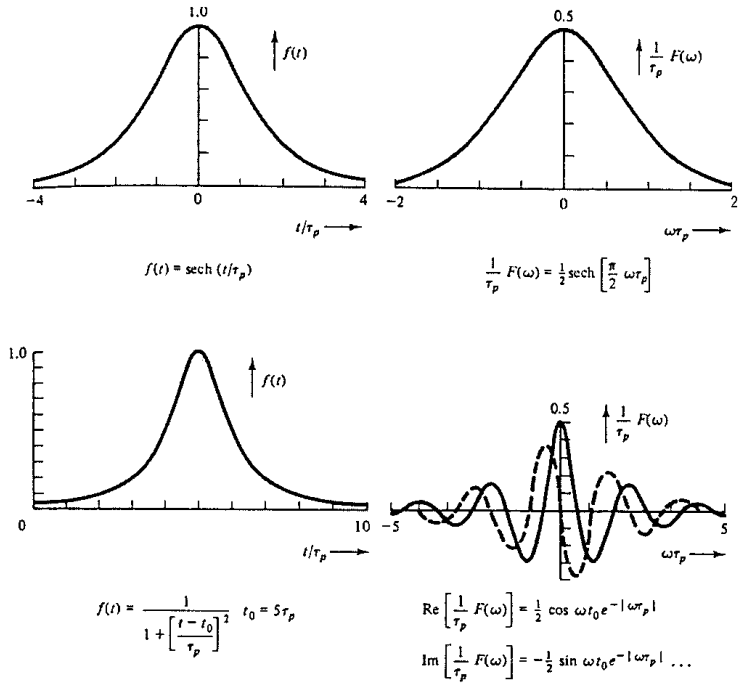

Figure 2.6: Fourier transforms to pulse shapes listed in table 2.2 [6].

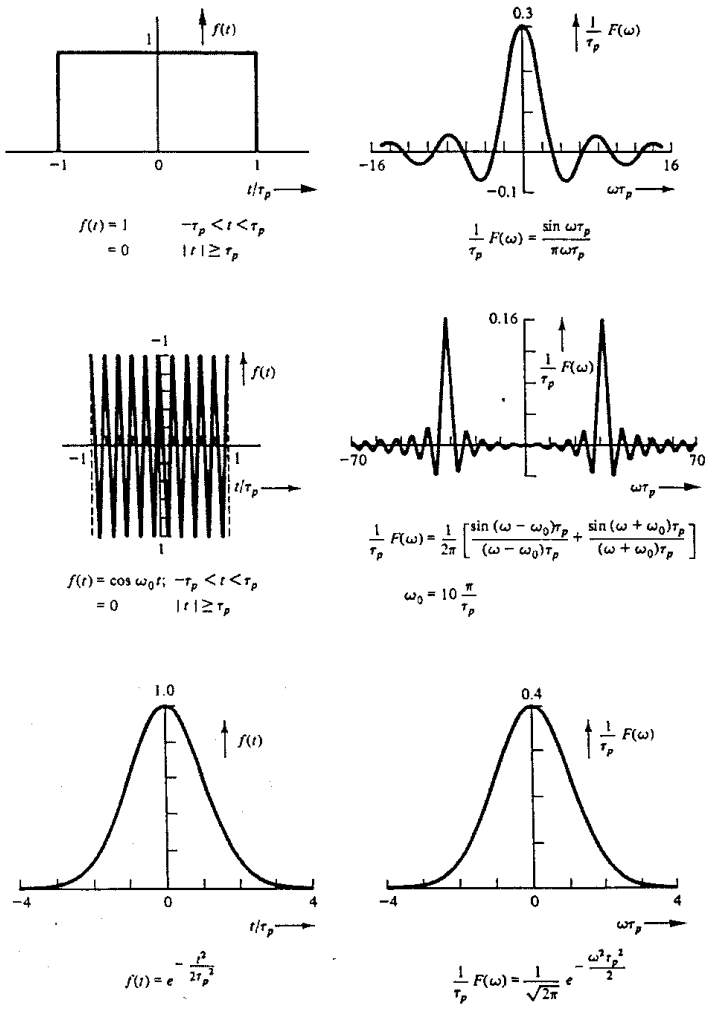

Figure 2.7: Fourier transforms to pulse shapes listed in table 2.2 continued [6].

2.1.6 Pulse Propagation

Having a basic model for the interaction of light and matter at hand, via section 2.1.4, we can investigate what happens if an electromagnetic wave packet, i.e. an optical pulse propagates through such a medium. We start from Eqs.(2.44) to evaluate the wave packet propagation for an arbitrary

propagation distance z

$$\underline{E}(z,t) = \frac{1}{2\pi}\int_0^\infty \underline{\tilde{E}}(\Omega)e^{\mathrm{j}(\Omega t - K(\Omega)z)}d\Omega. \tag{2.53}$$

Analogous to Eq. (2.50) for a pulse at a given position, we can separate an optical pulse into a carrier wave at frequency ω_0 and a complex envelope $A(z,t)$,

$$\underline{E}(z,t) = \underline{A}(z,t)e^{\mathrm{j}(\omega_0 t - K(\omega_0)z)}. \tag{2.54}$$

By introducing the offset frequency ω, the offset wavenumber $k(\omega)$ and spectrum of the envelope $\underline{\tilde{A}}(\omega)$

$$\begin{aligned} \omega &= \Omega - \omega_0, & (2.55)\\ k(\omega) &= K(\omega_0+\omega) - K(\omega_0), & (2.56)\\ \underline{\tilde{A}}(\omega) &= \underline{\tilde{E}}(\Omega = \omega_0 + \omega). & (2.57) \end{aligned}$$

the envelope at propagation distance z, see Fig.2.8, is expressed as

$$\underline{A}(z,t) = \frac{1}{2\pi}\int_{-\infty}^\infty \underline{\tilde{A}}(\omega)e^{\mathrm{j}(\omega t - k(\omega)z)}d\omega, \tag{2.58}$$

with the same constraints on the spectrum of the envelope as before, i.e. the spectrum of the envelope must be zero for negative frequencies beyond the carrier frequency. Depending on the dispersion relation $k(\omega)$, (see Fig. 2.9),.the pulse will be reshaped during propagation as discussed in the following section.

2.1.7 Dispersion

The dispersion relation indicates how much phase shift each frequency component experiences during propagation. These phase shifts, if not linear with respect to frequency, will lead to distortions of the pulse. If the propagation constant $k(\omega)$ is only slowly varying over the pulse spectrum, it is useful to represent the propagation constant, $k(\omega)$, or dispersion relation $K(\Omega)$ by its Taylor expansion, see Fig. 2.9,

$$k(\omega) = k'\omega + \frac{k''}{2}\omega^2 + \frac{k^{(3)}}{6}\omega^3 + O(\omega^4). \tag{2.59}$$

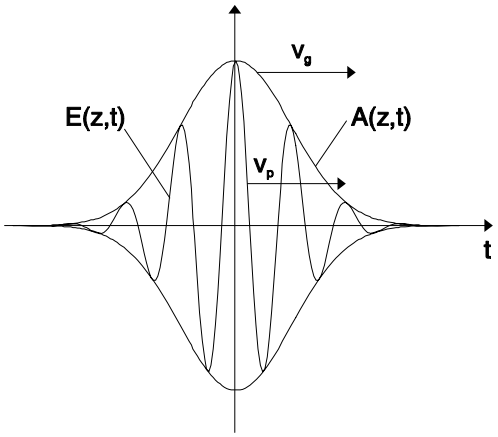

Figure 2.8: Electric field and pulse envelope in time domain.

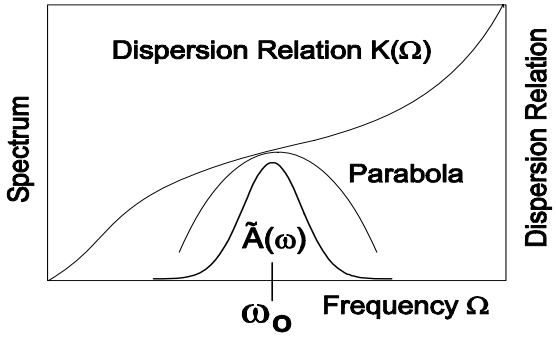

Figure 2.9: Taylor expansion of dispersion relation at the center frequency of the wave packet.

If the refractive index depends on frequency, the dispersion relation is no longer linear with respect to frequency, see Fig. 2.9 and the pulse propagation according to (2.58) can be understood most easily in the frequency domain

$$\frac{\partial \underline{\tilde{A}}(z,\omega)}{\partial z} = -\mathrm{j}k(\omega)\underline{\tilde{A}}(z,\omega). \tag{2.60}$$

Transformation of Eq.() into the time domain gives

$$\frac{\partial \underline{A}(z,t)}{\partial z} = -\mathrm{j}\sum_{n=1}^{\infty} \frac{k^{(n)}}{n!}\left(-\mathrm{j}\frac{\partial}{\partial t}\right)^n \underline{A}(z,t). \tag{2.61}$$

If we keep only the first term, the linear term, in Eq.(2.59), then we obtain for the pulse envelope from (2.58) by definition of the group velocity at frequency ω_0

$$v_{g0} = 1/k' = \left(\left. \frac{dk(\omega)}{d\omega} \right|_{\omega=0} \right)^{-1} \tag{2.62}$$

$$\underline{A}(z,t) = \underline{A}(0, t - z/v_{g0}). \tag{2.63}$$

Thus the derivative of the dispersion relation at the carrier frequency determines the propagation velocity of the envelope of the wave packet or group velocity, whereas the ratio between propagation constant and frequency determines the phase velocity of the carrier

$$v_{p0} = \omega_0/K(\omega_0) = \left(\frac{K(\omega_0)}{\omega_0} \right)^{-1}. \tag{2.64}$$

To get rid of the trivial motion of the pulse envelope with the group velocity, we introduce the retarded time $t' = t - z/v_{g0}$. With respect to this retarded time the pulse shape is invariant during propagation, if we approximate the dispersion relation by the slope at the carrier frequency

$$\underline{A}(z,t) = \underline{A}(0,t'). \tag{2.65}$$

Note, if we approximate the dispersion relation by its slope at the carrier frequency, i.e. we retain only the first term in Eq.(2.61), we obtain

$$\frac{\partial \underline{A}(z,t)}{\partial z} + \frac{1}{v_{g0}} \frac{\partial \underline{A}(z,t)}{\partial t} = 0, \tag{2.66}$$

and (2.63) is its solution. If, we transform this equation to the new coordinate system

$$z' = z, \tag{2.67}$$

$$t' = t - z/v_{g0}, \tag{2.68}$$

with

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial z'} - \frac{1}{v_{g0}} \frac{\partial}{\partial t'}, \tag{2.69}$$

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial t'} \tag{2.70}$$

the transformed equation is

$$\frac{\partial \underline{A}(z',t')}{\partial z'} = 0. \tag{2.71}$$

Since z is equal to z' we keep z in the following.

If the spectrum of the pulse is broad enough, so that the second order term in (2.59) becomes important, the pulse will no longer keep its shape. When keeping in the dispersion relation terms up to second order it follows from (2.58) and (2.69,2.70)

$$\frac{\partial \underline{A}(z,t')}{\partial z} = \mathrm{j}\frac{k''}{2}\frac{\partial^2 \underline{A}(z,t')}{\partial t'^2}. \tag{2.72}$$

This is the first non trivial term in the wave equation for the envelope. Because of the superposition principle, the pulse can be thought of to be decomposed into wavepackets (sub-pulses) with different center frequencies. Now, the group velocity depends on the spectral component of the pulse, see Figure 2.10, which will lead to broadening or dispersion of the pulse.

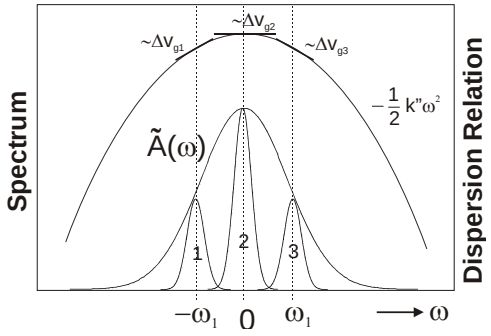

Figure 2.10: Decomposition of a pulse into wave packets with different center frequency. In a medium with dispersion the wavepackets move at different relative group velocity.

Fortunately, for a Gaussian pulse, the pulse propagation equation 2.72 can be solved analytically. The initial pulse is then of the form

$$\underline{E}(z = 0,t) = \underline{A}(z=0,t)e^{\mathrm{j}\omega_0 t} \tag{2.73}$$

$$\underline{A}(z = 0, t = t') = \underline{A}_0 \exp\left[-\frac{1}{2}\frac{t'^2}{\tau^2}\right] \tag{2.74}$$

Eq.(2.72) is most easily solved in the frequency domain since it transforms to

$$\frac{\partial \tilde{\underline{A}}(z,\omega)}{\partial z} = -\mathrm{j}\frac{k''\omega^2}{2}\tilde{\underline{A}}(z,\omega), \tag{2.75}$$

with the solution

$$\tilde{\underline{A}}(z,\omega) = \tilde{\underline{A}}(z=0,\omega)\exp\left[-\mathrm{j}\frac{k''\omega^2}{2}z\right]. \tag{2.76}$$

The pulse spectrum aquires a parabolic phase. Note, that here ω is the Fourier Transform variable conjugate to t' rather than t. The Gaussian pulse has the advantage that its Fourier transform is also a Gaussian

$$\tilde{\underline{A}}(z=0,\omega) = A_0\sqrt{2\pi}\tau\exp\left[-\frac{1}{2}\tau^2\omega^2\right] \tag{2.77}$$

and, therefore, in the spectral domain the solution at an arbitray propagation distance z is

$$\tilde{\underline{A}}(z,\omega) = A_0\sqrt{2\pi}\tau\exp\left[-\frac{1}{2}\left(\tau^2+\mathrm{j}k''z\right)\omega^2\right]. \tag{2.78}$$

The inverse Fourier transform is analogously

$$\underline{A}(z,t') = A_0\left(\frac{\tau^2}{(\tau^2+\mathrm{j}k''z)}\right)^{1/2}\exp\left[-\frac{1}{2}\frac{t'^2}{(\tau^2+\mathrm{j}k''z)}\right] \tag{2.79}$$

The exponent can be written as real and imaginary part and we finally obtain

$$\underline{A}(z,t') = A_0\left(\frac{\tau^2}{(\tau^2+\mathrm{j}k''z)}\right)^{1/2}\exp\left[-\frac{1}{2}\frac{\tau^2t'^2}{\left(\tau^4+\left(k''z\right)^2\right)}+\mathrm{j}\frac{1}{2}k''z\frac{t'^2}{\left(\tau^4+\left(k''z\right)^2\right)}\right] \tag{2.80}$$

As we see from Eq.(2.80) during propagation the FWHM of the Gaussian determined by

$$\exp\left[-\frac{\tau(\tau'_{FWHM}/2)^2}{\left(\tau^4+\left(k''z\right)^2\right)}\right] = 0.5 \tag{2.81}$$

changes from

$$\tau_{FWHM} = 2\sqrt{\ln 2}\,\tau \tag{2.82}$$

at the start to

$$\begin{aligned} \tau'_{FWHM} &= 2\sqrt{\ln 2}\,\tau\sqrt{1+\left(\frac{k''L}{\tau^2}\right)^2} \\ &= \tau_{FWHM}\sqrt{1+\left(\frac{k''L}{\tau^2}\right)^2} \end{aligned} \tag{2.83}$$

at $z = L$. For large stretching this result simplifies to

$$\tau'_{FWHM} = 2\sqrt{\ln 2}\,\left|\frac{k''L}{\tau}\right| \text{ for } \left|\frac{k''L}{\tau^2}\right| \gg 1. \tag{2.84}$$

The strongly dispersed pulse has a width equal to the difference in group delay over the spectral width of the pulse.

Figure 2.11 shows the evolution of the magnitude of the Gaussian wave packet during propagation in a medium which has no higher order dispersion in normalized units. The pulse spreads continuously.

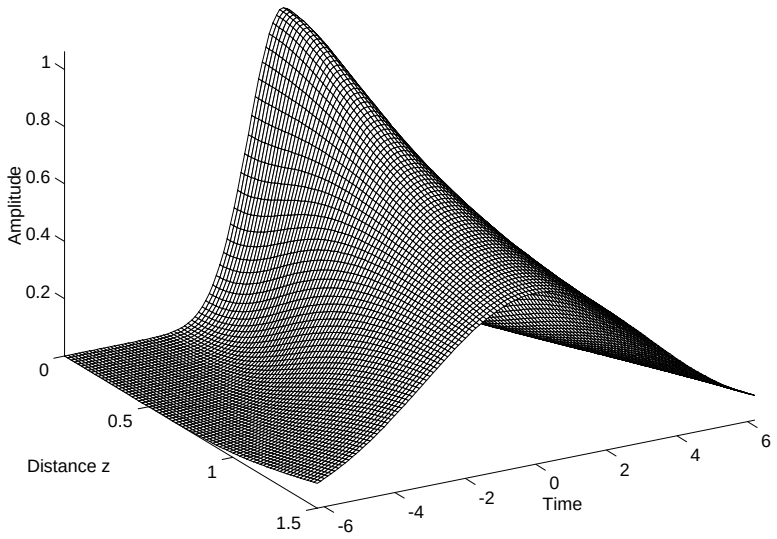

Figure 2.11: Magnitude of the complex envelope of a Gaussian pulse, $|\underline{A}(z,t')|$, in a dispersive medium.

As discussed before, the origin of this spreading is the group velocity dispersion (GVD), $k'' \neq 0$. The group velocity varies over the pulse spectrum significantly leading to a group delay dispersion (GDD) after a propagation distance $z = L$ of $k''L \neq 0$, for the different frequency components. This leads to the build-up of chirp in the pulse during propagation. We can understand this chirp by looking at the parabolic phase that develops over the pulse in time at a fixed propagation distance. The phase is, see Eq.(2.80)

$$\phi(z = L, t') = -\frac{1}{2}\arctan\left[\frac{k''L}{\tau^2}\right] + \frac{1}{2}k''L\frac{t'^2}{\left(\tau^4 + (k''L)^2\right)}. \tag{2.85}$$

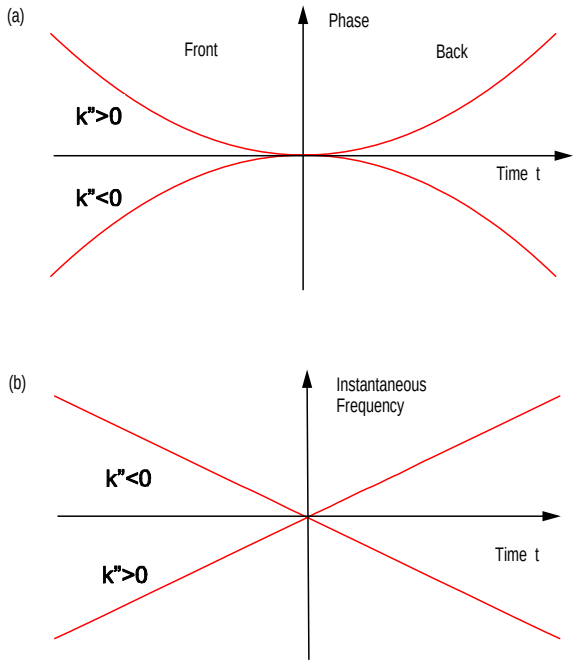

Figure 2.12: (a) Phase and (b) instantaneous frequency of a Gaussian pulse during propagation through a medium with positive or negative dispersion.

This parabolic phase, see Fig. 2.12 (a), can be understood as a localy varying frequency in the pulse, i.e. the derivative of the phase gives the

instantaneous frequency shift in the pulse with respect to the center frequency

$$\omega(z = L, t') = \frac{\partial}{\partial t'}\phi(L, t') = \frac{k''L}{\left(\tau^4 + (k''L)^2\right)}t' \tag{2.86}$$

see Fig.2.12 (b). The instantaneous frequency indicates that for a medium with positive GVD, ie. $k'' > 0$, the low frequencies are in the front of the pulse, whereas the high frequencies are in the back of the pulse, since the sub-pulses with lower frequencies travel faster than sub-pulses with higher frequencies. The opposite is the case for negative dispersive materials.

It is instructive for later purposes, that this behaviour can be completely understood from the center of mass motion of the sub-pulses, see Figure 2.10. Note, we can choose a set of sub-pulses, with such narrow bandwidth, that dispersion does not matter. In the time domain, these pulses are of course very long, because of the time bandwidth relationship. Nevertheless, since they all have different carrier frequencies, they interfere with each other in such a way that the superposition is a very narrow pulse. This interference, becomes destroyed during propagation, since the sub-pulses propagate at different speed, i.e. their center of mass propagates at different speed.

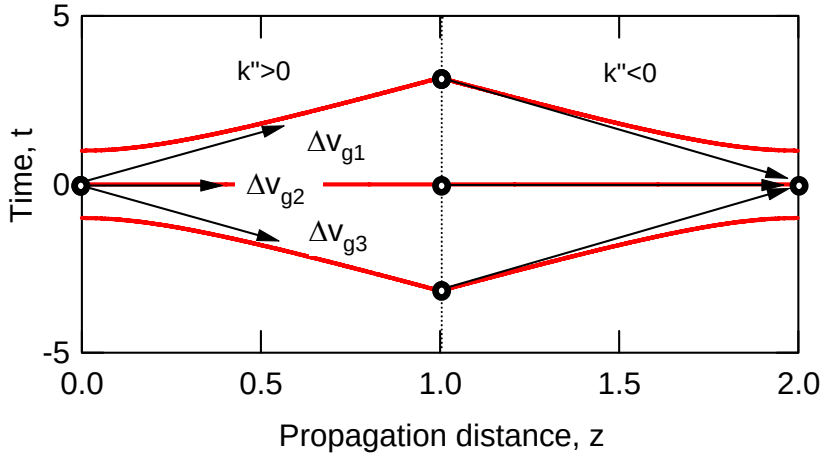

Figure 2.13: Pulse spreading by following the center of mass of sub-pulses according to Fig. 2.10. For $z < 1$, the pulses propagate in a medium with positive dispersion and for $z > 1$ in a medium with negative dispersion.

The differential group delay $\Delta T_g(\omega) = k''L\omega$ of a sub-pulse with its center frequency ω different from 0, is due to its differential group velocity $\Delta v_g(\omega) = -v_{g0}\Delta T_g(\omega)/T_{g0} = -v_{g0}^2 k''\omega$. Note, that $T_{g0} = L/v_{g0}$. This is illustrated in Figure 2.13 by ploting the trajectory of the relative motion of the center of mass of each sub-pulse as a function of propagation distance, which asymptotically approaches the formula for the pulse width of the highly dispersed pulse Eq.(2.84). If we assume that the pulse propagates through a negative dispersive medium following the positive dispersive medium, the group velocity of each sub-pulse is reversed. The sub-pulses propagate towards each other until they all meet at one point (focus) to produce again a short and unchirped initial pulse, see Figure 2.13. This is a very powerful technique to understand dispersive wave motion and as we will see in the next section is the connection between ray optics and physical optics.

2.1.8 Loss and Gain

If the medium considered has loss, described by the imaginary part of the dielectric susceptibility, see (2.43) and Fig. 2.3, we can incorporate this loss into a complex refractive index

$$\underline{n}(\Omega) = n_r(\Omega) + \mathrm{j}n_i(\Omega) \tag{2.87}$$

via

$$\underline{n}(\Omega) = \sqrt{1 + \underline{\widetilde{\chi}}(\Omega)}. \tag{2.88}$$

For an optically thin medium, i.e. $\underline{\widetilde{\chi}} \ll 1$ the following approximation is very useful

$$\underline{n}(\Omega) \approx 1 + \frac{\underline{\widetilde{\chi}}(\Omega)}{2}. \tag{2.89}$$

As one can show (in Recitations) the complex susceptibility (2.43) can be approximated close to resonance, i.e. $\Omega \approx \Omega_0$, by the complex Lorentzian lineshape

$$\underline{\widetilde{\chi}}(\Omega) = \frac{-\mathrm{j}\chi_0}{1 + \mathrm{j}Q\frac{\Omega-\Omega_0}{\Omega_0}}, \tag{2.90}$$

where $\chi_0 = Q\frac{\omega_p^2}{2\Omega_0^2}$ will turn out to be related to the peak absorption of the line, which is proportional to the density of atoms, Ω_0 is the center frequency

and $\Delta\Omega = \frac{\Omega_0}{Q}$ is the half width half maximum (HWHM) linewidth of the transition. The real and imaginary part of the complex Lorentzian are

$$\widetilde{\chi}_r(\Omega) = \frac{-\chi_0 \frac{(\Omega-\Omega_0)}{\Delta\Omega}}{1+\left(\frac{\Omega-\Omega_0}{\Delta\Omega}\right)^2}, \tag{2.91}$$

$$\widetilde{\chi}_i(\Omega) = \frac{-\chi_0}{1+\left(\frac{\Omega-\Omega_0}{\Delta\Omega}\right)^2},. \tag{2.92}$$

In the derivation of the wave equation for the pulse envelope (2.61) in section 2.1.7, there was no restriction to a real refractive index. Therefore, the wave equation (2.61) also treats the case of a complex refractive index. If we assume a medium with the complex refractive index (2.89), then the wavenumber is given by

$$\underline{K}(\Omega) = \frac{\Omega}{c_0}\left(1+\frac{1}{2}\left(\widetilde{\chi}_r(\Omega)+\mathrm{j}\widetilde{\chi}_i(\Omega)\right)\right). \tag{2.93}$$

Since we introduced a complex wavenumber, we have to redefine the group velocity as the inverse derivative of the real part of the wavenumber with respect to frequency. At line center, we obtain

$$\upsilon_g^{-1} = \left.\frac{\partial K_r(\Omega)}{\partial\Omega}\right|_{\Omega_0} = \frac{1}{c_0}\left(1-\frac{\chi_0}{2}\frac{\Omega_0}{\Delta\Omega}\right). \tag{2.94}$$

Thus, for a narrow absorption line, $\chi_0 > 0$ and $\frac{\Omega_0}{\Delta\Omega} \gg 1$, the absolute value of the group velocity can become much larger than the velocity of light in vacuum. The opposite is true for an amplifying medium, $\chi_0 < 0$. There is nothing wrong with this finding, since the group velocity only describes the motion of the peak of a Gaussian wave packet, which is not a causal wave packet. A causal wave packet is identical to zero for some earlier time $t < t_0$, in some region of space. A Gaussian wave packet fills the whole space at any time and can be reconstructed by a Taylor expansion at any time. Therefore, the tachionic motion of the peak of such a signal does not contradict special relativity.

The imaginary part in the wave vector (2.93) leads with $K = \frac{\Omega}{c_0}$ to absorption

$$\alpha(\Omega) = -K\widetilde{\chi}_i(\Omega). \tag{2.95}$$

In the envelope equation (2.60) for a wavepacket with carrier frequency $\omega_0 = \Omega_0$ and $K_0 = \frac{\Omega_0}{c_0}$ the loss leads to a term of the form

$$\left.\frac{\partial \tilde{\underline{A}}(z,\omega)}{\partial z}\right|_{(loss)} = -\alpha(\Omega_0+\omega)\tilde{A}(z,\omega) = \frac{-\chi_0 K_0}{1+\left(\frac{\omega}{\Delta\Omega}\right)^2}\tilde{\underline{A}}(z,\omega). \tag{2.96}$$

In the time domain, we obtain up to second order in the inverse linewidth

$$\left.\frac{\partial \underline{A}(z,t')}{\partial z}\right|_{(loss)} = -\chi_0 K_0\left(1+\frac{1}{\Delta\Omega^2}\frac{\partial^2}{\partial t^2}\right)\underline{A}(z,t'), \tag{2.97}$$

which corresponds to a parabolic approximation of the line shape at line center, (Fig. 2.3). As we will see later, for an amplifying optical transition we obtain a similar equation. We only have to replace the loss by gain

$$\left.\frac{\partial \underline{A}(z,t')}{\partial z}\right|_{(gain)} = g\left(1+\frac{1}{\Omega_g^2}\frac{\partial^2}{\partial t^2}\right)\underline{A}(z,t'), \tag{2.98}$$

where $g = -\chi_0 K_0$ is the peak gain at line center per unit length and Ω_g is the HWHM linewidth of a transition providing gain.

2.1.9 Sellmeier Equation and Kramers-Kroenig Relations

The linear susceptibility is the frequency response or impulse response of a linear system to an applied electric field, see Eq.(2.41). For a real physical system this response is causal, and therefore real and imaginary parts obey Kramers-Kroenig Relations

$$\chi_r(\Omega) = \frac{2}{\pi}\int_0^\infty \frac{\omega\chi_i(\omega)}{\omega^2-\Omega^2}d\omega = n_r^2(\Omega)-1, \tag{2.99}$$

$$\chi_i(\Omega) = -\frac{2}{\pi}\int_0^\infty \frac{\Omega\chi_r(\omega)}{\omega^2-\Omega^2}d\omega. \tag{2.100}$$

For optical media these relations have the consequence that the refractive index and absorption of a medium are not independent, which can often be exploited to compute the index from absorption data or the other way

around. The Kramers-Kroenig Relations also give us a good understanding of the index variations in transparent media, which means the media are used in a frequency range far away from resonances. Then the imaginary part of the susceptibility related to absorption can be approximated by

$$\chi_i(\Omega) = \sum_i A_i \delta(\omega - \omega_i) \tag{2.101}$$

and the Kramers-Kroenig relation results in the Sellmeier Equation for the refractive index

$$n^2(\Omega) = 1 + \sum_i A_i \frac{\omega_i}{\omega_i^2 - \Omega^2} = 1 + \sum_i a_i \frac{\lambda}{\lambda^2 - \lambda_i^2}. \tag{2.102}$$

This formula is very useful in fitting the refractive index of various media over a large frequency range with relatively few coefficients. For example Table 2.3 shows the sellmeier coefficients for fused quartz and sapphire.

	Fused Quartz	Sapphire
a_1	0.6961663	1.023798
a_2	0.4079426	1.058364
a_3	0.8974794	5.280792
λ_1^2	$4.679148 \cdot 10^{-3}$	$3.77588 \cdot 10^{-3}$
λ_2^2	$1.3512063 \cdot 10^{-2}$	$1.22544 \cdot 10^{-2}$
λ_3^2	$0.9793400 \cdot 10^{2}$	$3.213616 \cdot 10^{2}$

Table 2.3: Table with Sellmeier coefficients for fused quartz and sapphire.

In general, each absorption line contributes a corresponding index change to the overall optical characteristics of a material, see Fig. 2.14. A typical situation for a material having resonances in the UV and IR, such as glass, is shown in Fig. 2.15. As Fig. 2.15 shows, due to the Lorentzian line shape, that outside of an absorption line the refractive index is always decreasing as a function of wavelength. This behavior is called normal dispersion and the opposite behavior abnormal dispersion.

$$\frac{dn}{d\lambda} < 0 : \text{normal dispersion (blue refracts more than red)}$$
$$\frac{dn}{d\lambda} > 0 : \text{abnormal dispersion}$$

This behavior is also responsible for the mostly positive group delay dispersion over the transparency range of a material, as the group velocity or group delay dispersion is closely related to $\frac{dn}{d\lambda}$. Fig.2.16 shows the transparency range of some often used media.

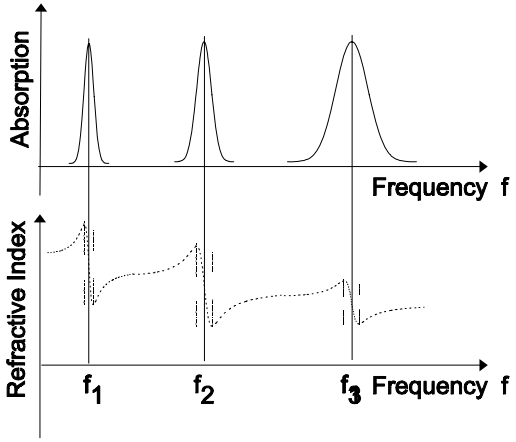

Figure 2.14: Each absorption line must contribute to an index change via the Kramers-Kroenig relations.

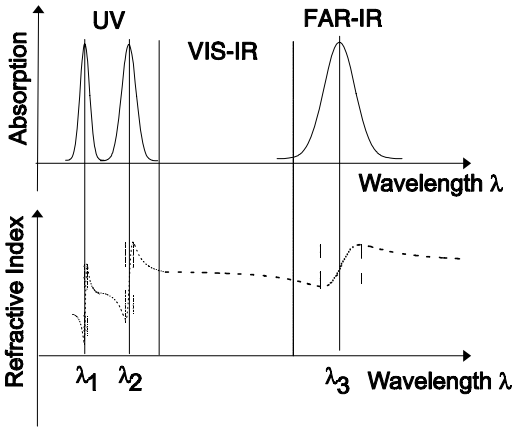

Figure 2.15: Typcial distribution of absorption lines in a medium transparent in the visible.

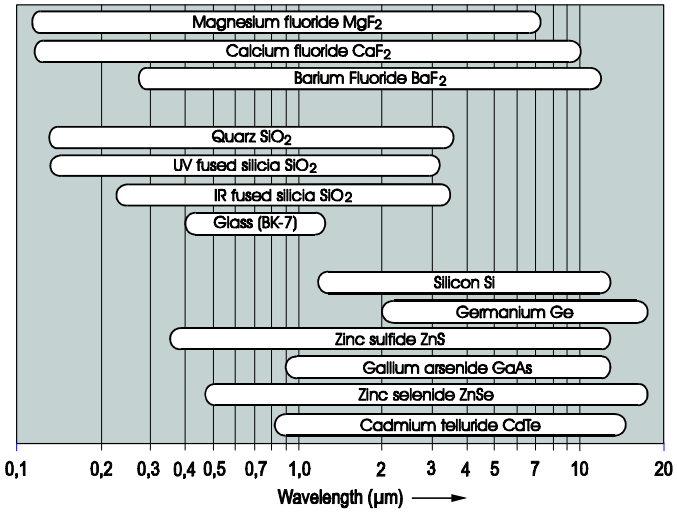

Figure 2.16: Transparency range of some materials according to [6], p. 175.

Often the dispersion GVD and GDD needs to be calculated from the Sellmeier equation, i.e. $n(\lambda)$. The corresponding quantities are listed in Table 2.4. The computations are done by substituting the frequency with the wavelength.

Dispersion Characteristic	Definition	Comp. from $n(\lambda)$
medium wavelength: λ_n	$\frac{\lambda}{n}$	$\frac{\lambda}{n(\lambda)}$
wavenumber: k	$\frac{2\pi}{\lambda_n}$	$\frac{2\pi}{\lambda}n(\lambda)$
phase velocity: υ_p	$\frac{\omega}{k}$	$\frac{c_0}{n(\lambda)}$
group velocity: υ_g	$\frac{d\omega}{dk}; d\lambda = \frac{-\lambda^2}{2\pi c_0}d\omega$	$\frac{c_0}{n}\left(1-\frac{\lambda}{n}\frac{dn}{d\lambda}\right)^{-1}$
group velocity dispersion: GVD	$\frac{d^2k}{d\omega^2}$	$\frac{\lambda^3}{2\pi c_0^2}\frac{d^2n}{d\lambda^2}$
group delay: $T_g = \frac{L}{\upsilon_g} = \frac{d\phi}{d\omega}$	$\frac{d\phi}{d\omega} = \frac{d(kL)}{d\omega}$	$\frac{n}{c_0}\left(1-\frac{\lambda}{n}\frac{dn}{d\lambda}\right)L$
group delay dispersion: GDD	$\frac{dT_g}{d\omega} = \frac{d^2(kL)}{d\omega^2}$	$\frac{\lambda^3}{2\pi c_0^2}\frac{d^2n}{d\lambda^2}L$

Table 2.4: Table with important dispersion characteristics and how to compute them from the wavelength dependent refractive index $n(\lambda)$.

2.2 Electromagnetic Waves and Interfaces

Many microwave and optical devices are based on the characteristics of electromagnetic waves undergoing reflection or transmission at interfaces between media with different electric or magnetic properties characterized by ϵ and μ, see Fig. 2.17. Without restriction we can assume that the interface is the (x-y-plane) and the plane of incidence is the (x-z-plane). An arbitrary incident plane wave can always be decomposed into two components. One component has its electric field parallel to the interface between the media, i.e. it is polarized parallel to the interface and it is called the transverse electric (TE)-wave or also s-polarized wave. The other component is polarized in the plane of incidence and its magnetic field is in the plane of the interface between the media. This wave is called the TM-wave or also p-polarized wave. The most general case of an incident monochromatic TEM-wave is a linear superposition of a TE and a TM-wave.

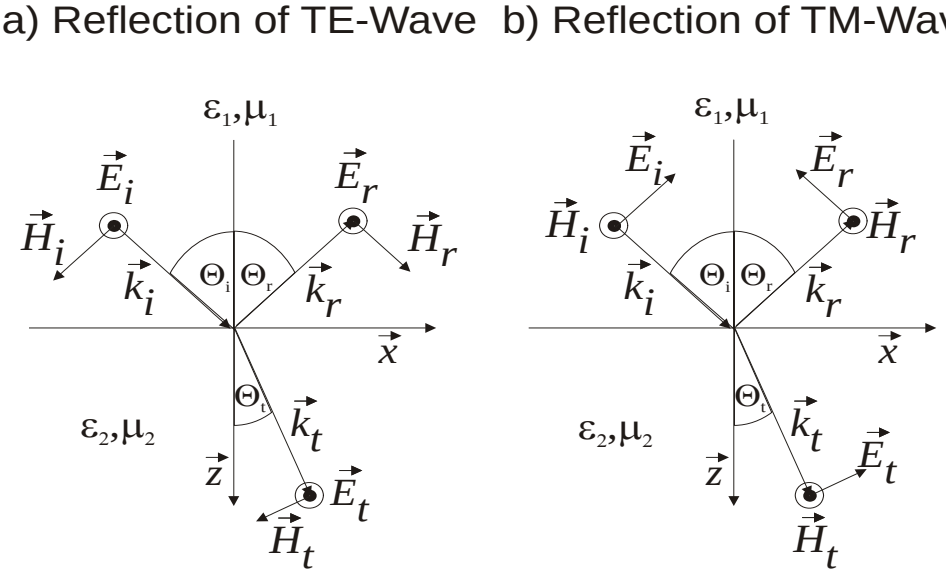

Figure 2.17: a) Reflection of a TE-wave at an interface, b) Reflection of aTM-wave at an interface

The fields for both cases are summarized in table 2.5

TE-wave	TM-wave
$\vec{\underline{E}}_i = \underline{E}_i\ e^{\mathrm{j}(\omega t - \vec{k}_i\cdot\vec{r})}\vec{e}_y$	$\vec{\underline{E}}_i = -\underline{E}_i\ e^{\mathrm{j}(\omega t - \vec{k}_i\cdot\vec{r})}\vec{e}_i$
$\vec{\underline{H}}_i = \underline{H}_i\ e^{\mathrm{j}(\omega t - \vec{k}_i\cdot\vec{r})}\vec{h}_i$	$\vec{\underline{H}}_i = \underline{H}_i\ e^{\mathrm{j}(\omega t - \vec{k}_i\cdot\vec{r})}\vec{e}_y$
$\vec{\underline{E}}_r = \underline{E}_r\ e^{\mathrm{j}(\omega t - \vec{k}_r\cdot\vec{r})}\vec{e}_y$	$\vec{\underline{E}}_r = \underline{E}_r\ e_r^{\mathrm{j}(\omega t - \vec{k}_r\cdot\vec{r})}\vec{e}_r$
$\vec{\underline{H}}_r = \underline{H}_r\ e^{\mathrm{j}(\omega t - \vec{k}_r\cdot\vec{r})}\vec{h}_r$	$\vec{\underline{H}}_r = \underline{E}_r\ e^{\mathrm{j}(\omega t - \vec{k}_r\cdot\vec{r})}\vec{e}_y$
$\vec{\underline{E}}_t = \underline{E}_t\ e^{\mathrm{j}(\omega t - \vec{k}_t\cdot\vec{r})}\vec{e}_y$	$\vec{\underline{E}}_t = \underline{E}_t\ e^{\mathrm{j}(\omega t - \vec{k}_t\cdot\vec{r})}\vec{e}_t$
$\vec{\underline{H}}_t = \underline{H}_t\ e^{\mathrm{j}(\omega t - \vec{k}_t\cdot\vec{r})}\vec{h}_t$	$\vec{\underline{H}}_t = \underline{H}_t\ e^{\mathrm{j}(\omega t - \vec{k}_t\cdot\vec{r})}\vec{e}_y$

Table 2.5: Electric and magnetic fields for TE- and TM-waves.

with wave vectors of the waves given by

$$\begin{aligned} k_i &= k_r = k_0\sqrt{\epsilon_1\mu_1}, \\ k_t &= k_0\sqrt{\epsilon_2\mu_2}, \\ \vec{k}_{i,t} &= k_{i,t}\left(\sin\theta_{i,t}\ \vec{e}_x + \cos\theta_{i,t}\ \vec{e}_z\right), \\ \vec{k}_r &= k_i\left(\sin\theta_r\ \vec{e}_x - \cos\theta_r\ \vec{e}_z\right), \end{aligned}$$

and unit vectors given by

$$\begin{aligned} \vec{h}_{i,t} &= -\cos\theta_{i,t}\ \vec{e}_x + \sin\theta_{i,t}\ \vec{e}_z, \\ \vec{h}_r &= \cos\theta_r\ \vec{e}_x + \sin\theta_r\ \vec{e}_z, \\ \vec{e}_{i,t} &= -\vec{h}_{i,t} = \cos\theta_{i,t}\ \vec{e}_x - \sin\ \theta_{i,t}\ \vec{e}_z, \\ \vec{e}_r &= -\vec{h}_r = -\cos\theta_r\ \vec{e}_x - \sin\theta_r\ \vec{e}_z. \end{aligned}$$

2.2.1 Boundary Conditions and Snell's law

From 6.013, we know that Stoke's and Gauss' Law for the electric and magnetic fields require constraints on some of the field components at media boundaries. In the absence of surface currents and charges, the tangential electric and magnetic fields as well as the normal dielectric and magnetic fluxes have to be continuous when going from medium 1 into medium 2 for all times at each point along the surface, i.e. $z = 0$

$$E/H_{i,x/y}e^{\mathrm{j}(\omega t - k_{i,x}x)} + E/H_{r,x/y}e^{\mathrm{j}(\omega t - k_{r,x}x)} = E/H_{i,x/y}e^{\mathrm{j}(\omega t - k_{t,x}x)}. \qquad (2.103)$$

This equation can only be fulfilled at all times if and only if the x-component of the k-vectors for the reflected and transmitted wave are equal to (match)

the corresponding component of the incident wave

$$k_{i,x} = k_{r,x} = k_{t,x} \tag{2.104}$$

This phase matching condition is shown in Fig. 2.18 for the case $\sqrt{\epsilon_2\mu_2} > \sqrt{\epsilon_1\mu_1}$ *or* $k_t > k_i$.

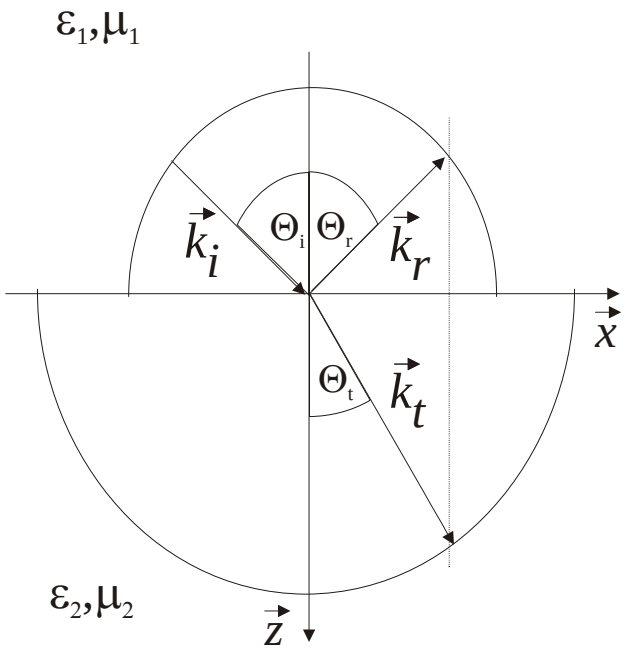

Figure 2.18: Phase matching condition for reflected and transmitted wave

The phase matching condition Eq(2.104) results in $\theta_r = \theta_i = \theta_1$ and Snell's law for the angle $\theta_t = \theta_2$ of the transmitted wave

$$\sin\theta_t = \frac{\sqrt{\epsilon_1\mu_1}}{\sqrt{\epsilon_2\mu_2}}\sin\theta_i \tag{2.105}$$

or for the case of non magnetic media with $\mu_1 = \mu_2 = \mu_0$

$$\sin\theta_t = \frac{n_1}{n_2}\sin\theta_i \tag{2.106}$$

2.2.2 Measuring Refractive Index with Minimum Deviation

Snell's law can be used for measuring the refractive index of materials. Consider a prism prepared from a material with unknown refractive index $n(\lambda)$, see Fig. 2.19 (a).

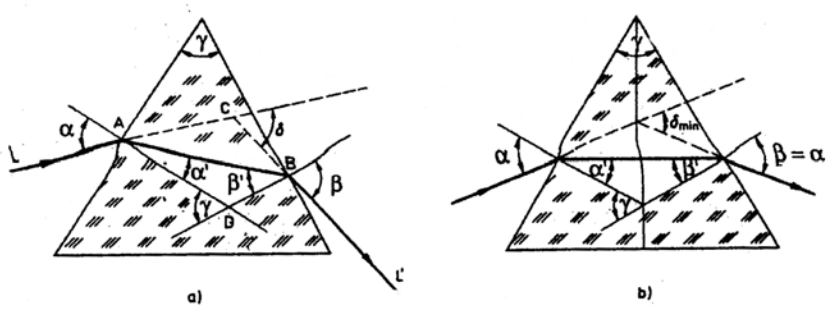

Figure 2.19: (a) Beam propagating through a prism. (b) For the case of minimum deviation [3] p. 65.

The prism is mounted on a rotation stage as shown in Fig. 2.20. The angle of incidence α is then varied with a fixed incident beam path and the transmitted light is observed on a screen. If one starts of with normal incidence on the first prism surface one notices that after turning the prism one goes through a minium for the deflection angle of the beam. This becomes obvious from Fig. 2.19 (b). There is an angle of incidence α where the beam path through the prism is symmetric. If the input angle is varied around this point, it would be identical to exchange the input and output beams. From that we conclude that the deviation δ must go through an extremum at the symmetry point, see Figure 2.21. It can be shown (Recitations), that the refractive index is then determined by

$$n = \frac{\sin\frac{\alpha(\delta_{\min})+\delta_{\min}}{2}}{\sin\frac{\alpha(\delta_{\min})}{2}}. \tag{2.107}$$

If the measurement is repeated for various wavelength of the incident radiation the complete wavelength dependent refractive index is characterized, see for example, Fig. 2.22.

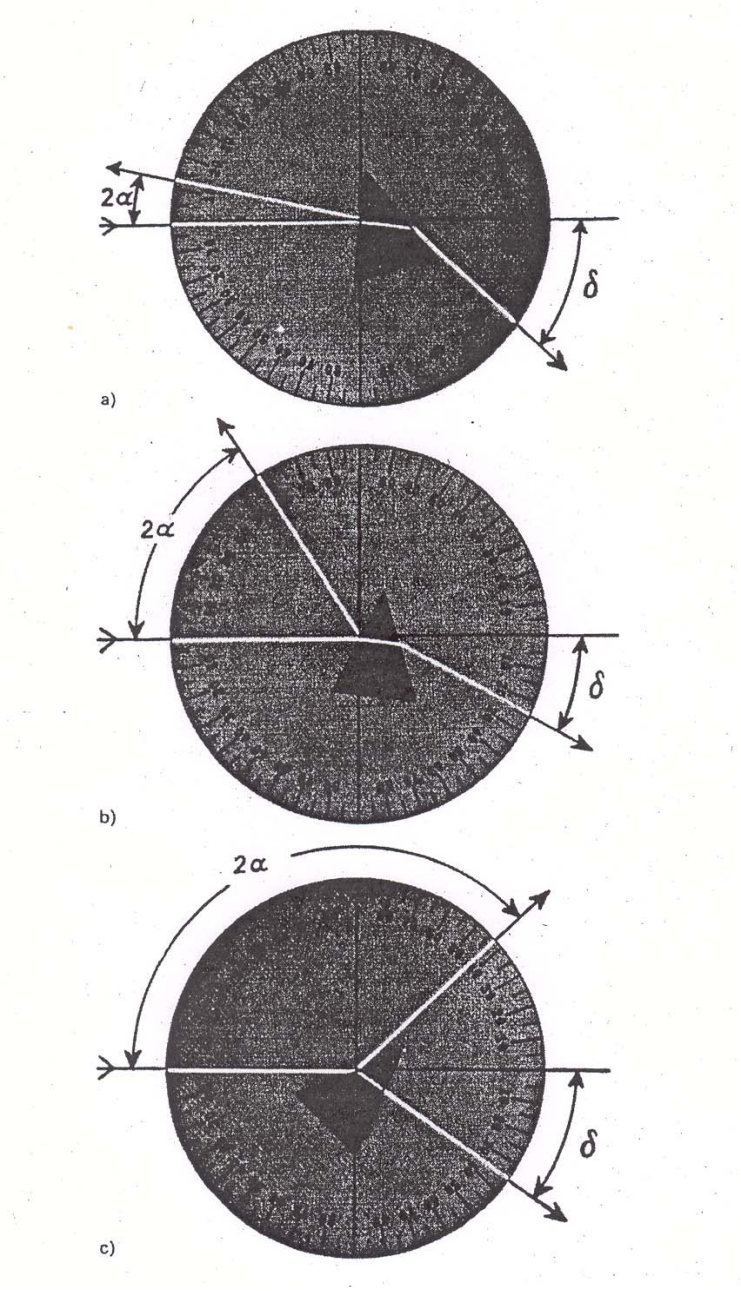

Figure 2.20: Refraction of a Prism with n=1.731 for different angles of incidence alpha. The angle of incidence is stepwise increased by rotating the prism clockwise. The angle of transmission first increases. After the angle for minimum deviation is reached the transmission angle starts to decrease [3] p67.

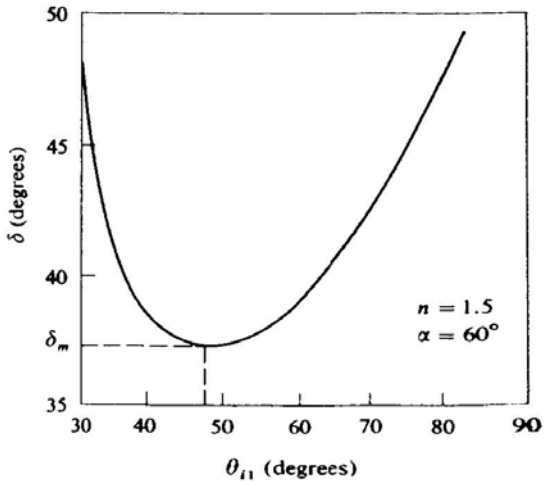

Figure 2.21: Deviation versus incident angel [1]

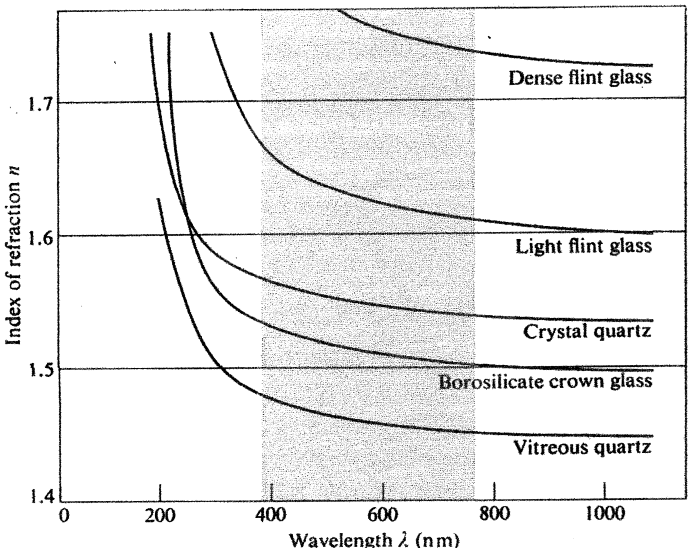

Figure 2.22: Refractive index as a function of wavelength for various media transmissive in the visible [1], p42.

2.2.3 Fresnel Reflection

After understanding the direction of the reflected and transmitted light, formulas for how much light is reflected and transmitted are derived by evaluating the boundary conditions for the TE and TM-wave. According to Eqs.(2.103) and (2.104) we obtain for the continuity of the tangential E and H fields:

TE-wave (s-pol.)	TM-wave (p-pol.)
$\underline{E}_i + \underline{E}_r = \underline{E}_t$	$\underline{E}_i \cos\theta_i - \underline{E}_r \cos\theta_r = \underline{E}_t \cos\theta_t$
$\underline{H}_i \cos\theta_i - \underline{H}_r \cos\theta_r = \underline{H}_t \cos\theta_t$	$\underline{H}_i + \underline{H}_r = \underline{H}_t$

(2.108)

Introducing the characteristic impedances in both half spaces $\underline{Z}_{1/2} = \sqrt{\frac{\mu_0\mu_{1/2}}{\epsilon_0\epsilon_{1/2}}}$, and the impedances that relate the tangential electric and magnetic field components $\underline{Z}_{1/2}^{TE/TM}$ in both half spaces the boundary conditions can be rewritten in terms of the electric or magnetic field components.

TE-wave (s-pol.)	TM-wave (p-pol.)
$\underline{Z}_{1/2}^{TE} = \frac{\underline{E}_{i/t}}{\underline{H}_{i/t} \cos\theta_{i/t}} = \frac{\underline{Z}_{1/2}}{\cos\theta_{1/2}}$	$\underline{Z}_{1/2}^{TM} = \frac{\underline{E}_{i/t} \cos\theta_{i/t}}{\underline{H}_{i/t}} = \underline{Z}_{1/2} \cos\theta_{1/2}$
$\underline{E}_i + \underline{E}_r = \underline{E}_t$	$\underline{H}_i - \underline{H}_r = \frac{\underline{Z}_2^{TM}}{\underline{Z}_1^{TM}} \underline{H}_t$
$\underline{E}_i - \underline{E}_r = \frac{\underline{Z}_1^{TE}}{\underline{Z}_2^{TE}} \underline{E}_t$	$\underline{H}_i + \underline{H}_r = \underline{H}_t$

(2.109)

Amplitude Reflection and Transmission coefficients

From these equations we can easily solve for the reflected and transmitted wave amplitudes in terms of the incident wave amplitudes. By dividing both equations by the incident wave amplitudes we obtain for the amplitude reflection and transmission coeffcients. Note, that reflection and transmission coefficients are defined in terms of the electric fields for the TE-wave and in terms of the magnetic fields for the TM-wave.

TE-wave (s-pol.)	TM-wave (p-pol.)
$\underline{r}^{TE} = \frac{\underline{E}_r}{\underline{E}_i}; \underline{t}^{TE} = \frac{\underline{E}_t}{\underline{E}_i}$	$\underline{r}^{TM} = \frac{\underline{H}_r}{\underline{H}_i}; \underline{t}^{TM} = \frac{\underline{H}_t}{\underline{H}_i}$
$1 + \underline{r}^{TE} = \underline{t}^{TE}$	$1 - \underline{r}^{TM} = \frac{\underline{Z}_2^{TM}}{\underline{Z}_1^{TM}} \underline{t}^{TM}$
$1 - \underline{r}^{TE} = \frac{\underline{Z}_1^{TE}}{\underline{Z}_2^{TE}} \underline{t}^{TE}$	$1 + \underline{r}^{TM} = \underline{t}^{TM}$

(2.110)

or in both cases the amplitude transmission and reflection coefficients are

$$\underline{t}^{TE/TM} = \frac{2}{1+\frac{\underline{Z}_{1/2}^{TE/TM}}{\underline{Z}_{2/1}^{TE/TM}}} = \frac{2\underline{Z}_{2/1}^{TE/TM}}{\underline{Z}_{1}^{TE/TM}+\underline{Z}_{2}^{TE/TM}} \tag{2.111}$$

$$\underline{r}^{TE/TM} = \frac{\underline{Z}_{2/1}^{TE/TM}-\underline{Z}_{1/2}^{TE/TM}}{\underline{Z}_{1}^{TE/TM}+\underline{Z}_{2}^{TE/TM}} \tag{2.112}$$

Despite the simplicity of these formulas, they describe already an enormous wealth of phenomena. To get some insight, consider the case of purely dielectric and lossless media characterized by its real refractive indices n_1 and n_2. Then Eqs.(2.111) and (2.112) simplify for the TE and TM case to

TE-wave (s-pol.)	TM-wave (p-pol.)
$\underline{Z}_{1/2}^{TE} = \frac{Z_{1/2}}{\cos\theta_{1/2}} = \frac{Z_0}{n_{1/2}\cos\theta_{1/2}}$	$\underline{Z}_{1/2}^{TM} = \underline{Z}_{1/2}\cos\theta_{1/2} = \frac{Z_0}{n_{1/2}}\cos\theta_{1/2}$
$\underline{r}^{TE} = \frac{n_1\cos\theta_1 - n_2\cos\theta_2}{n_1\cos\theta_1 + n_2\cos\theta_2}$	$\underline{r}^{TM} = \frac{\frac{n_2}{\cos\theta_2}-\frac{n_1}{\cos\theta_1}}{\frac{n_2}{\cos\theta_2}+\frac{n_1}{\cos\theta_1}}$
$\underline{t}^{TE} = \frac{2n_1\cos\theta_1}{n_1\cos\theta_1 + n_2\cos\theta_2}$	$\underline{t}^{TM} = \frac{2\frac{n_2}{\cos\theta_2}}{\frac{n_2}{\cos\theta_2}+\frac{n_1}{\cos\theta_1}}$

(2.113)

Figure 2.23 shows the evaluation of Eqs.(2.113) for the case of a reflection at the interface of air and glass with $n_2 > n_1$ and ($n_1 = 1$, $n_2 = 1.5$).

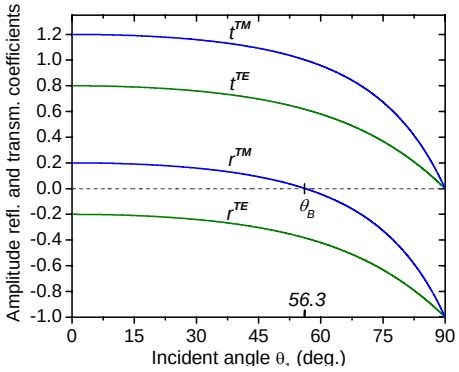

Figure 2.23: The amplitude coefficients of reflection and transmission as a function of incident angle. These correspond to external reflection $n_2 > n_1$ at an air-glas interface ($n_1 = 1$, $n_2 = 1.5$).

For TE-polarized light the reflected light changes sign with respect to the incident light (reflection at the optically more dense medium). This is not so for TM-polarized light under close to normal incidence. It occurs only for angles larger than θ_B, which is called the Brewster angle. So for TM-polarized light the amplitude reflection coefficient is zero at the Brewster angle. This phenomena will be discussed in more detail later.

This behavior changes drastically if we consider the opposite arrangement of media, i.e. we consider the glass-air interface with $n_1 > n_2$, see Figure 2.24. Then the TM-polarized light experiences a π-phase shift upon reflection close to normal incidence. For increasing angle of incidence this reflection coefficient goes through zero at the Brewster angle $\theta_B^{'}$ different from before. However, for large enough angle of incidence the reflection coefficient reaches magnitude 1 and stays there. This phenomenon is called total internal reflection and the angle where this occurs first is the critical angle for total internal reflection, θ_{tot}. Total internal reflection will be discussed in more detail later.

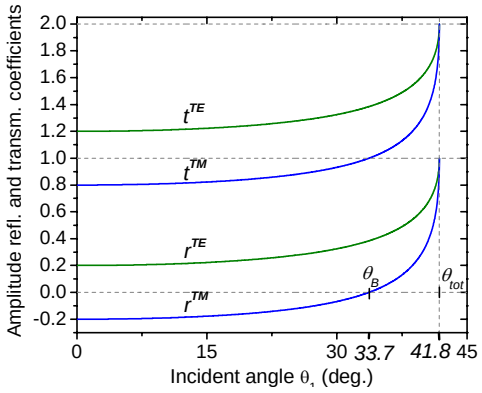

Figure 2.24: The amplitude coefficients of reflection and transmission as a function of incident angle. These correspond to internal reflection $n_1 > n_2$ at a glas-air interface ($n_1 = 1.5$, $n_2 = 1$).

Power reflection and transmission coefficients

Often we are not interested in the amplitude but rather in the optical power reflected or transmitted in a beam of finite size, see Figure 2.25.

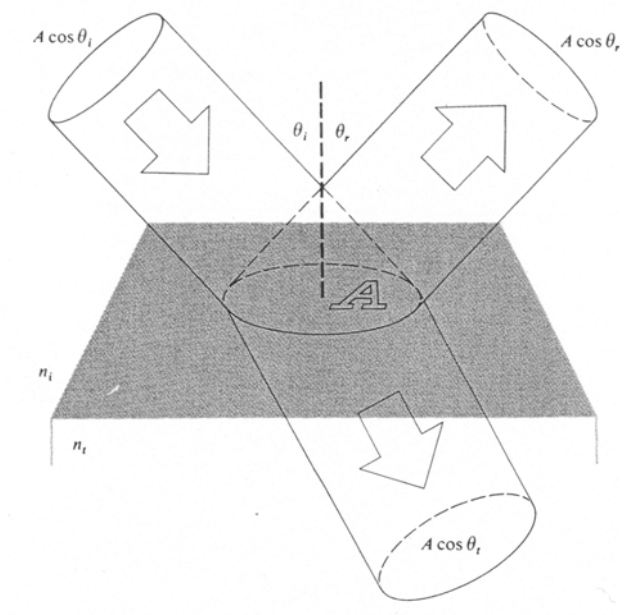

Figure 2.25: Reflection and transmission of an incident beam of finite size [1].

Note, that to get the power in a beam of finite size, we need to integrated the corresponding Poynting vector over the beam area, which means multiplication by the beam crosssectional area for a homogenous beam. Since the angle of incidence and reflection are equal, $\theta_i = \theta_r = \theta_1$ this beam crosssectional area drops out in reflection

$$R^{TE/TM} = \frac{I_r^{TE/TM} A\cos\theta_i}{I_i^{TE/TM} A\cos\theta_r} = \left|\underline{r}^{TE/TM}\right|^2 = \left|\frac{\underline{Z}_2^{TE/TM} - \underline{Z}_1^{TE/TM}}{\underline{Z}_1^{TE/TM} + \underline{Z}_2^{TE/TM}}\right|^2 \quad (2.114)$$

However, due to the different angles for the incident and the transmitted beam $\theta_t = \theta_2 \neq \theta_1$, we arrive at

$$\begin{aligned} T^{TE/TM} &= \frac{I_t^{TE/TM} A\cos\theta_t}{I_i^{TE/TM} A\cos\theta_r} \\ &= \frac{\cos\theta_2}{\cos\theta_1} \mathrm{Re}\left\{\frac{1}{\underline{Z}_{2/1}}\right\} \mathrm{Re}\left\{\frac{1}{\underline{Z}_{1/2}}\right\}^{-1} \left|\underline{t}^{TE/TM}\right|^2. \end{aligned} \quad (2.115)$$

Using in the case of TE-polarization $\frac{\underline{Z}_{1/2}}{\cos\theta_{1/2}} = \underline{Z}_{1/2}^{TE}$ and analogously for TM-polarization $\underline{Z}_{1/2}\cos\theta_{1/2} = \underline{Z}_{1/2}^{TM}$, we obtain

$$T^{TE/TM} = \mathrm{Re}\left\{\frac{1}{\underline{Z}_{1/2}^{TE/TM}}\right\}^{-1} \mathrm{Re}\left\{\frac{4\underline{Z}_{2/1}^{TE/TM}}{\left|\underline{Z}_{1}^{TE/TM} + \underline{Z}_{2}^{TE/TM}\right|^2}\right\} \tag{2.116}$$

Note, for the case where the characteristic impedances are complex this can not be further simplified. If the characteristic impedances are real, i.e. the media are lossless, the transmission coefficient simplifies to

$$T^{TE/TM} = \frac{4Z_{1/2}^{TE/TM} Z_{2/1}^{TE/TM}}{\left(Z_{1}^{TE/TM} + Z_{2}^{TE/TM}\right)^2}. \tag{2.117}$$

To summarize for lossless media the power reflection and transmission coefficients are

TE-wave (s-pol.)	TM-wave (p-pol.)
$Z_{1/2}^{TE} = \frac{Z_{1/2}}{\cos\theta_{1/2}} = \frac{Z_0}{n_{1/2}\cos\theta_{1/2}}$	$Z_{1/2}^{TM} = Z_{1/2}\cos\theta_{1/2} = \frac{Z_0}{n_{1/2}}\cos\theta_{1/2}$
$R^{TE} = \left\lvert\frac{n_2\cos\theta_2 - n_1\cos\theta_1}{n_1\cos\theta_1 + n_2\cos\theta_2}\right\rvert^2$	$R^{TM} = \left\lvert\frac{\frac{n_2}{\cos\theta_2} - \frac{n_1}{\cos\theta_1}}{\frac{n_2}{\cos\theta_2} + \frac{n_1}{\cos\theta_1}}\right\rvert^2$
$T^{TE} = \frac{4n_1\cos\theta_1 n_2\cos\theta_2}{\lvert n_1\cos\theta_1 + n_2\cos\theta_2\rvert^2}$	$T^{TM} = \frac{4\frac{n_2}{\cos\theta_2}\frac{n_1}{\cos\theta_1}}{\left\lvert\frac{n_2}{\cos\theta_2} + \frac{n_1}{\cos\theta_1}\right\rvert^2}$
$T^{TE} + R^{TE} = 1$	$T^{TM} + R^{TM} = 1$

(2.118)

A few phenomena that occur upon reflection at surfaces between different media are especially noteworthy and need a more indepth discussion because they enhance or enable the construction of many optical components and devices.

2.2.4 Brewster's Angle

As Figures 2.23 and 2.24 already show, for light polarized parallel to the plane of incidence, p-polarized light, the reflection coefficient vanishes at a given angle θ_B, called the Brewster angle. Using Snell's Law Eq.(2.106),

$$\frac{n_2}{n_1} = \frac{\sin\theta_1}{\sin\theta_2}, \tag{2.119}$$

we can rewrite the reflection and transmission coefficients in Eq.(2.118) only in terms of the angles. For example, we find for the reflection coefficient

$$R^{TM} = \left| \frac{\frac{n_2}{n_1} - \frac{\cos\theta_2}{\cos\theta_1}}{\frac{n_2}{n_1} + \frac{\cos\theta_2}{\cos\theta_1}} \right|^2 = \left| \frac{\frac{\sin\theta_1}{\sin\theta_2} - \frac{\cos\theta_2}{\cos\theta_1}}{\frac{\sin\theta_1}{\sin\theta_2} + \frac{\cos\theta_2}{\cos\theta_1}} \right|^2 \tag{2.120}$$

$$= \left| \frac{\sin 2\theta_1 - \sin 2\theta_2}{\sin 2\theta_1 + \sin 2\theta_2} \right|^2 \tag{2.121}$$

where we used in the last step in addition the relation $\sin 2\alpha = 2\sin\alpha\cos\alpha$. Thus by forcing $R^{TM} = 0$, the Brewster angle is reached for

$$\sin 2\theta_{1,B} - \sin 2\theta_{2,B} = 0 \tag{2.122}$$

or

$$2\theta_{1,B} = \pi - 2\theta_{2,B} \text{ or } \theta_{1,B} + \theta_{2,B} = \frac{\pi}{2} \tag{2.123}$$

This relation is illustrated in Figure 2.26. The reflected and transmitted beams are orthogonal to each other, so that the dipoles induced in the medium by the transmitted beam, shown as arrows in Fig. 2.26, can not radiate into the direction of the reflected beam. This is the physical origin of the zero in the reflection coefficient, only possible for a p-polarized or TM-wave.

The relation (2.123) can be used to express the Brewster angle as a function of the refractive indices, because if we substitute (2.123) into Snell's law we obtain

$$\begin{aligned} \frac{\sin\theta_1}{\sin\theta_2} &= \frac{n_2}{n_1} \\ \frac{\sin\theta_{1,B}}{\sin\left(\frac{\pi}{2} - \theta_{1,B}\right)} &= \frac{\sin\theta_{1,B}}{\cos\theta_{1,B}} = \tan\theta_{1,B}, \end{aligned}$$

or

$$\tan\theta_{1,B} = \frac{n_2}{n_1}. \tag{2.124}$$

Using the Brewster angle condition one can insert an optical component with a refractive index $n \neq 1$ into a TM-polarized beam in air without having reflections, see Figure 2.27. Note, this is not possible for a TE-polarized beam.

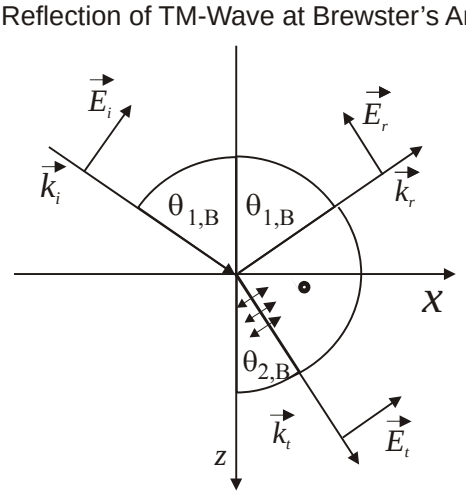

Figure 2.26: Conditions for reflection of a TM-Wave at Brewster's angle. The reflected and transmitted beams are orthogonal to each other, so that the dipoles excited in the medium by the transmitted beam can not radiate into the direction of the reflected beam.

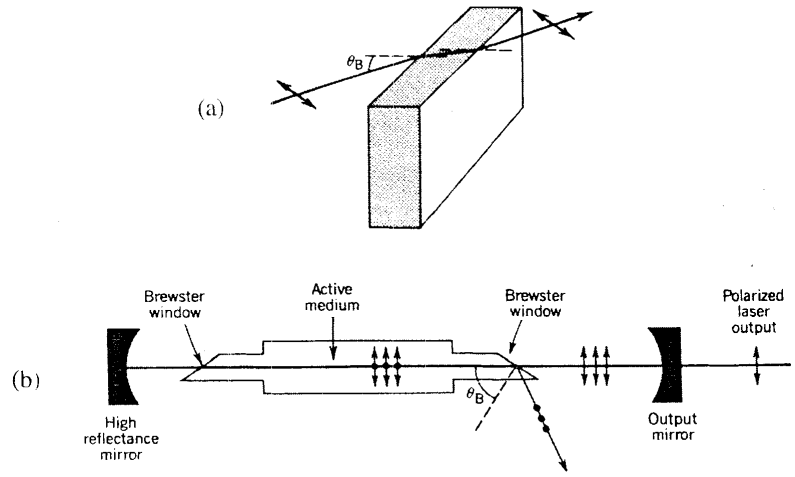

Figure 2.27: A plate under Brewster's angle does not reflect TM-light. The plate can be used as a window to introduce gas filled tubes into a laser beam without insertion loss (ideally), [6] p. 209.

2.2.5 Total Internal Reflection

Another striking phenomenon, see Figure 2.24, occurs for the case where the beam hits the surface from the side of the optically denser medium, i.e. $n_1 > n_2$. There is obviously a critical angle of incidence, beyond which all light is reflected. How can that occur? This is easy to understand from the phase matching diagram at the surface, see Figure 2.18, which is redrawn for this case in Figure 2.28.

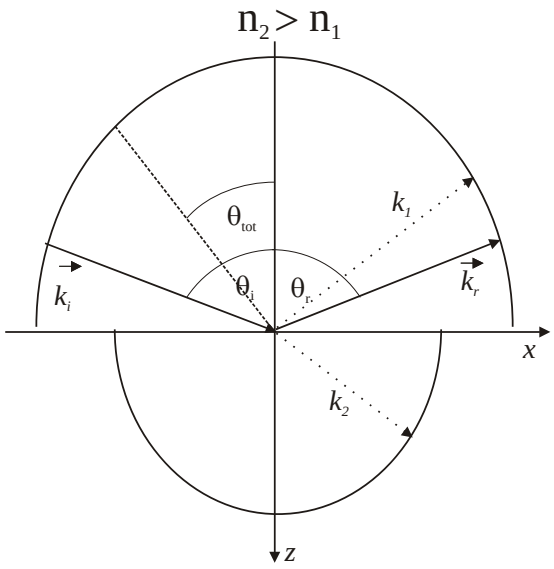

Figure 2.28: Phase matching diagram for total internal reflection.

There is no real wavenumber in medium 2 possible as soon as the angle of incidence becomes larger than the critical angle for total internal reflection

$$\theta_i > \theta_{tot} \tag{2.125}$$

with

$$\sin\theta_{tot} = \frac{n_2}{n_1}. \tag{2.126}$$

Figure 2.29 shows the angle of refraction and incidence for the two cases of external and internal reflection, when the angle of incidence approaches the critical angle.

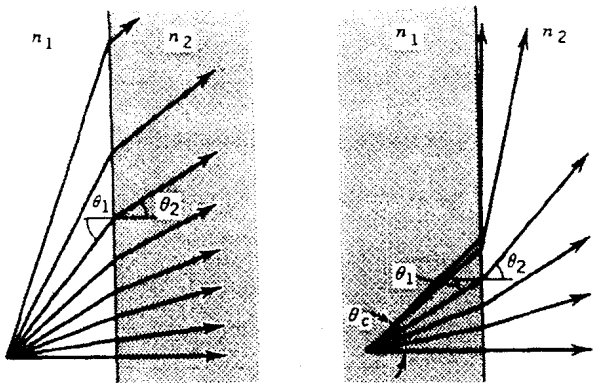

Figure 2.29: Relation between angle of refraction and incidence for external refraction and internal refraction ([6], p. 11).

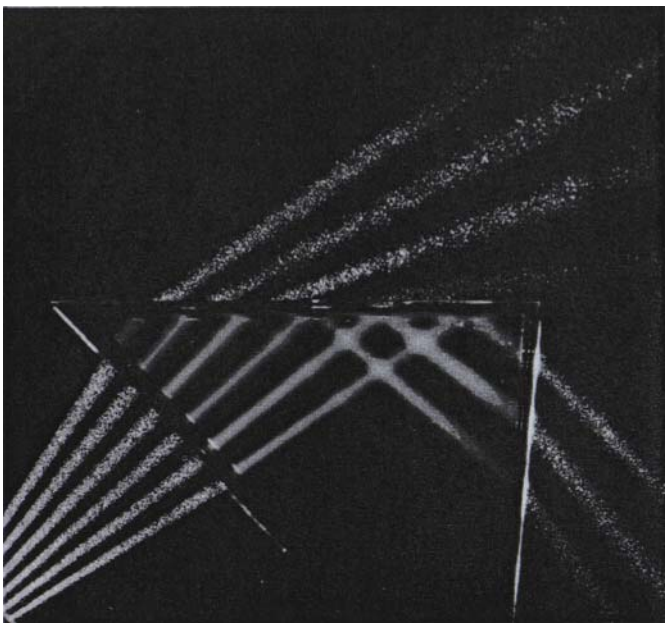

Figure 2.30: Relation between angle of refraction and incidence for external refraction and internal refraction ([1], p. 81).

Total internal reflection enables broadband reflectors. Figure 2.30 shows

again what happens when the critical angle of reflection is surpassed. Figure 2.31 shows how total internal reflection can be used to guide light via reflection at a prism or by multiple reflections in a waveguide.

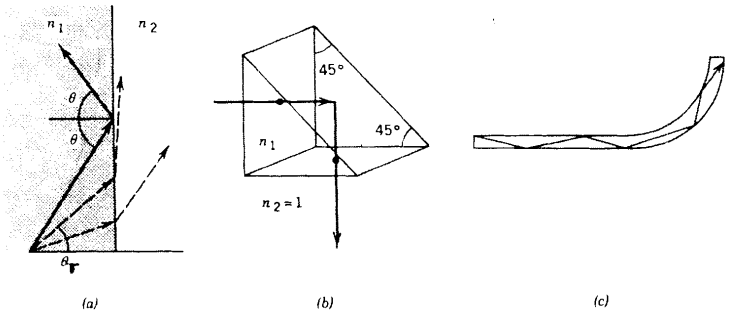

Figure 2.31: (a) Total internal reflection, (b) internal reflection in a prism, (c) Rays are guided by total internal reflection from the internal surface of an optical fiber ([6] p. 11).

Figure 2.32 shows the realization of a retro reflector, which always returns a parallel beam independent of the orientation of the prism (in fact the prism can be a real 3D-corner so that the beam is reflected parallel independent from the precise orientation of the corner cube). A surface patterned by little corner cubes constitute a "cats eye" used on traffic signs.

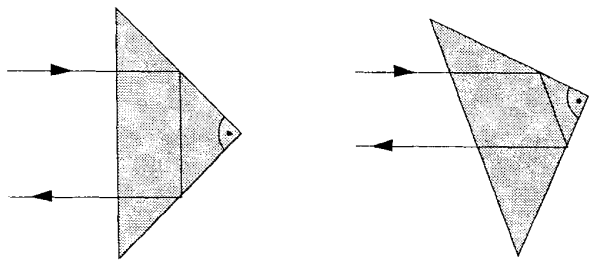

Figure 2.32: Total internal reflection in a retro reflector.

More on reflecting prisms and its use can be found in [1], pages 131-136.

Evanescent Waves

What is the field in medium 2 when total internal reflection occurs? Is it identical to zero? It turns out phase matching can still occur if the propagation constant in z-direction becomes imaginary, $k_{2z} = -\mathrm{j}\kappa_{2z}$, because then we can fulfill the wave equation in medium 2. This is equivalent to the dispersion relation

$$k_{2x}^2 + k_{2z}^2 = k_2^2,$$

or with $k_{2x} = k_{1x} = k_1 \sin\theta_1$, we obtain for the imaginary wavenumber

$$\begin{aligned} \kappa_{2z} &= \sqrt{k_1^2 \sin^2\theta_1 - k_2^2}, && (2.127) \\ &= k_1 \sqrt{\sin^2\theta_1 - \sin^2\theta_{tot}}. && (2.128) \end{aligned}$$

The electric field in medium 2 is then, for the example for a TE-wave, given by

$$\begin{aligned} \underline{\vec{E}}_t &= \underline{E}_t \, \vec{e}_y \, e^{\mathrm{j}(\omega t - \vec{k}_t \cdot \vec{r})}, && (2.129) \\ & \underline{E}_t \, \vec{e}_y \, e^{\mathrm{j}(\omega t - k_{2,x} x)} e^{-\kappa_{2z} z}. && (2.130) \end{aligned}$$

Thus the wave penetrates into medium 2 exponentially with a 1/e-depth δ, given by

$$\delta = \frac{1}{\kappa_{2z}} = \frac{1}{k_1 \sqrt{\sin^2\theta_1 - \sin^2\theta_{tot}}} \qquad (2.131)$$

Figure 2.33 shows the penetration depth as a function of angle of incidence for a silica/air interface and a silicon/air interface. The figure demonstrates that light from inside a semiconductor material with a relatively high index around n=3.5 is mostly captured in the semiconductor material (Problem of light extraction from light emitting diodes (LEDs)), see problem set 2.

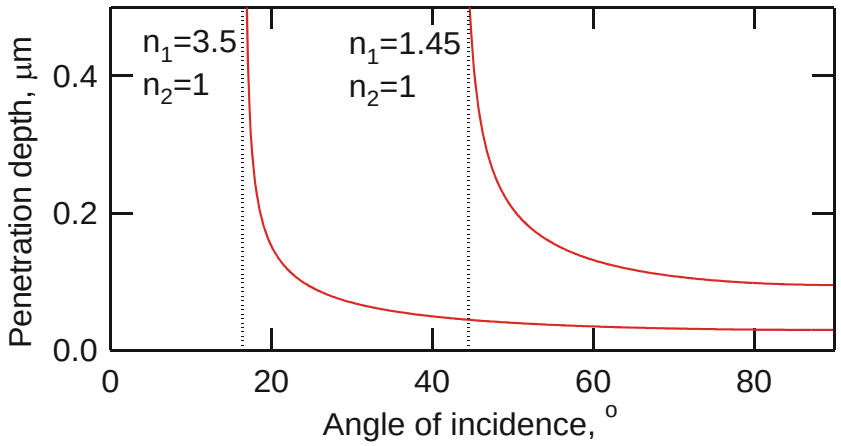

Figure 2.33: Penetration depth for total internal reflection at a silica/air and a silicon/air interface for $\lambda = 0.633$nm.

As the magnitude of the reflection coefficient is 1 for total internal reflection, the power flowing into medium 2 must vanish, i.e. the transmission is zero. Note, that the transmission and reflection coefficients in Eq.(2.113) can be used beyond the critical angle for total internal reflection. We only have to be aware that the electric field in medium 2 has an imaginary dependence in the exponent for the z-direction, i.e. $k_{2z} = k_2 \cos\theta_2 = -j\kappa_{2z}$. Thus $\cos\theta_2$ in all formulas for the reflection and transmission coefficients has to be replaced by the imaginary number

$$\begin{aligned} \cos\theta_2 &= \frac{k_{2z}}{k_2} = -\mathrm{j}\frac{k_1}{k_2}\sqrt{\sin^2\theta_1 - \sin^2\theta_{tot}} \\ &= -\mathrm{j}\frac{n_1}{n_2}\sqrt{\sin^2\theta_1 - \sin^2\theta_{tot}} \\ &= -\mathrm{j}\sqrt{\left(\frac{\sin\theta_1}{\sin\theta_{tot}}\right)^2 - 1}. \end{aligned} \tag{2.132}$$

Then the reflection coefficients in Eq.(2.113) change to all-pass functions

TE-wave (s-pol.)	TM-wave (p-pol.)
$\underline{r}^{TE} = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2}$	$\underline{r}^{TM} = \frac{\frac{n_2}{\cos\theta_2} - \frac{n_1}{\cos\theta_1}}{\frac{n_2}{\cos\theta_2} + \frac{n_1}{\cos\theta_1}}$
$\underline{r}^{TE} = \frac{\cos\theta_1 + \mathrm{j}\frac{n_2}{n_1}\sqrt{\left(\frac{\sin\theta_1}{\sin\theta_{tot}}\right)^2 - 1}}{\cos\theta_1 - \mathrm{j}\frac{n_2}{n_1}\sqrt{\left(\frac{\sin\theta_1}{\sin\theta_{tot}}\right)^2 - 1}}$	$\underline{r}^{TM} = \frac{\cos\theta_1 + \mathrm{j}\frac{n_1}{n_2}\sqrt{\left(\frac{\sin\theta_1}{\sin\theta_{tot}}\right)^2 - 1}}{\cos\theta_1 - \mathrm{j}\frac{n_1}{n_2}\sqrt{\left(\frac{\sin\theta_1}{\sin\theta_{tot}}\right)^2 - 1}}$
$\tan\frac{\phi^{TE}}{2} = \frac{1}{\cos\theta_1}\frac{n_2}{n_1}\sqrt{\left(\frac{\sin\theta_1}{\sin\theta_{tot}}\right)^2 - 1}$	$\tan\frac{\phi^{TM}}{2} = \frac{1}{\cos\theta_1}\frac{n_1}{n_2}\sqrt{\left(\frac{\sin\theta_1}{\sin\theta_{tot}}\right)^2 - 1}$

(2.133)

Thus the magnitude of the reflection coefficient is 1. However, there is a non-vanishing phase shift for the light field upon total internal reflection, denoted as ϕ^{TE} and ϕ^{TM} in the table above. Figure 2.34 shows these phase shifts for the glass/air interface and for both polarizations.

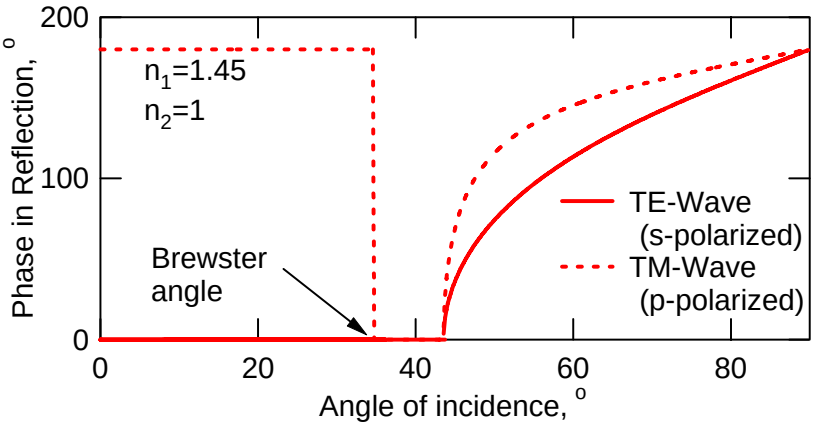

Figure 2.34: Phase shifts for TE- and TM- wave upon reflection from a silica/air interface, with $n_1 = 1.45$ and $n_2 = 1$.

Goos-Haenchen-Shift

So far, we looked only at plane waves undergoing reflection at surface due to total internal reflection. If a beam of finite transverse size is reflected from

such a surface it turns out that it gets displaced by a distance Δz, see Figure 2.35 (a), called Goos-Haenchen-Shift.

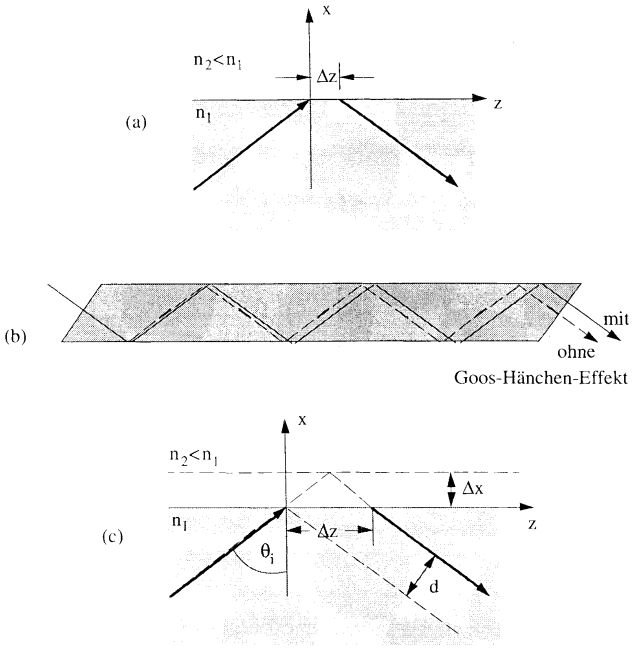

Figure 2.35: (a) Goos-Haenchen Shift and related beam displacement upon reflection of a beam with finite size; (b) Accumulation of phase shifts in a waveguide.

Detailed calculations show (problem set 2), that the displacement is given by

$$\Delta z = 2\delta^{TE/TM} \tan\theta_1, \tag{2.134}$$

as if the beam was reflected at a virtual layer with depth $\delta^{TE/TM}$ into medium 2. It turns out, that for TE-waves

$$\delta^{TE} = \delta, \tag{2.135}$$

where δ is the penetration depth according to Eq.(2.131) for evanescent waves. But for TM-waves

$$\delta^{TM} = \frac{\delta}{\left[1 + \left(\frac{n_1}{n_2}\right)^2\right] \sin^2\theta_1 - 1} \tag{2.136}$$

These shifts accumulate when the beam is propagating in a waveguide, see Figure 2.35 (b) and is important to understand the dispersion relations of waveguide modes. The Goos-Haenchen shift can be observed by reflection at a prism partially coated with a silver film, see Figure 2.36. The part reflected from the silver film is shifted with respect to the beam reflected due to total internal reflection, as shown in the figure.

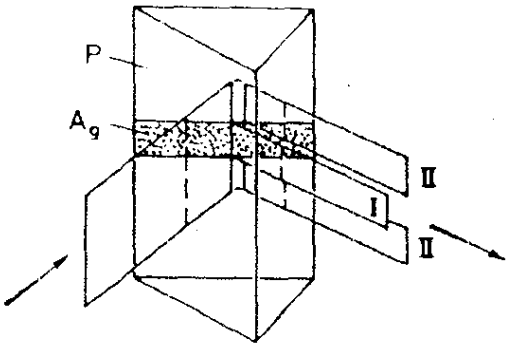

Figure 2.36: Experimental proof of the Goos-Haenchen shift by total internal reflection at a prism, that is partially coated with silver, where the penetration of light can be neglected. [3] p. 486.

Frustrated total internal reflection

Another proof for the penetration of light into medium 2 in the case of total internal reflection can be achieved by putting two prisms, where total internal reflection occurs back to back, see Figure 2.37. Then part of the light, depending on the distance between the two interfaces, is converted back into a propagating wave that can leave the second prism. This effect is called frustrated internal reflection and it can be used as a beam splitter as shown in Figure 2.37.

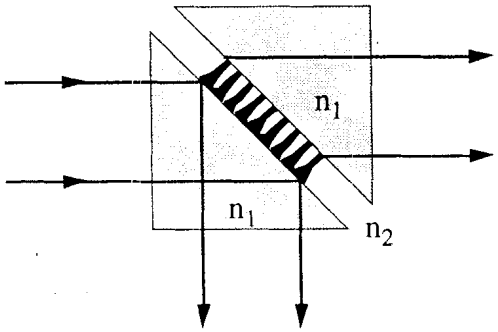

Figure 2.37: Frustrated total internal reflection. Part of the light is picked up by the second surface and converted into a propagating wave.

2.3 Mirrors, Interferometers and Thin-Film Structures

One of the most striking wave phenomena is interference. Many optical devices are based on the concept of interfering waves, such as low loss dielectric mirrors and interferometers and other thin-film optical coatings. After having a quick look into the phenomenon of interference, we will develope a powerful matrix formalism that enables us to evaluate efficiently many optical (also microwave) systems based on interference.

2.3.1 Interference and Coherence

Interference

Interference of waves is a consequence of the linearity of the wave equation (2.13). If we have two individual solutions of the wave equation

$$\vec{E}_1(\vec{r},t) = E_1 \cos(\omega_1 t - \vec{k}_1 \cdot \vec{r} + \varphi_1)\ \vec{e}_1, \tag{2.137}$$

$$\vec{E}_2(\vec{r},t) = E_2 \cos(\omega_2 t - \vec{k}_1 \cdot \vec{r} + \varphi_2)\ \vec{e}_2, \tag{2.138}$$

with arbitrary amplitudes, wave vectors and polarizations, the sum of the two fields (superposition) is again a solution of the wave equation

$$\vec{E}(\vec{r},t) = \vec{E}_1(\vec{r},t) + \vec{E}_2(\vec{r},t). \tag{2.139}$$

If we look at the intensity, wich is proportional to the amplitude square of the total field

$$\vec{E}(\vec{r},t)^2 = \left(\vec{E}_1(\vec{r},t) + \vec{E}_2(\vec{r},t)\right)^2, \tag{2.140}$$

we find

$$\vec{E}(\vec{r},t)^2 = \vec{E}_1(\vec{r},t)^2 + \vec{E}_2(\vec{r},t)^2 + 2\vec{E}_1(\vec{r},t)\cdot\vec{E}_2(\vec{r},t) \tag{2.141}$$

with

$$\vec{E}_1(\vec{r},t)^2 = \frac{E_1^2}{2}\left(1+\cos 2(\omega_1 t - \vec{k}_1\cdot\vec{r}+\varphi_1)\right), \tag{2.142}$$

$$\vec{E}_2(\vec{r},t)^2 = \frac{E_2^2}{2}\left(1+\cos 2(\omega_2 t - \vec{k}_2\cdot\vec{r}+\varphi_2)\right), \tag{2.143}$$

$$\begin{aligned}\vec{E}_1(\vec{r},t)\cdot\vec{E}_2(\vec{r},t) &= (\vec{e}_1\cdot\vec{e}_2)\,E_1E_2\cos(\omega_1 t - \vec{k}_1\cdot\vec{r}+\varphi_1)\cdot \\ &\quad\cdot\cos(\omega_2 t - \vec{k}_2\cdot\vec{r}+\varphi_2)\end{aligned} \tag{2.144}$$

$$\vec{E}_1(\vec{r},t)\cdot\vec{E}_2(\vec{r},t) = \frac{1}{2}(\vec{e}_1\cdot\vec{e}_2)\,E_1E_2\cdot \tag{2.145}$$

$$\cdot\left[\begin{array}{l}\cos\left((\omega_1-\omega_2)\,t-\left(\vec{k}_1-\vec{k}_2\right)\cdot\vec{r}+(\varphi_1-\varphi_2)\right) \\ +\cos\left((\omega_1+\omega_2)\,t-\left(\vec{k}_1+\vec{k}_2\right)\cdot\vec{r}+(\varphi_1+\varphi_2)\right)\end{array}\right] \tag{2.146}$$

Since at optical frequencies neither our eyes nor photo detectors, can ever follow the optical frequency itself and certainly not twice as large frequencies, we drop the rapidly oscillating terms. Or in other words we look only on the cycle-averaged intensity, which we denote by a bar

$$\begin{aligned}\overline{\vec{E}(\vec{r},t)^2} &= \frac{E_1^2}{2}+\frac{E_2^2}{2}+(\vec{e}_1\cdot\vec{e}_2)\,E_1E_2\cdot \\ &\quad\cdot\cos\left((\omega_1-\omega_2)\,t-\left(\vec{k}_1-\vec{k}_2\right)\cdot\vec{r}+(\varphi_1-\varphi_2)\right)\end{aligned} \tag{2.147}$$

Depending on the frequencies ω_1 and ω_2 and the deterministic and stochastic properties of the phases φ_1 and φ_2, we can detect this periodically varying intensity pattern called interference pattern. Interference of waves can be best visuallized with water waves, see Figure 2.38. Note, however, that water waves are a scalar field, whereas the EM-waves are vector waves. Therefore, the interference phenomena of EM-waves are much richer in nature than

for water waves. Notice, from Eq.(2.147), it follows immedicatly that the interference vanishes in the case of orthogonally polarized EM-waves, because of the scalar product involved. Also, if the frequencies of the waves are not identical, the interference pattern will not be stationary in time.

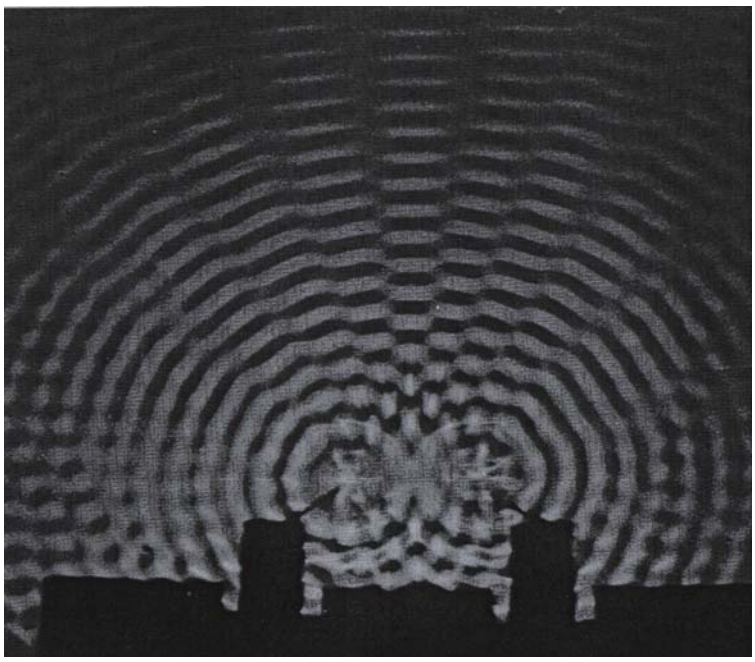

Figure 2.38: Interference of water waves from two point sources in a ripple tank [1] p. 276.

If the frequencies are identical, the interference pattern depends on the wave vectors, see Figure 2.39. The interference pattern which has itself a wavevector given by

$$\vec{k}_1 - \vec{k}_2 \tag{2.148}$$

shows a period of

$$\Lambda = \frac{2\pi}{\left|\vec{k}_1 - \vec{k}_2\right|}. \tag{2.149}$$

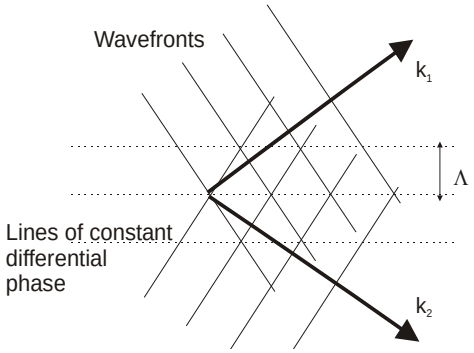

Figure 2.39: Interference pattern generated by two monochromatic plane waves.

Coherence

The ability of waves to generate an interference pattern is called coherence. Coherence can be quantified both temporally or spatially. For example, if we are at a certain position $\vec{r}$ in the interference pattern described by Eq.(2.147), we will only have stationary conditions over a time interval

$$T_{coh} << \frac{2\pi}{\omega_1 - \omega_2}.$$

Thus the spectral width of the waves determines the temporal coherence. However, it depends very often on the expermental arrangement whether a given situation can still lead to interference or not. Even so the interfering light may be perfectly temporally coherent, i.e. perfectly monochromatic, $\omega_1 = \omega_2$,yet the wave vectors may not be stable over time and the spatial inteference pattern may wash out, i.e. there is insufficient spatial coherence. So for stable and maximum interference three conditions must be fulfilled:

- stable and identical polarization
- small change in the relative phase between the beams involved over the observation time, temporal coherence, often achieved by using narrrow linewidth light
- stable beam propagation or guiding of light to achieve spatial coherence.

It is by no means trivial to arrive at a light source and an experimental setup that enables good coherence and strong interference of the beams involved.

Interference of beams can be used to measure relative phase shifts between them which may be proportional to a physical quantity that needs to be measured. Such phase shifts between two beams can also be used to modulate the light output at a given position in space via interference. In 6.013, we have already encountered interference effects between forward and backward traveling waves on transmission lines. This is very closely related to what we use in optics, therefore, we quickly relate the TEM-wave progagation to the transmission line formalism developed in Chapter 5 of 6.013.

2.3.2 TEM-Waves and TEM-Transmission Lines

The motion of voltage V and current I along a TEM transmission line with an inductance L' and a capacitance $C\prime$ per unit length is satisfies

$$\frac{\partial V(t,z)}{\partial z} = -L'\frac{\partial I(t,z)}{\partial t} \tag{2.150}$$

$$\frac{\partial I(t,z)}{\partial z} = -C'\frac{\partial V(t,z)}{\partial t} \tag{2.151}$$

Substitution of these equations into each other results in wave equations for either the voltage or the current

$$\frac{\partial^2 V(t,z)}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 V(t,z)}{\partial t^2} = 0, \tag{2.152}$$

$$\frac{\partial^2 I(t,z)}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 I(t,z)}{\partial t^2} = 0, \tag{2.153}$$

where $c = 1/\sqrt{L'C'}$ is the speed of wave propagation on the transmission line. The ratio between voltage and current for monochomatic waves is the characteristic impedance $Z = \sqrt{L'/C'}$.

The equations of motion for the electric and magnetic field of a x-polarized TEM wave according to Figure 2.1, with $E-$field along the x-axis and H-fields along the y- axis follow directly from Faraday's and Ampere's law

$$\frac{\partial E(t,z)}{\partial z} = -\mu\frac{\partial H(t,z)}{\partial t}, \tag{2.154}$$

$$\frac{\partial H(t,z)}{\partial z} = -\varepsilon\frac{\partial E(t,z)}{\partial t}, \tag{2.155}$$

which are identical to the transmission line equations (2.150) and (2.151). Substitution of these equations into each other results again in wave equations for electric and magnetic fields propagating at the speed of light $c = 1/\sqrt{\mu\varepsilon}$ and with characteristic impedance $Z_F = \sqrt{\mu/\varepsilon}$.

The solutions of the wave equation are forward and backward traveling waves, which can be decoupled by transforming the fields to the forward and backward traveling waves

$$a(t,z) = \sqrt{\frac{A_{eff}}{2Z_F}}\left(E(t,z) + Z_{Fo}H(t,z)\right), \tag{2.156}$$

$$b(t,z) = \sqrt{\frac{A_{eff}}{2Z_F}}\left(E(t,z) - Z_{Fo}H(t,z)\right), \tag{2.157}$$

which fulfill the equations

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)a(t,z) = 0, \tag{2.158}$$

$$\left(\frac{\partial}{\partial z} - \frac{1}{c}\frac{\partial}{\partial t}\right)b(t,z) = 0. \tag{2.159}$$

Note, we introduced that cross section A_{eff} such that $|a|^2$ is proportional to the total power carried by the wave. Clearly, the solutions are

$$a(t,z) = f(t - z/c_0), \tag{2.160}$$

$$b(t,z) = g(t + z/c_0), \tag{2.161}$$

which resembles the D'Alembert solutions of the wave equations for the electric and magnetic field

$$E(t,z) = \sqrt{\frac{Z_{Fo}}{2A_{eff}}}\left(a(t,z) + b(t,z)\right), \tag{2.162}$$

$$H(t,z) = \sqrt{\frac{1}{2Z_{Fo}A_{eff}}}\left(a(t,z) - b(t,z)\right). \tag{2.163}$$

Here, the forward and backward propagating fields are already normalized such that the Poynting vector multiplied with the effective area gives already the total power transported by the fields in the effective cross section A_{eff}

$$P = \vec{S}\cdot\left(A_{eff}\ \vec{e}_z\right) = A_{eff}E(t,z)H(t,z) = |a(t,z)|^2 - |b(t,z)|^2. \tag{2.164}$$

In 6.013, it was shown that the relation between sinusoidal current and voltage waves

$$V(t,z) = \mathrm{Re}\left\{\underline{V}(z)e^{j\omega t}\right\} \text{ and } I(t,z) = \mathrm{Re}\left\{\underline{I}(z)e^{j\omega t}\right\} \tag{2.165}$$

along the transmission line or corresponding electric and magnetic fields in one dimensional wave propagation is described by a generalized complex impedance $\underline{Z}(z)$ that obey's certain transformation rules, see Figure 2.40 (a).

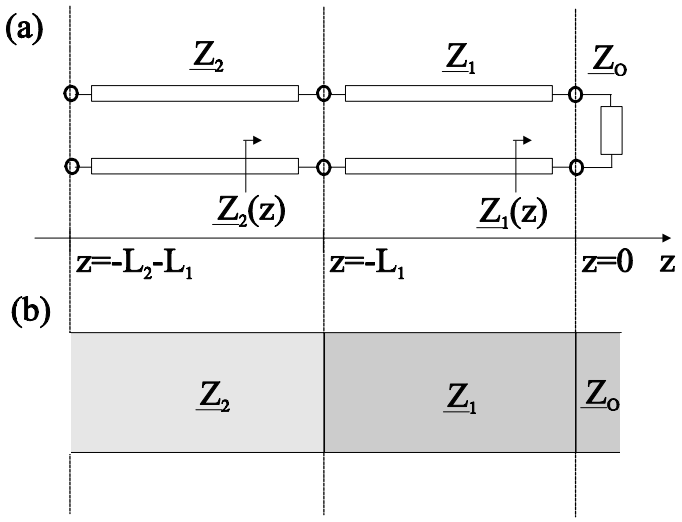

Figure 2.40: (a) Transformation of generalized impedance along transmission lines, (b) Transformation of generalized impedance accross free space sections with different characterisitc wave impedances in each section.

Along the first transmission line, which is terminated by a load impedance, the generalized impedance transforms according to

$$\underline{Z}_1(z) = \underline{Z}_1 \cdot \frac{\underline{Z}_0 - j\underline{Z}_1 \tan(k_1 z)}{\underline{Z}_1 - j\underline{Z}_0 \tan(k_1 z)} \tag{2.166}$$

with $k_1 = k_0\underline{n}_1$ and along the second transmission line the same rule applies as an example

$$\underline{Z}_2(z) = \underline{Z}_2 \cdot \frac{\underline{Z}_1(-L_1) - j\underline{Z}_2 \tan(k_2 z)}{\underline{Z}_2 - j\underline{Z}_1(-L_1) \tan(k_2 z)} \tag{2.167}$$

with $k_2 = k_0 n_2$. Note, that the media can also be lossy, then the characteristic impedances of the transmission lines and the propagation constants are already themselves complex numbers. The same formalism can be used to solve corresponding one dimensional EM-wave propagation problems.

Antireflection Coating

The task of an antireflection (AR-)coating, analogous to load matching in transmission line theory, is to avoid reflections between the interface of two media with different optical properties. One method of course could be to place the interface at Brewster's angle. However, this is not always possible. Let's assume we want to put a medium with index n into a beam under normal incidence, without having reflections on the air/medium interface. The medium can be for example a lens. This is exactly the situation shown in Figure 2.40 (b). $\underline{Z}_2$ describes the refractive index of the lense material, e.g. $n_2 = 3.5$ for a silicon lense, we can deposit on the lens a thin layer of material with index n_1 corresponding to $\underline{Z}_1$ and this layer should match to the free space index $n_0 = 1$ or impedance $\underline{Z}_0 = 377\Omega$. Using (2.166) we obtain

$$\underline{Z}_2 = \underline{Z}_1(-L_1) = \underline{Z}_1 \frac{\underline{Z}_0 - j\underline{Z}_1 \tan(-k_1 L_1)}{\underline{Z}_1 - j\underline{Z}_0 \tan(-k_1 L_1)} \tag{2.168}$$

If we choose a quarter wave thick matching layer $k_1 L_1 = \pi/2$, this simplifies to the famous result

$$\underline{Z}_2 = \frac{\underline{Z}_1^2}{\underline{Z}_0}, \tag{2.169}$$

$$\text{or } n_1 = \sqrt{n_2 n_0} \text{ and } L_1 = \frac{\lambda}{4 n_1}. \tag{2.170}$$

Thus a quarter wave AR-coating needs a material which has an index corresponding to the geometric mean of the two media to be matched. In the current example this would be $n_2 = \sqrt{3.5} \approx 1.87$

2.3.3 Scattering and Transfer Matrix

Another formalism to analyze optical systems (or microwave circuits) can be formulated using the forward and backward propagating waves, which transform much simpler along a homogenous transmission line than the total fields, i.e. the sum of forward adn backward waves. However, at interfaces

scattering of these waves occurs whereas the total fields are continuous. For monochromatic forward and backward propagating waves

$$\underline{a}(t,z) = \underline{a}(z)e^{j\omega t} \text{ and } \underline{b}(t,z) = \underline{b}(z)e^{j\omega t} \tag{2.171}$$

propagating in z-direction over a distance z with a propagation constant k, we find from Eqs.(2.158) and (2.159)

$$\begin{pmatrix} \underline{a}(z) \\ \underline{b}(z) \end{pmatrix} = \begin{pmatrix} e^{-jkz} & 0 \\ 0 & e^{jkz} \end{pmatrix} \begin{pmatrix} \underline{a}(0) \\ \underline{b}(0) \end{pmatrix}. \tag{2.172}$$

A piece of transmission line is a two port. The matrix transforming the amplitudes of the waves at the input port (1) to those of the output port (2) is called the transfer matrix, see Figure 2.41

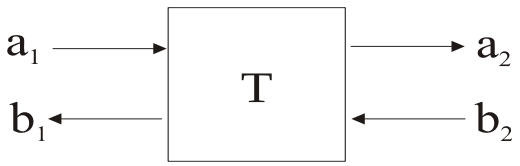

Figure 2.41: Definition of the wave amplitudes for the transfer matrix T.

For example, from Eq.(2.172) follows that the transfer matrix for free space propagation is

$$\mathbf{T} = \begin{pmatrix} e^{-jkz} & 0 \\ 0 & e^{jkz} \end{pmatrix}. \tag{2.173}$$

This formalism can be expanded to arbitrary multiports. Because of its mathematical properties the scattering matrix that describes the transformation between the incoming and outgoing wave amplitudes of a multiport is often used, see Figure 2.42.

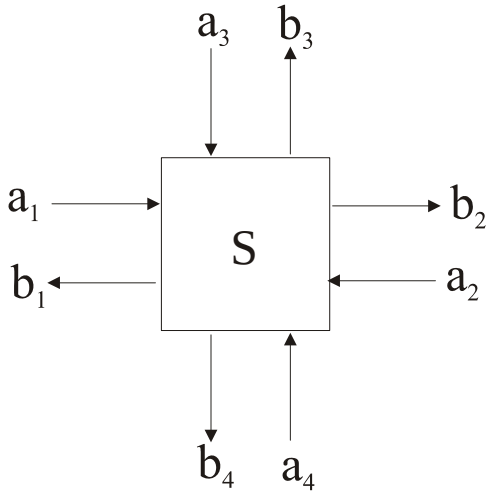

Figure 2.42: Scattering matrix and its port definition.

The scattering matrix defines a linear transformation from the incoming to the outgoing waves

$$\vec{\underline{b}} = \underline{\mathbf{S}}\vec{\underline{a}}, \text{ with } \underline{a} = (a_1, a_2, ...)^T \,, \; \underline{b} = (b_1, b_2, ...)^T \,. \tag{2.174}$$

Note, that the meaning between forward and backward waves no longer coincides with $\underline{a}$ and $\underline{b}$, a connection, which is difficult to maintain if several ports come in from many different directions.

The transfer matrix T has advantages, if many two ports are connected in series with each other. Then the total transfer matrix is the product of the individual transfer matrices.

2.3.4 Properties of the Scattering Matrix

Physial properties of the system reflect itself in the mathematical properties of the scattering matrix.

Reciprocity

A system with constant scalar dielectric and magnetic properties must have a symmetric scattering matrix (without proof)

$$\underline{\mathbf{S}} = \underline{\mathbf{S}}^T. \tag{2.175}$$

Losslessness

In a lossless system the total power flowing into the system must be equal to the power flowing out of the system in steady state

$$|\vec{\underline{a}}|^2 = \left|\vec{\underline{b}}\right|^2, \tag{2.176}$$

i.e.

$$\underline{\mathbf{S}}^+\underline{\mathbf{S}} = 1 \text{ or } \underline{\mathbf{S}}^{-1}=\underline{\mathbf{S}}^+. \tag{2.177}$$

The scattering matrix of a lossless system must be unitary.

Time Reversal

To find the scattering matrix of the time reversed system, we realize that incoming waves become outgoing waves under time reversal and the other way around, i.e. the meaning of $\underline{a}$ and $\underline{b}$ is exchanged and on top of it the waves become negative frequency waves.

$$\underline{a}e^{j(\omega t-kz)} \overset{\text{time reversal}}{\rightarrow} \underline{a}e^{j(-\omega t-kz)}. \tag{2.178}$$

To obtain the complex amplitude of the corresponding positive frequency wave, we need to take the complex conjugate value. So to obtain the equations for the time reversed system we have to perform the following substitutions

$$\begin{array}{cc} \text{Original system} & \text{Time reversed system} \\ \vec{\underline{b}} = \underline{\mathbf{S}}\vec{\underline{a}} & \vec{\underline{a}}^* = \underline{\mathbf{S}}\vec{\underline{b}}^* \rightarrow \vec{\underline{b}} = \left(\underline{\mathbf{S}}^{-1}\right)^* \vec{\underline{a}} \end{array}. \tag{2.179}$$

2.3.5 Beamsplitter

As an example, we look at the scattering matrix for a partially transmitting mirror, which could be simply formed by the interface between two media with different refractive index, which we analyzed in the previous section, see Figure 2.43. (Note, for brevity we neglect the reflections at the normal surface input to the media, or we put an AR-coating on them.) In principle, this device has four ports and should be described by a 4x4 matrix. However, most often only one of the waves is used at each port, as shown in Figure 2.43.

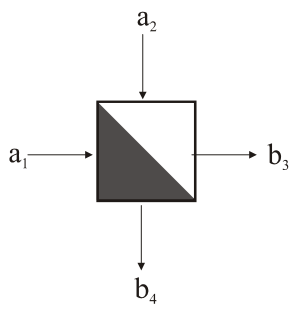

Figure 2.43: Port definitions for the beam splitter

The scattering matrix is determined by

$$\vec{\underline{b}} = \underline{\mathbf{S}}\vec{\underline{a}}, \text{ with } \underline{a} = (a_{1,}a_2)^T \ , \ \vec{\underline{b}} = (b_3, b_4)^T \tag{2.180}$$

and

$$\mathbf{S}= \begin{pmatrix} r & jt \\ jt & r \end{pmatrix}, \text{ with } r^2 + t^2 = 1. \tag{2.181}$$

The matrix $\mathbf{S}$ was obtained using using the S-matrix properties described above. From Eqs.(2.113) we could immediately identify r as a function of the refractive indices, angle of incidence and the polarization used. Note, that the off-diagonal elements of $\mathbf{S}$ are identical, which is a consequence of reciprocity. That the main diagonal elements are identical is a consequence of unitarity for a lossless beamsplitter and furthermore $t = \sqrt{1 - r^2}$. For a given frequency r and t can always be made real by choosing proper reference planes at the input and the output of the beam splitter. Beamsplitters can be made in many ways, see for example Figure 2.37.

2.3.6 Interferometers

Having a valid description of a beamsplitter at hand, we can build and analyze various types of interferometers, see Figure 2.44.

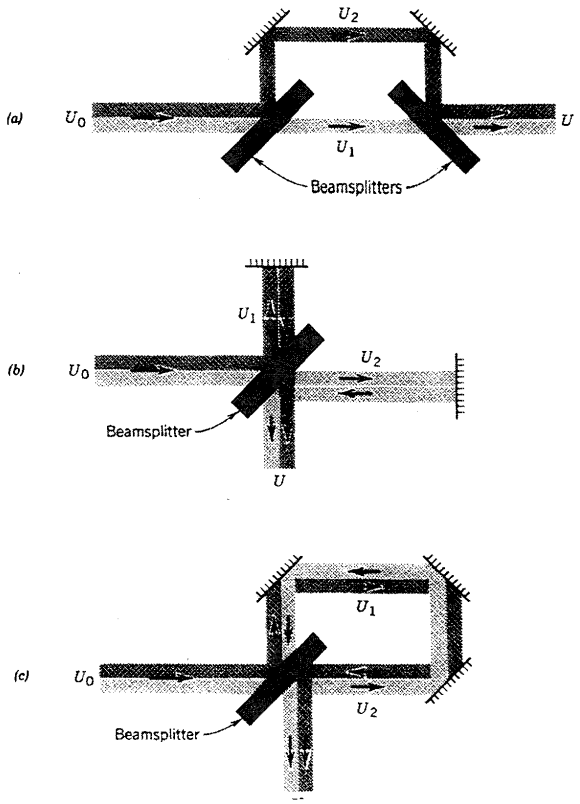

Figure 2.44: Different types of interferometers: (a) Mach-Zehnder Interferometer; (b) Michelson Interferometer; (c) Sagnac Interferometer [6] p. 66.

Each of these structures has advantages and disadvantages depending on the technology they are realized. The interferometer in Figure 2.44 (a) is called Mach-Zehnder interferometer, the one in Figure 2.44 (b) is called Michelson Interferometer. In the Sagnac interferometer , Figure 2.44 (c) both beams see identical beam path and therefore errors in the beam path can be balance out and only differential changes due to external influences lead to an output signal, for example rotation, see problem set 3.

To understand the light transmission through an interferometer we analyze as an example the Mach-Zehnder interferometer shown in Figure 2.45. If we excite input port 1 with a wave with complex amplitude $\underline{a}_0$ and no input at port 2 and assume 50/50 beamsplitters, the first beam splitter will

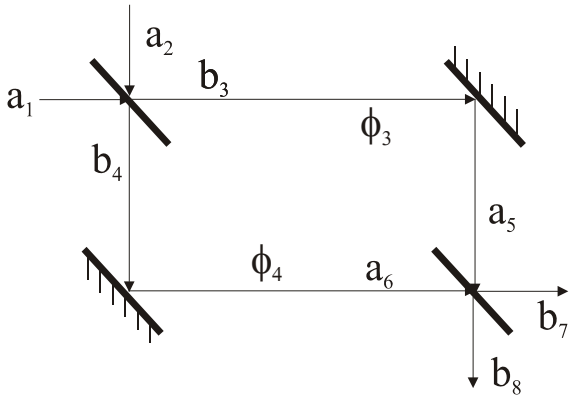

Figure 2.45: Mach-Zehnder Interferometer

produce two waves with complex amplitudes

$$\begin{aligned} \underline{b}_3 &= \tfrac{1}{\sqrt{2}}\underline{a}_0 \\ \underline{b}_4 &= j\tfrac{1}{\sqrt{2}}\underline{a}_0 \end{aligned} \tag{2.182}$$

During propagation through the interferometer arms, both waves pick up a phase delay $\phi_3 = kL_3$ and $\phi_4 = kL_4$, respectively

$$\begin{aligned} a_5 &= \tfrac{1}{\sqrt{2}}\underline{a}_0 e^{-j\phi_3}, \\ a_6 &= j\tfrac{1}{\sqrt{2}}\underline{a}_0 e^{-j\phi_4}. \end{aligned} \tag{2.183}$$

After the second beam splitter with the same scattering matrix as the first one, we obtain

$$\begin{aligned} b_7 &= \tfrac{1}{2}\underline{a}_0 \left(e^{-j\phi_3} - e^{-j\phi_4}\right), \\ b_8 &= j\tfrac{1}{2}\underline{a}_0 \left(e^{-j\phi_3} + e^{-j\phi_4}\right). \end{aligned} \tag{2.184}$$

The transmitted power to the output ports is

$$\begin{aligned} |b_7|^2 &= \tfrac{|\underline{a}_0|^2}{4}\left|1 - e^{-j(\phi_3-\phi_4)}\right|^2 = \tfrac{|\underline{a}_0|^2}{2}\left[1 - \cos\left(\phi_3 - \phi_4\right)\right], \\ |b_8|^2 &= \tfrac{|\underline{a}_0|^2}{4}\left|1 + e^{-j(\phi_3-\phi_4)}\right|^2 = \tfrac{|\underline{a}_0|^2}{2}\left[1 + \cos\left(\phi_3 - \phi_4\right)\right]. \end{aligned} \tag{2.185}$$

The total output power is equal to the input power, as it must be for a lossless system. However, depending on the phase difference $\Delta\phi = \phi_3 - \phi_4$ between both arms, the power is split differently between the two output ports, see Figure 2.46.With proper biasing, i.e. $\phi_3 - \phi_4 = \pi/2 + \Delta\phi$, the difference in

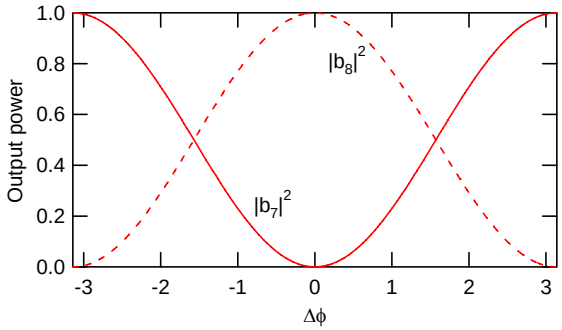

Figure 2.46: Output power from the two arms of an interfereometer as a function of phase difference.

output power between the two arms can be made directly proportional to the phase difference $\Delta\phi$.

Opening up the beam size in the interferometer and placing optics into the beam enables to visualize beam distortions due to imperfect optical components, see Figures 2.47 and 2.48.

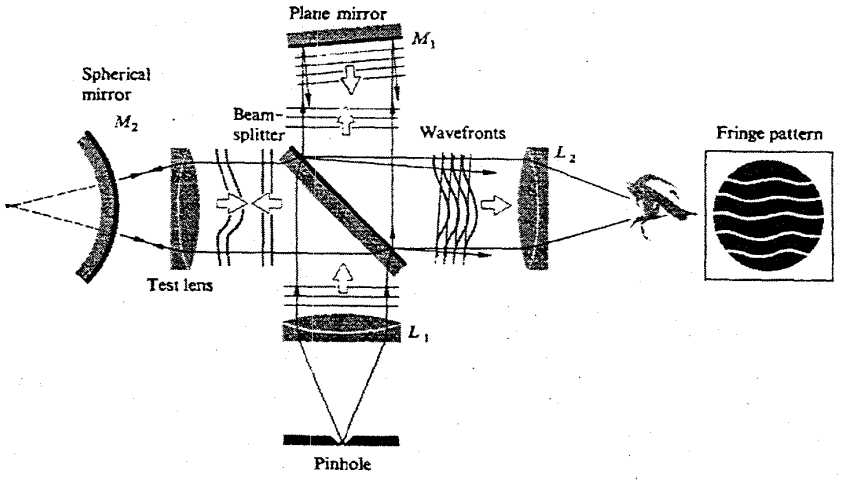

Figure 2.47: Twyman-Green Interferometer to test optics quality [1] p. 324.

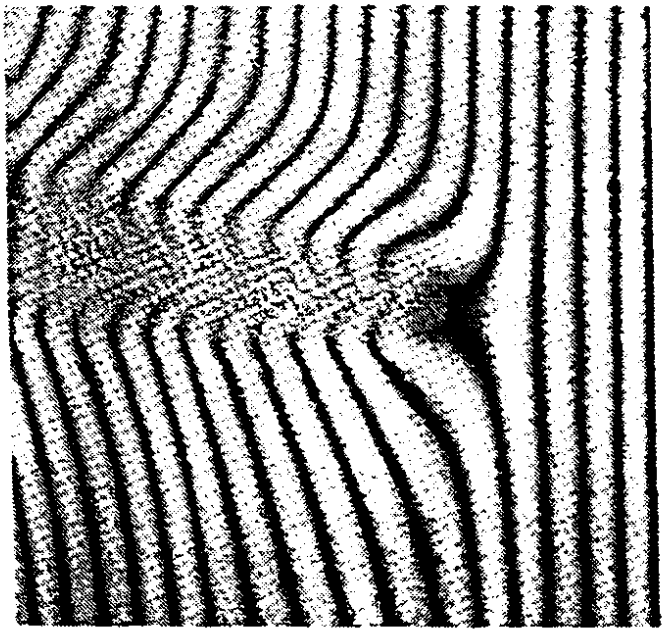

Figure 2.48: Interference pattern with a hot iron placed in one arm of the interferometer ([1], p. 395).

2.3.7 Fabry-Perot Resonator

Interferometers can act as filters. The phase difference between the interferometer arms depends on frequency, therefore, the transmission from input to output depends on frequency, see Figure 2.46. However, the filter function is not very sharp. The reason for this is that only a two beam interference is used. Much more narrowband filters can be constructed by multipass interferences such as in a Fabry-Perot Resonator, see Figure 2.49. The simplest Fabry Perot is described by a sequence of three layers where at least the middle layer has an index different from the other two layers, such that reflections occur on these interfaces.

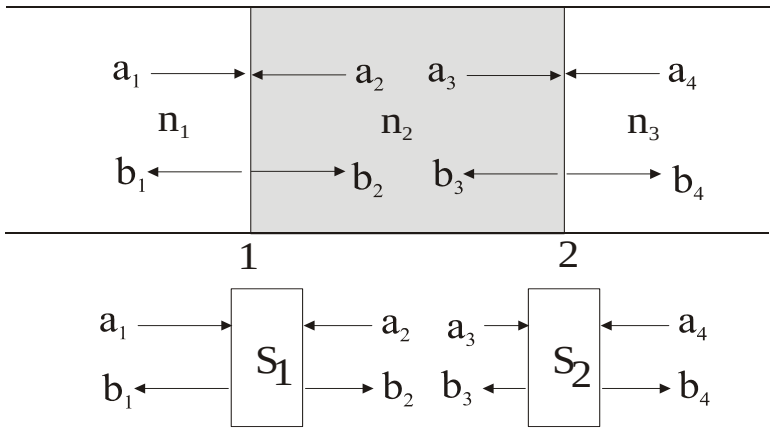

Figure 2.49: Multiple intereferences in a Fabry Perot resonator. In the simplest implementation a Fabry Perot only consists of a sequence of three layers with different refractive index so that two reflections occur with multiple interferences. Each of this discontinuites can be described by a scattering matrix.

Any kind of device that has reflections at two parallel interfaces may act as a Fabry Perot such as two semitransparent mirrors. A thin layer of material against air can act as a Fabry-Perot and is often called etalon. Given the reflection and transmission coefficients at the interfaces 1 and 2, we can write down the scattering matrices for both interfaces according to Eqs.(2.180) and (2.181).

$$\begin{pmatrix} \tilde{\underline{b}}_1 \\ \tilde{\underline{b}}_2 \end{pmatrix} = \begin{pmatrix} r_1 & jt_1 \\ jt_1 & r_1 \end{pmatrix} \begin{pmatrix} \tilde{\underline{a}}_1 \\ \tilde{\underline{a}}_2 \end{pmatrix} \text{ and } \begin{pmatrix} \tilde{\underline{b}}_3 \\ \tilde{\underline{b}}_4 \end{pmatrix} = \begin{pmatrix} r_2 & jt_2 \\ jt_2 & r_2 \end{pmatrix} \begin{pmatrix} \tilde{\underline{a}}_3 \\ \tilde{\underline{a}}_4 \end{pmatrix}. \tag{2.186}$$

If we excite the Fabry-Perot with a wave from the right with amplitude. $\tilde{\underline{a}}_1 \neq 0$, then a fraction of that wave will be transmitted to the interface into the Fabry-Perot as wave $\tilde{\underline{b}}_2$ and part will be already reflected into $\tilde{\underline{b}}_1$,

$$\tilde{\underline{b}}_1^{(0)} = r_1 \tilde{\underline{a}}_1. \tag{2.187}$$

The transmitted wave will then propagate and pick up a phase factor $e^{-j\phi/2}$, with $\phi = 2k_2 L$ and $k_2 = \frac{2\pi}{\lambda} n_2$,

$$\tilde{\underline{a}}_3 = \mathrm{j}t\tilde{\underline{a}}_1 e^{-j\phi/2}. \tag{2.188}$$

After propagation it will be reflected off from the second interface which has a reflection coefficient

$$\Gamma_2 = \left.\frac{\tilde{\underline{b}}_3}{\tilde{\underline{a}}_3}\right|_{\underline{a}_4=0} = r_2. \tag{2.189}$$

Then the reflected wave $\tilde{\underline{b}}_3$ propagates back to interface 1, picking up another phase factor $e^{-j\phi/2}$ resulting in an incoming wave after one roundtrip of $\tilde{\underline{a}}_2^{(1)} = jt_1r_2e^{-j\phi}\tilde{\underline{a}}_1$. Upon reflection on interface 1, part of this wave is transmitted leading to an output

$$\tilde{\underline{b}}_1^{(1)} = jt_1jt_1r_2e^{-j\phi}\tilde{\underline{a}}_1. \tag{2.190}$$

The partial wave $\underline{a}_2^{(1)}$ is reflected again and after another roundtrip it arrives at interface 1 as $\tilde{\underline{a}}_2^{(2)} = (r_1r_2)\,e^{-j\phi}\cdot jt_1r_2e^{-j\phi}\tilde{\underline{a}}_1$. Part of this wave is transmitted and part of it is reflected back to go through another cycle. Thus in total if we sum up all partial waves that contribute to the output at port 1 of the Fabry-Perot filter, we obtain

$$\begin{aligned}
\tilde{\underline{b}}_1 &= \sum_{n=0}^{\infty}\tilde{\underline{b}}_1^{(n)} \\
&= \left(r_1 - t_1^2r_2e^{-j\phi}\sum_{n=0}^{\infty}r_1r_2e^{-j\phi}\right)\tilde{\underline{a}}_1 \\
&= \left(r_1 - t_1^2r_2\frac{e^{-j\phi}}{1-r_1r_2e^{-j\phi}}\right)\tilde{\underline{a}}_1 \\
&= \frac{r_1 - r_2e^{-j\phi}}{1-r_1r_2e^{-j\phi}}\tilde{\underline{a}}_1
\end{aligned} \tag{2.191}$$

Note, that the coefficient in front of Eq.(2.191) is the coefficient S_{11} of the scattering matrix of the Fabry-Perot. In a similar manner, we obtain

$$\begin{pmatrix} \tilde{\underline{b}}_3 \\ \tilde{\underline{b}}_4 \end{pmatrix} = \underline{\mathbf{S}}\begin{pmatrix} \tilde{\underline{a}}_1 \\ \tilde{\underline{a}}_2 \end{pmatrix} \tag{2.192}$$

and

$$\underline{\mathbf{S}} = \frac{1}{1-r_1r_2e^{-j\phi}}\begin{pmatrix} r_1 - r_2e^{-j\phi} & -t_1t_2e^{-j\phi/2} \\ -t_1t_2e^{-j\phi/2} & r_2 - r_1e^{-j\phi} \end{pmatrix} \tag{2.193}$$

In the following, we want to analyze the properties of the Fabry-Perot for the case of symmetric reflectors, i.e. $r_1 = r_2$ and $t_1 = t_2$. Then we obtain for

the power transmission coefficient of the Fabry-Perot, $|S_{21}|^2$ in terms of the power reflectivity of the interfaces $R = r^2$

$$|S_{21}|^2 = \left| \frac{1-R}{1-Re^{-j\phi}} \right|^2 = \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2(\phi/2)} \tag{2.194}$$

Figure 2.50 shows the transmission $|S_{21}|^2$ of the Fabry-Perot interferometer for equal reflectivities $|r_1|^2 = |r_2|^2 = R$.

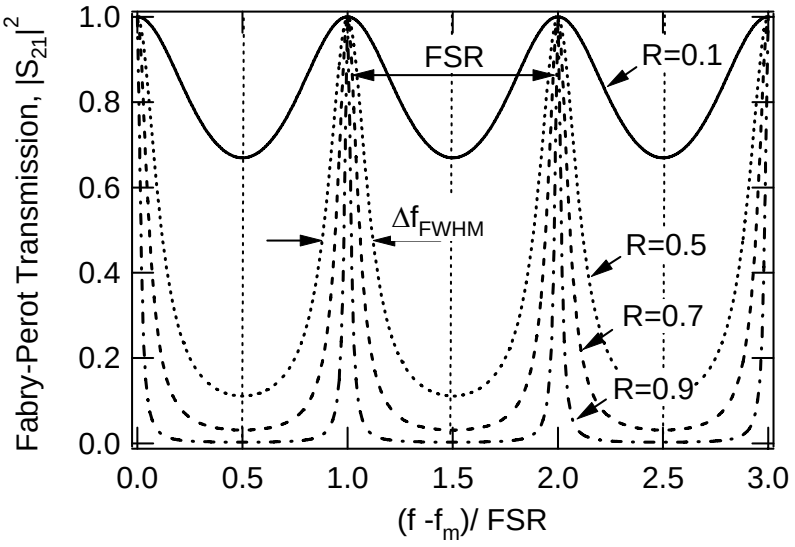

Figure 2.50: Transmission of a lossless Fabry-Perot interferometer with $|r_1|^2 = |r_2|^2 = R$

At very low reflectivity R of the mirror the transmission is almost everywhere 1, there is only a slight sinusoidal modulation due to the first order interferences which are periodically in phase and out of phase, leading to 100% transmission or small reflection. For large reflectivity R, due to the then multiple interference operation of the Fabry-Perot Interferometer, very narrow transmission resonances emerge at frequencies, where the roundtrip phase in the resonator is equal to a multiple of 2π

$$\phi = \frac{2\pi f}{c_0} n_2 2L = 2\pi m, \tag{2.195}$$

which occurs at a comb of frequencies, see Figure 2.51

$$f_m = m\frac{c_0}{2n_2L}. \tag{2.196}$$

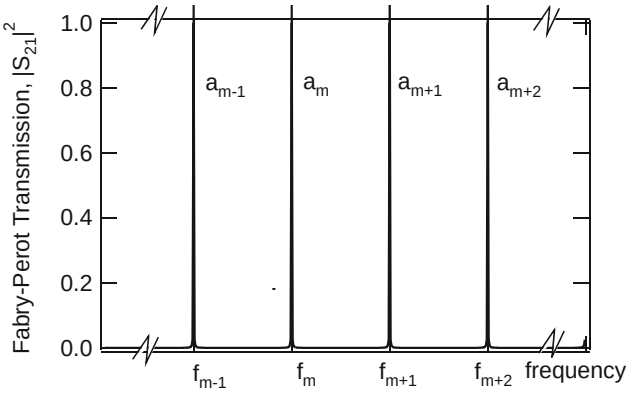

Figure 2.51: Developement of a set of discrete resonances in a one-dimensionsal resonator.

On a large frequency scale, a set of discrete frequencies, resonances or modes arise. The frequency range between resonances is called free spectral range (FSR) of the Fabry-Perot Interferometer

$$FSR = \frac{c_0}{2n_2L} = \frac{1}{T_R}, \tag{2.197}$$

which is the inverse roundtrip time T_R of the light in the one-dimensonal cavity or resonator formed by the mirrors. The filter characteristic of each resonance can be approximately described by a Lorentzian line derived from Eq.(2.194) by substituting $f = f_m + \Delta f$ with $\Delta f \ll FSR$,

$$\begin{aligned} |S_{21}|^2 &= \frac{(1-R)^2}{(1-R)^2 + 4R\sin^2\left(\left[m2\pi + 2\pi\frac{\Delta f}{FSR}\right]/2\right)} \\ &\approx \frac{1}{1+\left(\frac{2\pi\sqrt{R}}{1-R}\frac{\Delta f}{FSR}\right)^2}, \qquad (2.198) \\ &\approx \frac{1}{1+\left(\frac{\Delta f}{\Delta f_{FWHM/2}}\right)^2}, \qquad (2.199) \end{aligned}$$

where we introduced the FWHM of the transmission filter

$$\Delta f_{FWHM} = \frac{FSR}{F}, \tag{2.200}$$

with the finesse of the interferometer defined as

$$F = \frac{\pi\sqrt{R}}{1-R} \approx \frac{\pi}{T}. \tag{2.201}$$

The last simplification is valid for a highly reflecting mirror $R \approx 1$ and T is the mirror transmission. From this relation it is immediately clear that the finesse has the additional physical meaning of the optical power enhancement inside the Fabry-Perot at resonance besides the factor of π, since the power inside the cavity must be larger by $1/T$, if the transmission through the Fabry-Perot is unity.

2.3.8 Quality Factor of Fabry-Perot Resonances

Another quantity often used to characterize a resonator or a resonance is its quality factor Q, which is defined as the ratio between the resonance frequency and the decay rate for the energy stored in the resonator, which is also often called inverse photon lifetime, τ_{ph}^{-1}

$$Q = \tau_{ph} f_m. \tag{2.202}$$

Lets assume, energy is stored in one of the resonator modes which occupies a range of frequencies $[f_m - FSR/2, f_m + FSR/2]$ as indicated in Figure 2.52. Then the fourier integral

$$\underline{a}_m(t) = \int_{-FSR/2}^{+FSR/2} \underline{\tilde{b}}_2(f) e^{j2\pi(f-f_m)t}\, df, \tag{2.203}$$

where $\left|\underline{\tilde{b}}_2(f)\right|^2$ is normalized such that it describes the power spectral density of the forward traveling wave in the resonator gives the mode amplitude of the m-th mode and its magnitude square is the energy stored in the mode. Note, that we could have taken any of the internal waves $\underline{\tilde{a}}_2, \underline{\tilde{b}}_2, \underline{\tilde{a}}_3$, and $\underline{\tilde{b}}_3$. The time dependent field we create corresponds to the field of the forward or backward traveling wave at the corresponding reference plane in the resonator.

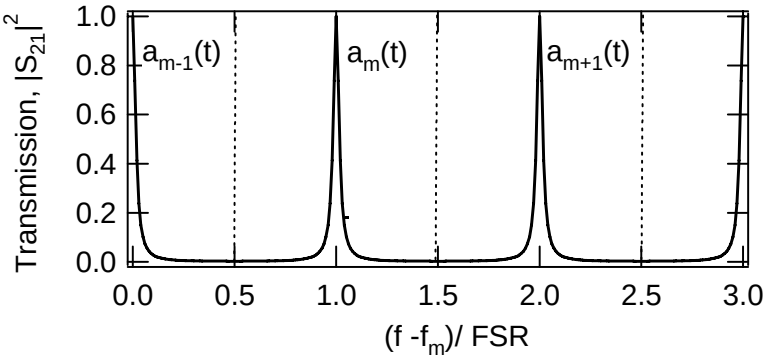

Figure 2.52: Integration over all frequency components within the frequency range $[f_m - FSR/2, f_m + FSR/2]$ defines a mode amplitude $a(t)$ with a slow time dependence

We now make a "Gedanken-Experiment". We switch on the incoming waves $\tilde{\underline{a}}_1(\omega)$ and $\tilde{\underline{a}}_4(\omega)$ to load the cavity with energy and evaluate the internal wave $\tilde{\underline{b}}_2(\omega)$. Instead of summing up all the multiple reflections like we did in constructing the scattering matrix (2.192), we exploit our skills in analyzing feedback systems, which the Fabry-Perot filter is. The scattering equations set force by the two scattering matrices characterizing the resonator mirrors in the Fabry-Perot can be visuallized by the signal flow diagram in Figure 2.53

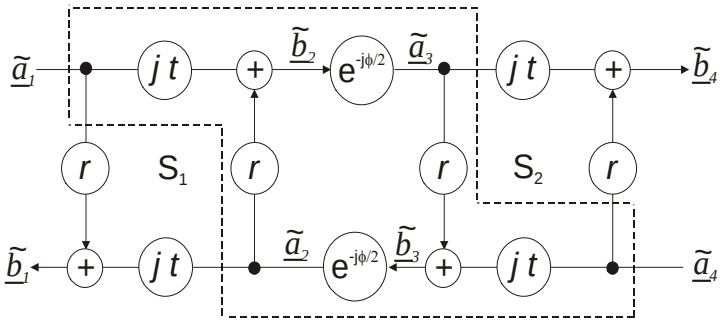

Figure 2.53: Representation of Fabry-Perot resonator by a signal flow diagram

For the task to find the relationship between the internal waves feed by the incoming wave only the dashed part of the signal flow is important. The internal feedback loop can be clearly recognized with a closed loop transfer function

$$r^2 e^{-j\phi},$$

which leads to the resonance denominator

$$1 - r^2 e^{-j\phi}$$

in every element of the Fabry-Perot scattering matrix (2.192). Using Blacks formula from 6.003 and the superposition principle we immediately find for the internal wave

$$\tilde{\underline{b}}_2 = \frac{jt}{1 - r^2 e^{-j\phi}} \left(\tilde{\underline{a}}_1 + r e^{-j\phi/2} \tilde{\underline{a}}_4 \right). \tag{2.204}$$

Close to one of the resonance frequencies, $\Omega = 2\pi f_m + \omega$, using $t = 1 - r^2$, (2.204) can be approximated by

$$\begin{aligned} \tilde{\underline{b}}_2(\omega) &\approx \frac{j}{1 + j\frac{R}{1-R}\omega T_R} \left(\tilde{\underline{a}}_1(\omega) + r(-1)^m e^{-j\omega T_R/2} \tilde{\underline{a}}_4(\omega) \right), & (2.205) \\ &\approx \frac{j}{1 + j\omega T_R/T} \left(\tilde{\underline{a}}_1(\omega) + r(-1)^m e^{-j\omega T_R/2} \tilde{\underline{a}}_4(\omega) \right) & (2.206) \end{aligned}$$

for high reflectivity R. Multiplication of this equation with the resonant denominator

$$(1 + j\omega T_R/T)\, \tilde{\underline{b}}_2(\omega) \approx j \left(\tilde{\underline{a}}_1(\omega) + r(-1)^m e^{-j\omega T_R/2} \tilde{\underline{a}}_4(\omega) \right) \tag{2.207}$$

and inverse Fourier-Transform in the time domain, while recognizing that the internal fields vanish far off resonance, i.e.

$$\underline{a}_m(t) = \int_{-\pi \cdot FSR}^{+\pi \cdot FSR} \tilde{\underline{b}}_2(\omega) e^{j\omega t}\, d\omega = \int_{-\infty}^{+\infty} \tilde{\underline{b}}_2(\omega) e^{j\omega t}\, d\omega, \tag{2.208}$$

we obtain the following differential equation for the mode amplitude slowly varying in time

$$T_R \frac{d}{dt} \underline{a}_m(t) = -T \left(\underline{a}_m(t) + j\underline{a}_1(t) + j(-1)^m \underline{a}_4(t - T_R/2) \right) \tag{2.209}$$

with the input fields

$$\underline{a}_{1/4}(t) = \int_{-\pi \cdot FSR}^{+\pi \cdot FSR} \tilde{\underline{a}}_{1/4}(\omega) e^{j\omega t}\, d\omega. \tag{2.210}$$

Despite the pain to derive this equation the physical interpretation is remarkably simple and far reaching as we will see when we apply this equation later on to many different situations. Lets assume, we switch off the loading of the cavity at some point, i.e. $\underline{a}_{1/4}(t) = 0$, then Eq.(2.209) results in

$$\underline{a}_m(t) = \underline{a}_m(0) e^{-t/(T_R/T)} \tag{2.211}$$

And the power decays accordingly

$$|\underline{a}_m(t)|^2 = |\underline{a}_m(0)|^2 e^{-t/(T_R/2T)} \tag{2.212}$$

twice as fast as the amplitude. The energy decay time of the cavity is often called the cavity energy decay time, or photon lifetime, τ_{ph}, which is here

$$\tau_{ph} = \frac{T_R}{2T}.$$

Note, the factor of two comes from the fact that each mirror of the Fabry-Perot has a transmission T per roundtrip time. For exampl a $L = 1.5m$ long cavity with mirrors of 0.5% transmission, i.e. $T_R = 10$ns and $2T = 0.01$ has a photon lifetime of 1μs. It needs hundred bounces on the mirror for a photon to be essentially lost from the cavity.

Highest quality dielectric mirrors may have a reflection loss of only $10^{-5\ldots-6}$, this is not really transmission but rather scattering loss in the mirror. Such high reflectivity mirrors may lead to the construction of cavities with photon lifetimes on the order of milliseconds.

Now, that we have an expression for the energy decay time in the cavity, we can evaluate the quality factor of the resonator

$$Q = f_m \cdot \tau_{ph} = \frac{m}{2T}. \tag{2.213}$$

Again for a resonator with the same parameters as before and at optical frequencies of 300THz corresponding to 1μm wavelength, we obtain $Q = 2 \cdot 10^8$.

2.3.9 Thin-Film Filters

Transfer matrix formlism is an efficient method to analyze the reflection and transmission properties of layered dielectric media, such as the one shown in Figure 2.54. Using the transfer matrix method, it is an easy task to compute the transmission and reflection coefficients of a structure composed of layers with arbitrary indices and thicknesses. A prominent example of a thin-film filter are Bragg mirrors. These are made of a periodic arrangement of two layers with low and high index n_1 and n_2, respectively. For maximum reflection bandwidth, the layer thicknesses are chosen to be quarter wave for the wavelength maximum reflection occures, $n_1 d_1 = \lambda_0/4$ and $n_2 d_2 = \lambda_0/4$

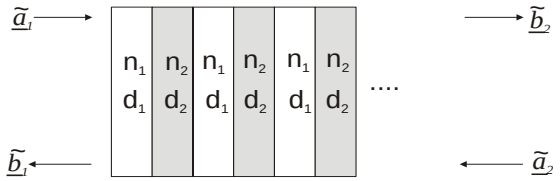

Figure 2.54: Thin-Film dielectric mirror composed of alternating high and low index layers.

As an example Figure 2.55 shows the reflection from a Bragg mirror with $n_1 = 1.45, n_2 = 2.4$ for a center wavelength of $\lambda_0 = 800$nm. The layer thicknesses are then $d_1 = 134$nm and $d_2 = 83$nm.

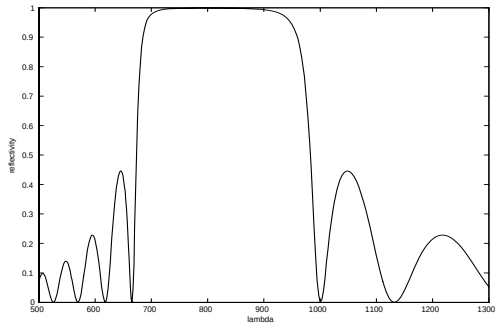

Figure 2.55: Reflectivity of an 8 pair quarter wave Bragg mirror with $n_1 = 1.45$ and $n_2 = 2.4$ designed for a center wavelength of 800nm. The mirror is embedded in the same low index material.

2.4 Paraxial Wave Equation and Gaussian Beams

So far, we have only treated optical systems operating with plane waves, which is an idealization. In reality plane waves are impossible to generate because of there infinite amount of energy required to do so. The simplest (approximate) solution of Maxwell's equations describing a beam of finite size is the Gaussian beam. In fact many optical systems are based on Gaussian beams. Most lasers are designed to generate a Gaussian beam as output. Gaussian beams stay Gaussian beams when propagating in free space. However, due to its finite size, diffraction changes the size of the beam and lenses are imployed to reimage and change the cross section of the beam. In this section, we want to study the properties of Gaussian beams and its propagation and modification in optical systems.

2.4.1 Paraxial Wave Equation

We start from the Helmholtz Equation (2.18)

$$\left(\Delta + k_0^2\right) \widetilde{\vec{E}}(x,y,z,\omega) = 0, \tag{2.214}$$

with the free space wavenumber $k_0 = \omega/c_0$. This equation can easily be solved in the Fourier domain, and one set of solutions are of course the plane waves with wave vector $|\vec{k}|^2 = k_0^2$. We look for solutions which are polarized in x-direction

$$\widetilde{\vec{E}}(x,y,z,\omega) = \widetilde{E}(x,y,z)\ \vec{e}_x. \tag{2.215}$$

We want to construct a beam with finite transverse extent into the x-y-plane and which is mainly propagating into the positive z-direction. As such we may try a superposition of plane waves with a dominant z-component of the k-vector, see Figure 2.56. The k-vectors can be written as

$$\begin{aligned} k_z &= \sqrt{k_0^2 - k_x^2 - k_y^2}, \\ &\approx k_0 \left(1 - \frac{k_x^2 - k_y^2}{2k_0^2}\right). \end{aligned} \tag{2.216}$$

with $k_x, k_y << k_0$.

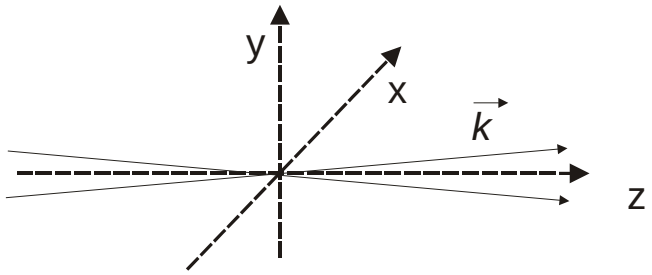

Figure 2.56: Construction of a paraxial beam by superimposing many plane waves with a dominante k-component in z-direction.

Then we obtain for the propagating field

$$\begin{aligned}\widetilde{E}(x,y,z) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\widetilde{E}_0(k_x,k_y)\cdot \\ &\quad \exp\left[-jk_0\left(1-\frac{k_x^2+k_y^2}{2k_0^2}\right)z - jk_x x - jk_y y\right]dk_x dk_y, \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\widetilde{E}_0(k_x,k_y)\cdot \\ &\quad \exp\left[j\left(\frac{k_x^2+k_y^2}{2k_0}\right)z - jk_x x - jk_y y\right]dk_x dk_y e^{-jk_0 z}, \quad (2.217)\end{aligned}$$

where $\widetilde{E}_0(k_x,k_y)$ is the amplitude for the waves with the corresponding transverse k-component. This function should only be nonzero within a small range $k_x, k_y \ll k_0$. The function

$$\widetilde{E}_0(x,y,z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\widetilde{E}_0(k_x,k_y)\exp\left[j\left(\frac{k_x^2+k_y^2}{2k_0}\right)z - jk_x x - jk_y y\right]dk_x dk_y \quad (2.218)$$

is a slowly varying function in the transverse directions x and y, and it can be easily verified that it fulfills the paraxial wave equation

$$\frac{\partial}{\partial z}\widetilde{E}_0(x,y,z) = \frac{-j}{2k_0}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial x^2}\right)\widetilde{E}_0(x,y,z). \quad (2.219)$$

Note, that this equation is in its structure identical to the dispersive spreading of an optical pulse. The difference is that this spreading occurs now in the two transverse dimensions and is called diffraction.

2.4.2 Gaussian Beams

Since the kernel in Eq.(2.218) is quadratic in the transverse k-components using a two-dimensional Gaussian for the amplitude distribution leads to a beam in real space which is also Gaussian in the radial direction because of the resulting Gaussian integral. By choosing for the transverse amplitude distribution

$$\widetilde{E}_0(k_x, k_y) = \exp\left[-\frac{k_x^2 + k_y^2}{2k_T^2}\right], \tag{2.220}$$

Eq.(2.218) can be rewritten as

$$\widetilde{E}_0(x, y, z) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left[j\left(\frac{k_x^2 + k_y^2}{2k_0}\right)(z + jz_R) - jk_x x - jk_y y\right] dk_x dk_y, \tag{2.221}$$

with the parameter $z_R = k_0/k_T^2$, which we will later identify as the Rayleigh range. Thus, Gaussian beam solutions with different finite transverse width in k-space and real space behave as if they propagate along the z-axis with different imaginary z-component z_R. Carrying out the Fourier transformation results in the Gaussian Beam in real space

$$\widetilde{E}_0(x, y, z) = \frac{j}{z + jz_R} \exp\left[-jk_0\left(\frac{x^2 + y^2}{2(z + jz_R)}\right)\right]. \tag{2.222}$$

The Gaussian beam is often formulated in terms of the complex beam parameter or q-parameter.

The propagation of the beam in free space and later even through optical imaging systems can be efficiently described by a proper transformation of the q-parameter

$$\widetilde{E}_0(r, z) = \frac{1}{q(z)} \exp\left[-jk_0\left(\frac{r^2}{2q(z)}\right)\right]. \tag{2.223}$$

Free space propagation is then described by

$$q(z) = z + jz_R \tag{2.224}$$

Using the inverse q-parameter, decomposed in real and imagniary parts,

$$\frac{1}{q(z)} = \frac{1}{R(z)} - j\frac{\lambda}{\pi w^2(z)}. \tag{2.225}$$

leads to

$$\widetilde{E}_0(r,z) = \frac{\sqrt{2P}}{\sqrt{\pi}w(z)} \exp\left[-\frac{r^2}{w^2(z)} - jk_0\frac{r^2}{2R(z)} + j\zeta(z)\right]. \tag{2.226}$$

Thus $w(z)$ is the waist of the beam and $R(z)$ is the radius of the phase fronts. We normalized the beam such that the Gaussian beam intensity $I(z,r) = \left|\widetilde{E}_0(r,z)\right|^2$ expressed in terms of the power P carried by the beam is given by

$$I(r,z) = \frac{2P}{\pi w^2(z)} \exp\left[-\frac{2r^2}{w^2(z)}\right], \tag{2.227}$$

$$\text{i.e.} \quad P = \int_0^{\infty}\int_0^{2\pi} I(r,z)\ r dr\ d\varphi. \tag{2.228}$$

The use of the q-parameter simplifies the description of Gaussian beam propagation. In free space propagation from z_1 to z_2, the variation of the beam parameter q is simply governed by

$$q_2 = q_1 + z_2 - z_1. \tag{2.229}$$

where q_2 and q_1 are the beam parameters at z_1 and z_2.

If the beam waist, at which the beam has a minimum spot size w_0 and a planar wavefront ($R = \infty$), is located at $z = 0$, the variations of the beam spot size and the radius of curvature of the phase fronts are explicitly expressed as

$$w(z) = w_o\left[1 + \left(\frac{z}{z_R}\right)^2\right]^{1/2}, \tag{2.230}$$

and

$$R(z) = z\left[1 + \left(\frac{z_R}{z}\right)^2\right], \tag{2.231}$$

where z_R is called the Rayleigh range. The Rayleigh range is the distance over which the cross section of the beam doubles. The Rayleigh range is related to the initial beam waist and the wavelength of light according to

$$z_R = \frac{\pi w_o^2}{\lambda}. \tag{2.232}$$

Intensity

Figure 2.57 shows the intensity of the Gaussian beam according to Eq.(2.227) for different propagation distances.

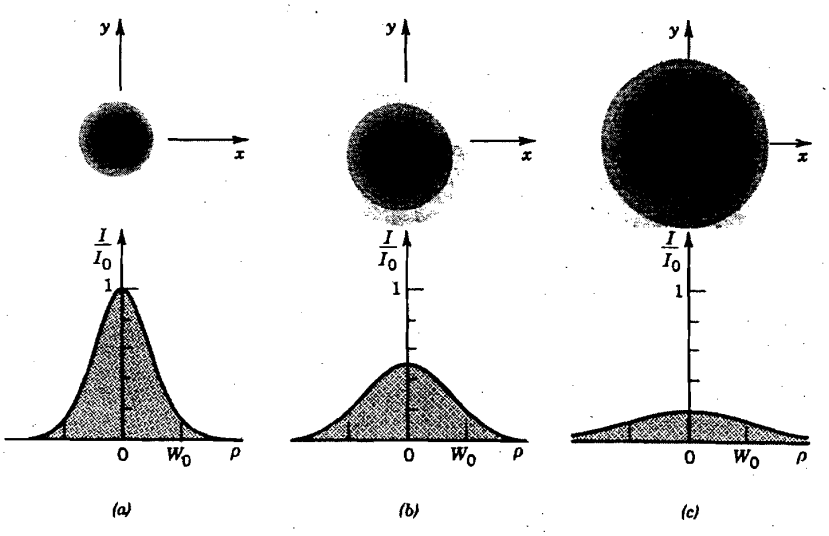

Figure 2.57: The normalized beam intensity I/I_0 as a function of the radial distance r at different axial distances: (a) z=0, (b) z=z_R, (c) z=2z_R.

The beam intensity can be rewritten as

$$I(r,z) = I_0 \frac{w_0^2}{w^2(z)} \exp\left[-\frac{2r^2}{w^2(z)}\right], \text{ with } I_0 = \frac{2P}{\pi w_0^2}. \tag{2.233}$$

For $z > z_R$ the beam radius growth linearly and therefore the area expands quadratically, which brings down the peak intensity quadratically with propagation distance.

On the beam axis ($r = 0$) the intensity is given by

$$I(r,z) = I_0 \frac{w_0^2}{w^2(z)} = \frac{I_0}{1 + \left(\frac{z}{z_R}\right)^2}. \tag{2.234}$$

The normalized beam intensity as a function of propagation distance is shown in Figure 2.58

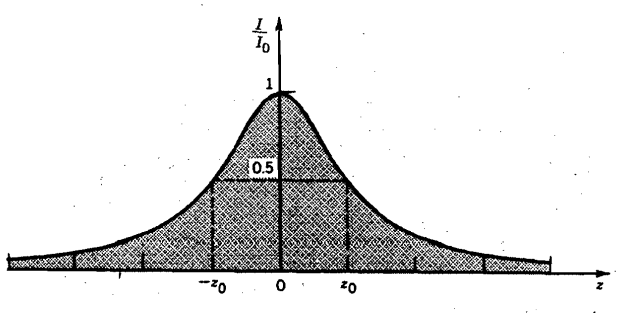

Figure 2.58: The normalized Beam intensity $I(r=0)/I_0$ on the beam axis as a function of propagation distance z [6], p. 84.

Power

The fraction of the total power contained in the beam up to a certain radius is

$$\begin{aligned} \frac{P(r<r_0)}{P} &= \frac{2\pi}{P}\int_0^{r_0} I(r,z)rdr \\ &= \frac{4}{w^2(z)}\int_0^{r_0} \exp\left[-\frac{2r^2}{w^2(z)}\right] rdr \\ &= 1-\exp\left[-\frac{2r_0^2}{w^2(z)}\right]. \end{aligned} \tag{2.235}$$

Thus, there is a certain fraction of power within a certain radius of the beam

$$\frac{P(r<w(z))}{P} = 0.86, \tag{2.236}$$

$$\frac{P(r<1.5w(z))}{P} = 0.99. \tag{2.237}$$

Beam radius

Due to diffraction, the smaller the spot size at the beam waist, the faster the beam diverges according to 2.230 as illustrated in Figure **??**.

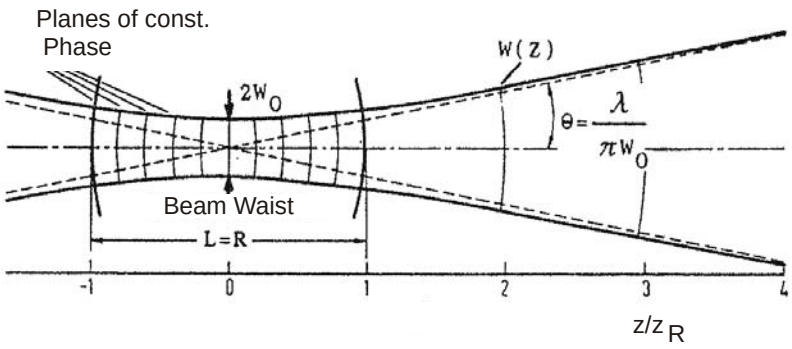

Figure 2.59: Gaussian beam and its characteristics.

Beam divergence

The angular divergence of the beam is inversely proportional to the beam waist. In the far field, the half angle divergence is given by

$$\theta = \frac{\lambda}{\pi w_o}, \tag{2.238}$$

see Figure 2.59.

Confocal parameter and depth of focus

In linear microscopy, only a layer which has the thickness over which the beam is focused, called depth of focus, will contribute to a sharp image. In nonlinear microscopy (see problem set) only a volume on the order of beam cross section times depth of focus contributes to the signal. Therefore, the depth of focus or confocal parameter of the Gaussian beam, is the distance over which the beam stays focused and is defined as twice the Rayleigh range

$$b = 2z_R = \frac{2\pi w_o^2}{\lambda}. \tag{2.239}$$

The confocal parameter depends linear on the spot size (area) of the beam and is inverse to the wavelength of light. At a wavelength of $1\mu m$ a beam with a radius of $w_o = 1cm$,.the beam will stay focussed ove distances as long

600m. However, if the beam is stronlgy focussed down to $w_o = 10\mu m$ the field of depth is only 600μm.

Phase

The phase delay of the Gaussian beam is

$$\Phi(r,z) = k_0 z - \zeta(z) + k_0 \frac{r^2}{2R(z)} \tag{2.240}$$

$$\zeta(z) = \arctan\left(\frac{z}{z_R}\right). \tag{2.241}$$

On beam axis, there is the additional phase $\zeta(z)$ when the beam undergoes focussing as shown in Figure 2.60. This is in addition to the phase shift that a uniform plane wave already aquires.

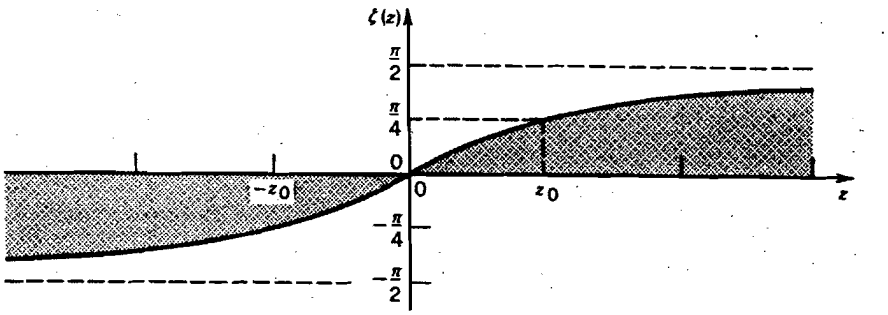

Figure 2.60: Phase delay of a Gaussian beam relative to a uniform plane wave on the beam axis [6], p. 87. This phase shift is known as Guoy-Phase-Shift.

This effect is known as Guoy-Phase-Shift. The third term in the phase shift is parabolic in the radius and describes the wavefront (planes of constant phase) bending due to the focusing, i.e. distortion from the uniform plane wave.

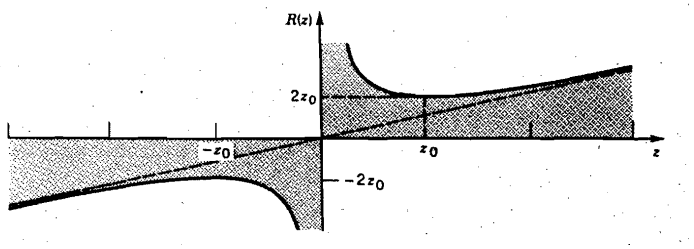

Figure 2.61: The radius of curvature $R(z)$ of the wavefronts of a Gaussian beam [6], p. 89.

The surfaces of constant phase are detemined by $k_0 z - \zeta(z) + k_0 \frac{r^2}{2R(z)} = const.$ Since the radius of curvature $R(z)$ and the additional phase $\zeta(z)$ are slowly varying functions of z, i.e. they are constant over the radial variation of the wavefront, the wavefronts are paraboloidal surfaces with radius $R(z)$, see Figures 2.61 and 2.62.

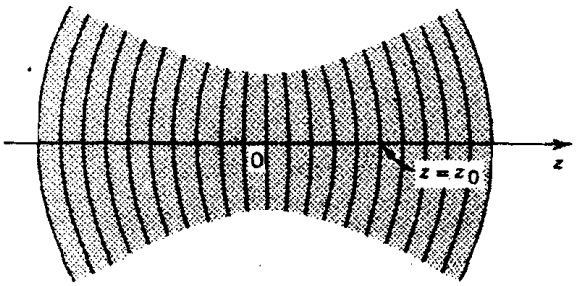

Figure 2.62: Wavefronts of a Gaussian beam, [6] p. 88.

For comparison, Figure 2.63 shows the wavefront of (a) a uniform plane wave, (b) a spherical wave and (c) a Gaussian beam. At points near the beam center, the Gaussian beam resembles a plane wave. At large z, the beam behaves like a spherical wave except that the phase fronts are delayed by a quarter of the wavlength due to the Guoy-Phase-Shift.

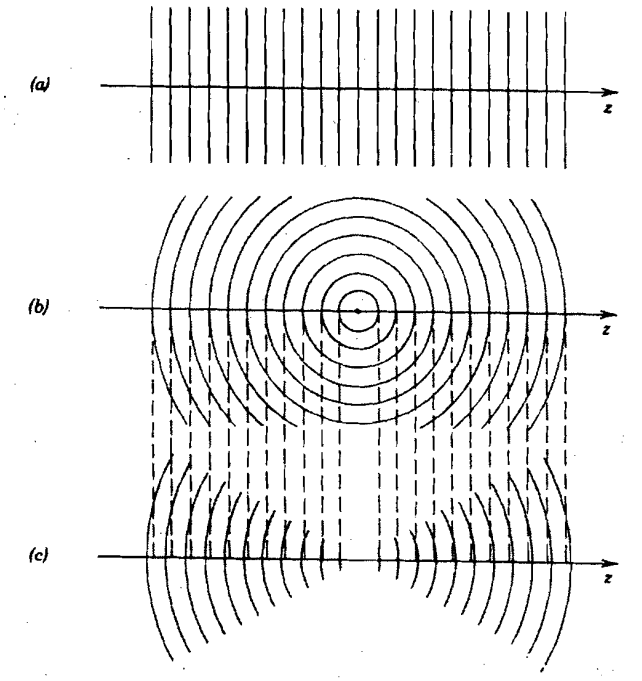

Figure 2.63: Wavefronts of (a) a uniform plane wave;(b) a spherical wave; (c) a Gaussian beam [5], p. 89.

2.5 Rays and Optical Systems

Now, that we understand how a beam of finite size as a solution of Maxwell's Equations can be constructed, we are interested how such a beam can be imaged by an optical system. Propagation of a Gaussian beam in free space leads to spreading of the beam because of the diffraction. We need means to focus the beam again. The output beam from a laser may have a certain size but we may need a different size for a given experiment. We can change the size or focus the beam by an optical imaging system. Optical systems are studied and analyzed using ray optics. What is a ray? We have already discussed that diffraction of a beam is similar to dispersion of an optical pulse. Dispersion of a pulse we understood because of the different group velocity of different frequency components or sub-pulses. It turns out that

these sub-pulses are the temporal analog to the rays. In the same way we can construct a short pulse by a superposition of sub-pulses with different center frequencies, we can construct a Gaussian beam by sub-beams with different center transverse k-vectors and a very narrow spread in transverse k-vectors. These are Gaussian beams with a large beam diameter such that diffraction is not any longer important. These beams are called rays. The ray only experiences a phase shift during propagation depending on the local refractive index $n(r)$. Therefore, we can completely understand the imaging of Gaussian beams in paraxial optical systems by the imaging properties of rays.

2.5.1 Ray Propagation

A ray propagating in an optical system, see Figure 2.64, can be described by its position r with respect to the optical axis and its inclination with respect to the optical axis r'. It is advantageous to use not (r, r') as the ray coordinates but the combination $(r, n\ r')$, where n is the local refractive index at the position of the ray. Due to propagation, the ray coordinates may change, which can be desribed by a marix, that maps initial position and inclination into the corresponding quantitaties after the propagation

$$\begin{pmatrix} r_2 \\ n_2 r_2' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_1 \\ n_1 r_1' \end{pmatrix}. \tag{2.242}$$

This imaging matrix is called an ABCD-matrix.

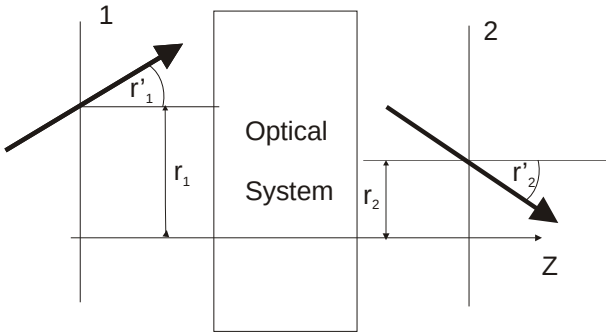

Figure 2.64: Description of optical ray propagation by its distance and inclination from the optical axis

The advantage in using $(r, n\ r')$ as the ray coordinates is that it preserves the phase space volume, i.e. for lossless optical systems the determinant of the ABCD-matrix must be 1. Also Snell's law for paraxial rays has then a simple form, see Figure 2.65. For paraxial rays the angles to the interface normal, θ_1 and θ_2, are much smaller than 1, and we can write

$$r'_1 = \tan\theta_1 \approx \sin\theta_1 \approx \theta_1, \text{ and } r'_2 = \tan\theta_2 \approx \sin\theta_2 \approx \theta_2.$$

Then Snell's law is

$$n_1\ r'_1 = n_2\ r'_2. \tag{2.243}$$

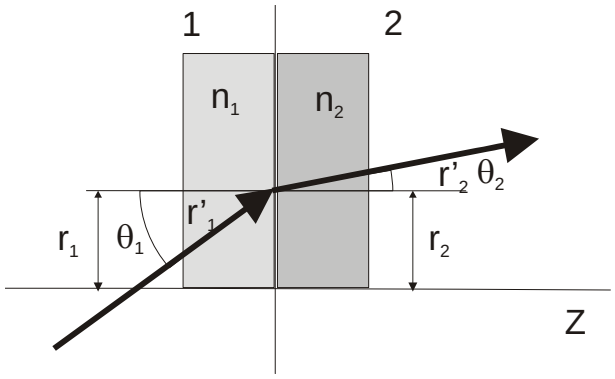

Figure 2.65: Snell's law for paraxial rays

The ABCD-matrix describing a ray going from a medium with index n_1 to a medium with index n_2 is the unity matrix

$$r_2 = r_1 \tag{2.244}$$
$$n_2\ r'_2 = n_1\ r'_1. \tag{2.245}$$

Free space propagation

For propagation in free space, see Figure 2.66, the relationship between input and output ray parameters is

$$r_2 = r_1 + r'_1 \cdot L$$
$$r'_2 = r'_1$$

or the propagation matrix is

$$\mathbf{M} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}. \tag{2.246}$$

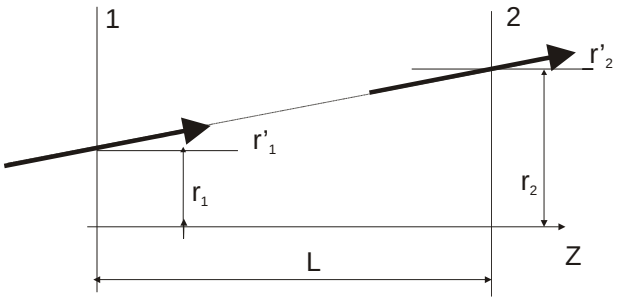

Figure 2.66: Free space propagation

Propagation in medium with length L and index n

Free propagation through a medium with index n does result in a reduced position shift with respect to the optical axis in comparison to free space, because the beam is first bent to the optical axis according to Snell's law, see Figure 2.67. Therefore the corresponding ABCD-matrix is

$$\mathbf{M} = \begin{pmatrix} 1 & L/n \\ 0 & 1 \end{pmatrix}. \tag{2.247}$$

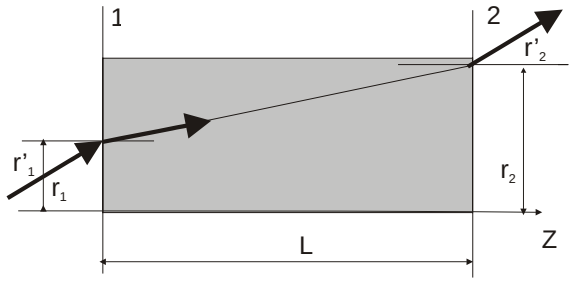

Figure 2.67: Ray propagation through a medium with refractive index n, shortens the path length of the beam by a factor of n.

Parbolic surface or thin lens

Plano-Convex Lens When a ray penetrates a parabolic surface between two media with refractive indices n_1 and n_2, it changes its inclination. A parabolic surface can be closely approximated by the surface of a sphere, see Figure 2.68. Snells law in paraxial approximation is

$$n_1 \ (r_1' + \alpha) = n_2 \ (r_2' + \alpha) \,. \tag{2.248}$$

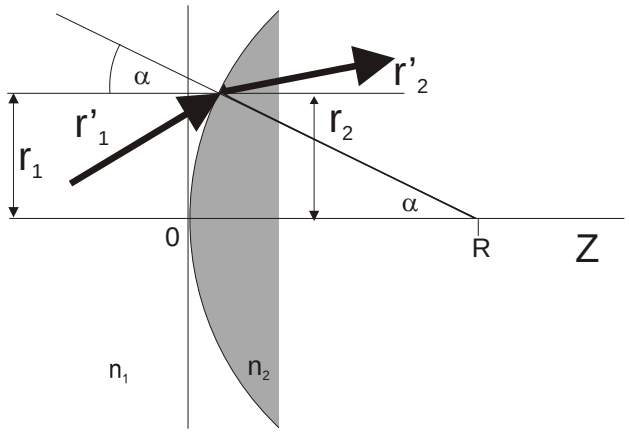

Figure 2.68: Derivation of ABCD-matrix of a thin plano-convex lens.

The small angle α can be approximated by $\alpha \approx r_1/R$. In total we then obtain the mapping

$$r_2 = r_1 \tag{2.249}$$

$$n_2 \, r_2' = n_1 \, r_1' + \frac{n_1 - n_2}{R} r_1 \tag{2.250}$$

or

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ \frac{n_1 - n_2}{R} & 1 \end{pmatrix} . \tag{2.251}$$

Note, the second normal interface does not change the ray propagation matrix and therefore Eq.(2.251) describes correctly the ray propagation through a thin plano-convex lens.

Biconvex Lens If the lens would have a second convex surface, this would refract the ray twice as strongly and we would obtain

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ 2\frac{n_1-n_2}{R} & 1 \end{pmatrix}. \tag{2.252}$$

The quantity $2\frac{n_2-n_1}{R}$ is called the refractive strength of the biconvex lense or inverse focal length $1/f$.Because the system of a thin lens plus free space propagation results in the matrix (calculated in the reverse order)

$$\mathbf{M}_{tot} = \begin{pmatrix} 1 & f \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} = \begin{pmatrix} 0 & f \\ -\frac{1}{f} & 1 \end{pmatrix}, \tag{2.253}$$

which ensures that each ray parallel to the optical axis goes through the on axis focal point at the end of the free space section, see Figure 2.69.

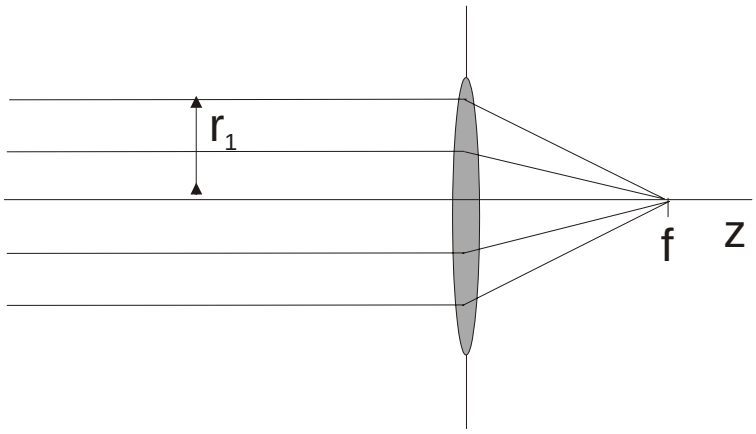

Figure 2.69: Imaging of parallel rays through a lens with focal length f.

Curved Mirrors

Other often used optical components in imaging systems are curved mirrors with radius of curvature $ROC = R$, see Figure 2.70. The advantage of reflective optics is that the rays don't have to pass through dispersive material like through a lense, which is very disturbing for ultrashort pulses.

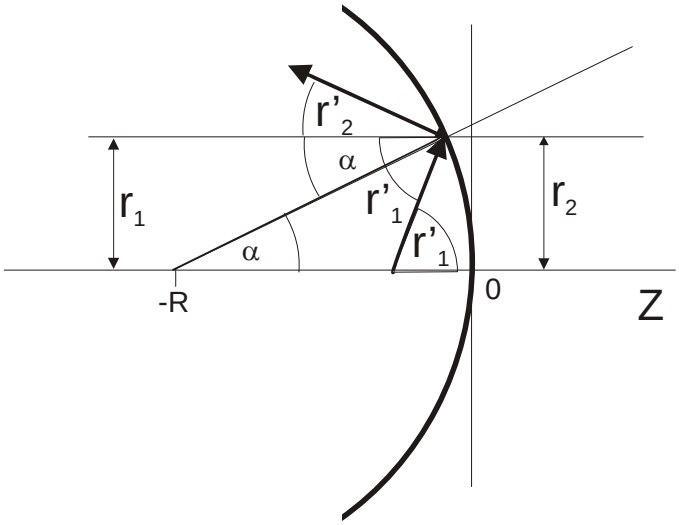

Figure 2.70: Derivation of ray matrix for concave mirror with Radius R.

As in the case of the thin lens,e the imaging does not change the distance of the ray from the optical axis, however, the slope of the rays obey

$$r_1' - \alpha = r_2' + \alpha. \tag{2.254}$$

with $\alpha \approx r_1/R$ in paraxial approximation. Therefore the ABCD matrix describing the reflection of rays at a curved mirror with $ROC = R$ is

$$\mathbf{M} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}, \text{ with } f = \frac{R}{2}. \tag{2.255}$$

2.5.2 Gauss' Lens Formula

As a simple application of the ray matrices for optical system design, we derive Gauss' lens formula, which says that all rays emitted from an orignial placed a distance d_1 from a lens with focal length f form an image at a distance d_2, which is related to d_1 by

$$\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{f}, \tag{2.256}$$

see Figure 2.71.

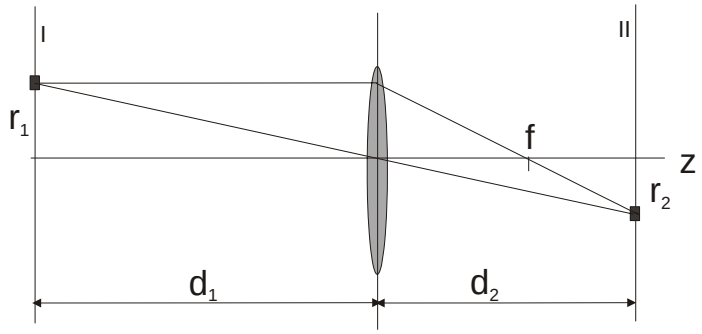

Figure 2.71: Gauss' lens formula.

The magnification of the lens system is $M_r = \frac{r_2}{r_1} = \frac{d_2}{d_1} = \left|\frac{f}{d_1 - f}\right|$. The ray matrix that describes the imagaing from the orignal plane I to the image plane II is described by the product

$$\begin{aligned}\begin{pmatrix} A & B \\ C & D \end{pmatrix} &= \begin{pmatrix} 1 & d_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} 1 & d_1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - \frac{d_2}{f} & \left(1 - \frac{d_2}{f}\right) d_1 + d_2 \\ -\frac{1}{f} & 1 - \frac{d_1}{f} \end{pmatrix}. \end{aligned} \tag{2.257}$$

In order that the distance r_2 only depends on r_1, but not on r_1', B must be 0, which is Eq. (2.256). Thus in total we have

Magnification	$M_r = \left\|\frac{f}{d_1 - f}\right\|$	
Distance to focus	$d_2 - f = M_r^2 (d_1 - f)$	

(2.258)

More complicated imaging systems, such as thick lenses, can be described by ray matrices and arbitrary paraxial optical systems can be analyzed with them, which shall not be pursued further here. Rather, we want to study how Gaussian beams are imaged by paraxial optical systems

2.6 Gaussian Beams and Resonators

2.6.1 Gaussian Beam Propagation

The propagation of Gaussian beams through paraxial optical systems can be efficiently evaluated using the ABCD-law [4], which states that the q-

parameter of a Gaussian beam passing a optical system described by an ABCD-marix is given by

$$q_2 = \frac{Aq_1 + B}{Cq_1 + D}, \tag{2.259}$$

where q_1 and q_2 are the beam parameters at the input and the output planes of the optical system or component, see Figure 2.72

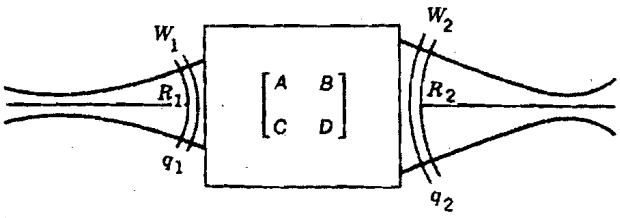

Figure 2.72: Gaussian beam transformation by ABCD law, [6], p. 99.

To proove this law, we realize that it is true for the case of free space propagation, i.e. pure diffraction, comparing (2.259) with (2.229) and (2.246). If we can proove that it is additionally true for a thin lens, then we are finished, because every ABCD matrix (2x2 matrix) can be written as a product of a lower and upper triangular matrix (LR-decomposition) like the one for free space propagation and the thin lens. Note, the action of the lens is identical to the action of free space propagation, but in the Fourier-domain. In the Fourier domain the Gaussian beam parameter is replaced by its inverse (2.222)

$$\widetilde{E}_0(x, y, z) = \frac{j}{q(z)} \exp\left[-jk_0\left(\frac{x^2 + y^2}{2q(z)}\right)\right]. \tag{2.260}$$

$$\widetilde{E}_0(k_z, k_y, z) = 2\pi j \exp\left[-jq(z)\left(\frac{k_z^2 + k_y^2}{2k_0}\right)\right] \tag{2.261}$$

But the inverse q-parameter transforms according to (2.259)

$$\frac{1}{q_2} = \frac{D\frac{1}{q_1} + C}{B\frac{1}{q_1} + A}, \tag{2.262}$$

which leads for a thin lens to

$$\frac{1}{q_2} = \frac{1}{q_1} - \frac{1}{f}. \tag{2.263}$$

This is exactly what a thin lens does, see Eq.(2.225), it changes the radius of curvature of the phase front but not the waist of the beam according to

$$\frac{1}{R_2} = \frac{1}{R_1} - \frac{1}{f}. \tag{2.264}$$

With that finding, we have proven the ABCD law for Gaussian beam propagation through paraxial optical systems.

The ABCD-matrices of the optical elements discussed so far including nonnomal incidence are summarized in Table 2.6. As an application of the

Optical Element	ABCD-Matrix
Propagation in Medium with index n and length L	$\begin{pmatrix} 1 & L/n \\ 0 & 1 \end{pmatrix}$
Thin Lens with focal length f	$\begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}$
Mirror under Angle θ to Axis and Radius R Sagittal Plane	$\begin{pmatrix} 1 & 0 \\ \frac{-2\cos\theta}{R} & 1 \end{pmatrix}$
Mirror under Angle θ to Axis and Radius R Tangential Plane	$\begin{pmatrix} 1 & 0 \\ \frac{-2}{R\cos\theta} & 1 \end{pmatrix}$
Brewster Plate under Angle θ to Axis and Thickness d, Sagittal Plane	$\begin{pmatrix} 1 & \frac{d}{n} \\ 0 & 1 \end{pmatrix}$
Brewster Plate under Angle θ to Axis and Thickness d, Tangential Plane	$\begin{pmatrix} 1 & \frac{d}{n^3} \\ 0 & 1 \end{pmatrix}$

Table 2.6: ABCD matrices for commonly used optical elements.

Gaussian beam propagation, lets consider the imaging of a Gaussian beam with a waist w_{01} by a thin lens at a distance d_1 away from the waist to a beam with a different size w_{02}, see Figure 2.73.

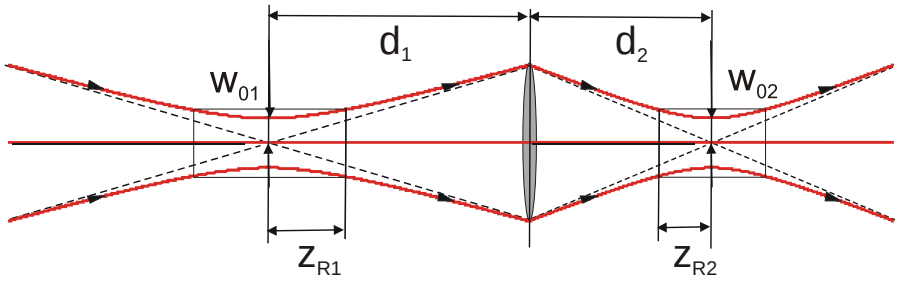

Figure 2.73: Focusing of a Gaussian beam by a lens.

There will be a new focus at a distance d_2.The corresponding ABCD matrix is of course the one from Eq.(2.257), which is repeated here

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} 1-\frac{d_2}{f} & \left(1-\frac{d_2}{f}\right) d_1 + d_2 \\ -\frac{1}{f} & 1-\frac{d_1}{f} \end{pmatrix}. \tag{2.265}$$

The q-parameter of the Gaussian beam at the position of minimum waist is purely imaginary $q_1 = jz_{R1} = j\frac{\pi w_{01}^2}{\lambda}$ and $q_2 = jz_{R2} = j\frac{\pi w_{02}^2}{\lambda}$, where

$$q_2 = \frac{A\ q_1 + B}{C\ q_1 + D} = \frac{jz_{R1}A + B}{jz_{R1}C + D} = \frac{jz_{R1}A + B}{jz_{R1}C + D} = jz_{R2}. \tag{2.266}$$

In the limit of ray optics, where the beam waists can be considered to by zero, i.e. $z_{R1} = z_{R2} = 0$ we obtain $B = 0$, i.e. the imaging rule of classical ray optics Eq.(2.256). It should not come at a surprise that for the Gaussian beam propagation this law does not determine the exact distance d_2 of the position of the new waist. Because, in the ray analysis we neglected diffraction. Therefore, the Gaussian beam analysis, although it uses the same description of the optical components, gives a slightly different and improved answer for the position of the focal point. To find the position d_2, we request that the real part of the right hand side of (2.266) is zero,

$$BD - z_{R1}^2 AC = 0 \tag{2.267}$$

which can be rewritten as

$$\frac{1}{d_2} = \frac{1}{f} - \frac{1}{d_1 + \frac{z_{R1}^2}{d_1 - f}}. \tag{2.268}$$

Again for $z_{R1} \to 0$, we obtain the ray optics result. And the imaginary part of Eq.(2.266) leads to

$$\frac{1}{z_{R2}} = \frac{1}{z_{R1}} \left(D^2 + z_{R1}^2 C^2 \right), \tag{2.269}$$

or

$$\frac{1}{w_{02}^2} = \frac{1}{w_{01}^2} \left(1 - \frac{d_1}{f} \right)^2 \left[1 + \left(\frac{z_{R1}}{d_1 - f} \right)^2 \right]. \tag{2.270}$$

With the magnification M for the spot size, with is closely related to the Magnification M_r of ray optics, we can rewrite the results as

Magnification	$M = M_r / \sqrt{1+\xi^2}$, with $\xi = \frac{z_{R1}}{d_1 - f}$ and $M_r = \left\lvert \frac{f}{d_1 - f} \right\rvert$
Beam waist	$w_{02} = M \cdot w_{01}$
Confocal parameter	$2z_{R2} = M^2\ 2z_{R2}$
Distance to focus	$d_2 - f = M^2 (d_1 - f)$
Divergence	$\theta_{02} = \theta_{01} / M$

(2.271)

2.6.2 Resonators

With the Gaussian beam solutions, we can finally construct optical resonators with finite transverse extent, i.e. real Fabry-Perots, by inserting into the Gaussian beam, see Figure 2.74, curved mirrors with the proper radius of curvature, such that the beam is imaged upon itself.

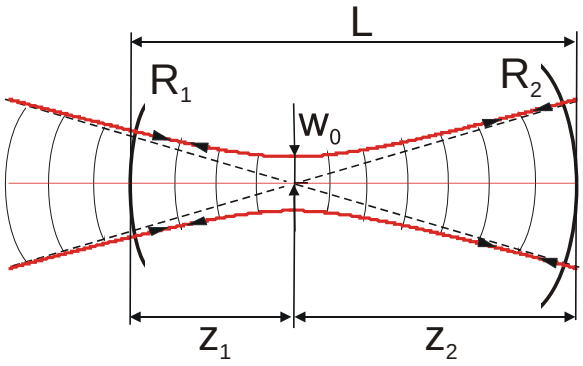

Figure 2.74: Fabry-Perot resonator with finite beam cross section by inserting curved mirrors into the beam to back reflect the beam onto itself.

Any resonator can be unfolded into a sequence of lenses and free space propagation. Here, we replace the curved mirrors by equivalent lenses with $f_1 = R_1/2$, and $f_2 = R_2/2$, see Figure 2.75.

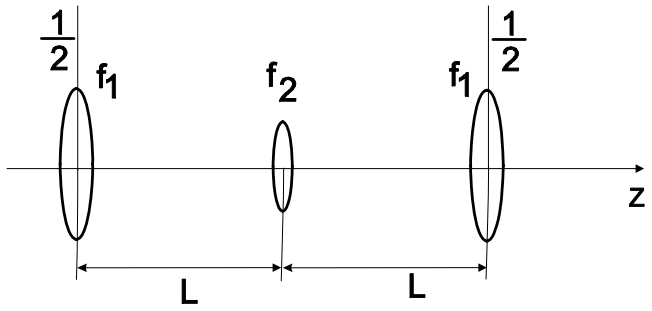

Figure 2.75: Two-mirror resonator unfolded. Note, only one half of the focusing strength of mirror 1 belongs to a fundamental period describing one resonator roundtrip.

The product of ABCD matrices describing one roundtrip of the beam in the resonator according to Figure 2.75 is

$$M = \begin{pmatrix} 1 & 0 \\ \frac{-1}{2f_1} & 1 \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{-1}{f_2} & 1 \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \frac{-1}{2f_1} & 1 \end{pmatrix}. \quad (2.272)$$

To carry out this product and to formulate the cavity stability criteria, it is convenient to use the cavity parameters $g_i = 1 - L/R_i, i = 1, 2$. The resulting cavity roundtrip ABCD-matrix can be written in the form

$$M = \begin{pmatrix} (2g_1g_2 - 1) & 2g_2L \\ 2g_1\left(g_1g_2 - 1\right)/L & (2g_1g_2 - 1) \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \quad (2.273)$$

Resonator Stability

The ABCD matrices describe the dynamics of rays propagating inside the resonator. The resonator is stable if no ray escapes after many round-trips, which is the case when the magnitude of the eigenvalues of the matrix M are less than one. Since we have a lossless resonator, i.e. $\det|M| = 1$, the product of the eigenvalues has to be 1 and, therefore, the stable resonator

corresponds to the case of a complex conjugate pair of eigenvalues with a magnitude of 1. The eigenvalue equation to M is given by

$$\det|M-\lambda\cdot 1| = \det\left|\begin{pmatrix} (2g_1g_2-1)-\lambda & 2g_2L \\ 2g_1\left(g_1g_2-1\right)/L & (2g_1g_2-1)-\lambda \end{pmatrix}\right| = 0, \quad (2.274)$$

$$\lambda^2 - 2\left(2g_1g_2-1\right)\lambda + 1 = 0. \quad (2.275)$$

The eigenvalues are

$$\lambda_{1/2} = (2g_1g_2-1) \pm \sqrt{(2g_1g_2-1)^2-1}, \quad (2.276)$$

$$= \begin{cases} \exp(\pm\theta), \cosh\theta = 2g_1g_2-1, \text{ for } |2g_1g_2-1|>1 \\ \exp(\pm j\psi), \cos\psi = 2g_1g_2-1, \text{ for } |2g_1g_2-1|\leq 1 \end{cases} \quad (2.277)$$

The case of a complex conjugate pair corresponds to a stable resontor. Therefore, the stability criterion for a stable two mirror resonator is

$$|2g_1g_2-1| \leq 1. \quad (2.278)$$

The stable and unstable parameter ranges are given by

$$\text{stable} : 0 \leq g_1\cdot g_2 = S \leq 1 \quad (2.279)$$

$$\text{unstable} : g_1g_2 \leq 0; \text{ or } g_1g_2 \geq 1. \quad (2.280)$$

where $S = g_1\cdot g_2$, is the stability parameter of the cavity. The stability criterion can be easily interpreted geometrically. Of importance are the distances between the mirror mid-points M_i and the cavity end points, i.e. $g_i = (R_i - L)/R_i = -S_i/R_i$, as shown in Figure 2.76.

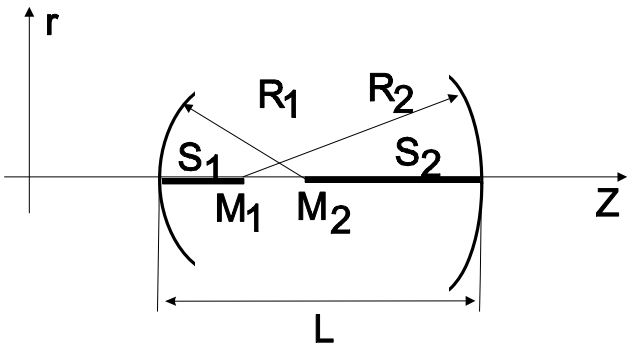

Figure 2.76: The stability criterion involves distances between the mirror mid-points M_i and the cavity end points. i.e. $g_i = (R_i - L)/R_i = -S_i/R_i$.

The following rules for a stable resonator can be derived from Figure 2.76 using the stability criterion expressed in terms of the distances S_i. Note, that the distances and radii can be positive and negative

$$\text{stable} : 0 \leq \frac{S_1 S_2}{R_1 R_2} \leq 1. \tag{2.281}$$

The rules are:

- A resonator is stable if the mirror radii, laid out along the optical axis, overlap.
- A resonator is unstable if the radii do not overlap or one lies within the other.

Figure 2.77 shows stable and unstable resonator configurations.

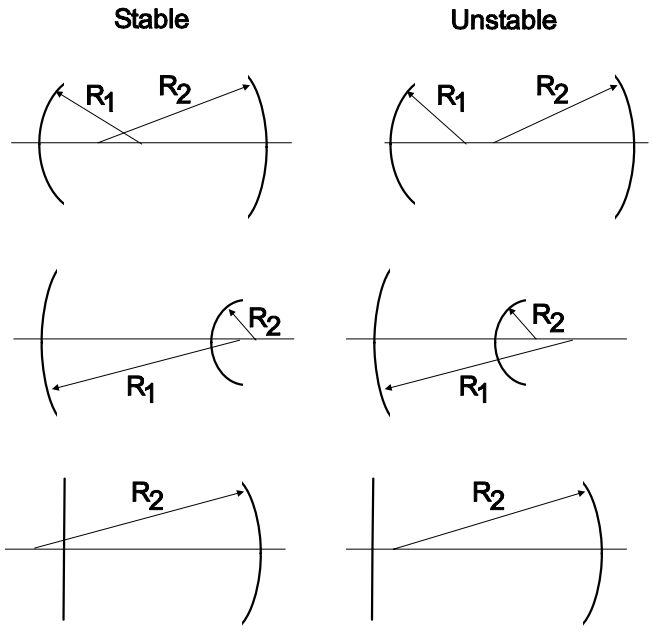

Figure 2.77: Illustration of stable and unstable resonator configurations.

For a two-mirror resonator with concave mirrors and $R_1 \leq R_2$, we obtain the general stability diagram as shown in Figure 2.78.

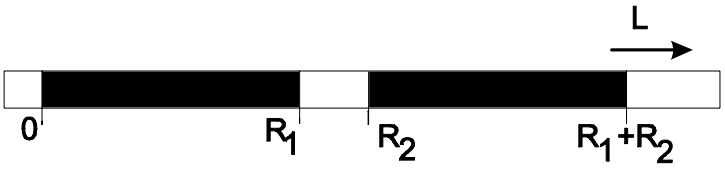

Figure 2.78: Stabile regions (black) for the two-mirror resonator.

There are two ranges for the mirror distance L, within which the cavity is stable, $0 \leq L \leq R_1$ and $R_2 \leq L \leq R_1 + R_2$. It is interesting to investigate the spot size at the mirrors and the minimum spot size in the cavity as a function of the mirror distance L.

Resonator Mode Characteristics

The stable modes of the resonator reproduce themselves after one round-trip, i.e.

$$q_1 = \frac{Aq_1 + B}{Cq_1 + D} \tag{2.282}$$

The inverse q-parameter, which is directly related to the phase front curvature and the spot size of the beam, is determined by

$$\left(\frac{1}{q}\right)^2 + \frac{A-D}{B}\left(\frac{1}{q}\right) + \frac{1-AD}{B^2} = 0. \tag{2.283}$$

The solution is

$$\left(\frac{1}{q}\right)_{1/2} = -\frac{A-D}{2B} \pm \frac{j}{2\,|B|}\sqrt{(A+D)^2 - 1} \tag{2.284}$$

If we apply this formula to (2.273), we find the spot size on mirror 1

$$\left(\frac{1}{q}\right)_{1/2} = -\frac{j}{2\,|B|}\sqrt{(A+D)^2 - 1} = -j\frac{\lambda}{\pi w_1^2}. \tag{2.285}$$

or

$$w_1^4 = \left(\frac{2\lambda L}{\pi}\right)^2 \frac{g_2}{g_1}\frac{1}{1-g_1g_2} \tag{2.286}$$

$$= \left(\frac{\lambda R_1}{\pi}\right)^2 \frac{R_2 - L}{R_1 - L}\left(\frac{L}{R_1 + R_2 - L}\right). \tag{2.287}$$

By symmetry, we find the spot size on mirror 3 by switching index 1 and 2:

$$w_2^4 = \left(\frac{2\lambda L}{\pi}\right)^2 \frac{g_1}{g_2}\frac{1}{1-g_1g_2} \tag{2.288}$$

$$= \left(\frac{\lambda R_2}{\pi}\right)^2 \frac{R_1-L}{R_2-L}\left(\frac{L}{R_1+R_2-L}\right). \tag{2.289}$$

The intracavity focus can be found by transforming the focused Gaussian beam with the propagation matrix

$$M = \begin{pmatrix} 1 & z_1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \frac{-1}{2f_1} & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1-\frac{z_1}{2f_1} & z_1 \\ \frac{-1}{2f_1} & 1 \end{pmatrix}, \tag{2.290}$$

to its new focus by properly choosing z_1, see Figure 2.74. A short calculation results in

$$z_1 = L\frac{g_2\,(g_1-1)}{2g_1g_2-g_1-g_2} \tag{2.291}$$

$$= \frac{L(L-R_2)}{2L-R_1-R_2}, \tag{2.292}$$

and, again, by symmetry

$$z_2 = L\frac{g_1\,(g_2-1)}{2g_1g_2-g_1-g_2} \tag{2.293}$$

$$= \frac{L(L-R_1)}{2L-R_1-R_2} = L-z_1. \tag{2.294}$$

The spot size in the intracavity focus is

$$w_o^4 = \left(\frac{\lambda L}{\pi}\right)^2 \frac{g_1g_2\,(1-g_1g_2)}{(2g_1g_2-g_1-g_2)^2} \tag{2.295}$$

$$= \left(\frac{\lambda}{\pi}\right)^2 \frac{L(R_1-L)(R_2-L)(R_1+R_2-L)}{(R_1+R_2-2L)^2}. \tag{2.296}$$

All these quantities for the two-mirror resonator are shown in Figure 2.79.

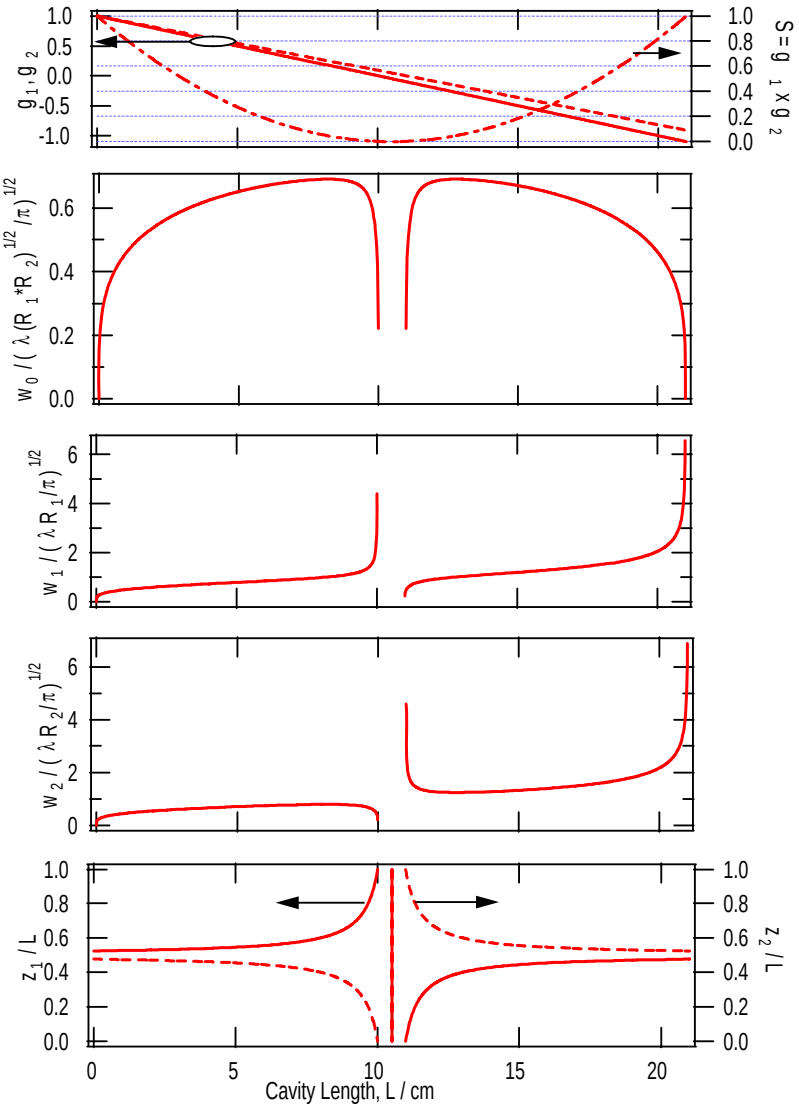

Figure 2.79: From top to bottom: Cavity parameters, g_1, g_2, S, w_0, w_1, w_2, z_1 and z_2 for the two-mirror resonator with $R_1 = 10$ cm and $R_2 = 11$ cm.

Hermite-Gaussian-Beams ($\mathbf{TEM}_{pq}$-Beams)

It turns out that the Gaussian Beams are not the only solution to the paraxial wave equation (2.219). The stable modes of the resonator reproduce themselves after one round-trip,

$$\begin{aligned}\widetilde{E}_{l,m}(x,y,z) &= A_{l,m}\left[\frac{w_0}{w(z)}\right] G_l\left[\frac{\sqrt{2x}}{w(z)}\right] G_m\left[\frac{\sqrt{2y}}{w(z)}\right] \cdot \\ &\quad \exp\left[-jk_0\left(\frac{x^2+y^2}{2R(z)}\right) + j(l+m+1)\zeta(z)\right]\end{aligned} \tag{2.297}$$

where

$$G_l\left[u\right] = H_l\left[u\right]\exp\left[-\frac{u^2}{2}\right], \text{ for } l = 0,1,2,... \tag{2.298}$$

are the Hermite-Gaussians with the Hermite-Polynomials

$$\begin{aligned}H_0\left[u\right] &= 1, \\ H_1\left[u\right] &= 2u, \\ H_2\left[u\right] &= 4u^2 - 1, \\ H_3\left[u\right] &= 8u^3 - 12u,\end{aligned} \tag{2.299}$$

and $\zeta(z)$ is the Guoy-Phase-Shift according to Eq.(2.241). The lower order Hermite Gaussians are depicted in Figure 2.80

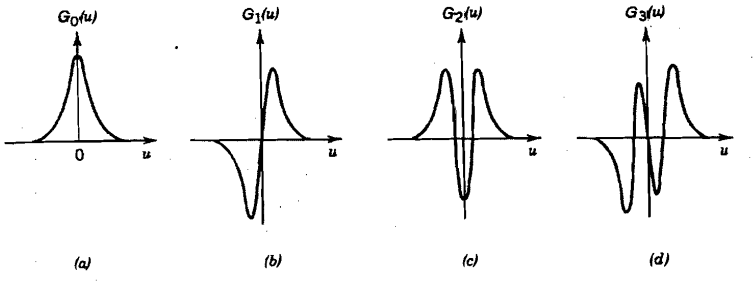

Figure 2.80: Hermite-Gauissians $G_l(u)$ for $l = 0, 1, 2$ and 3.

and the intensity profile of the first higher order resonator modes are shown in Figure 2.81.

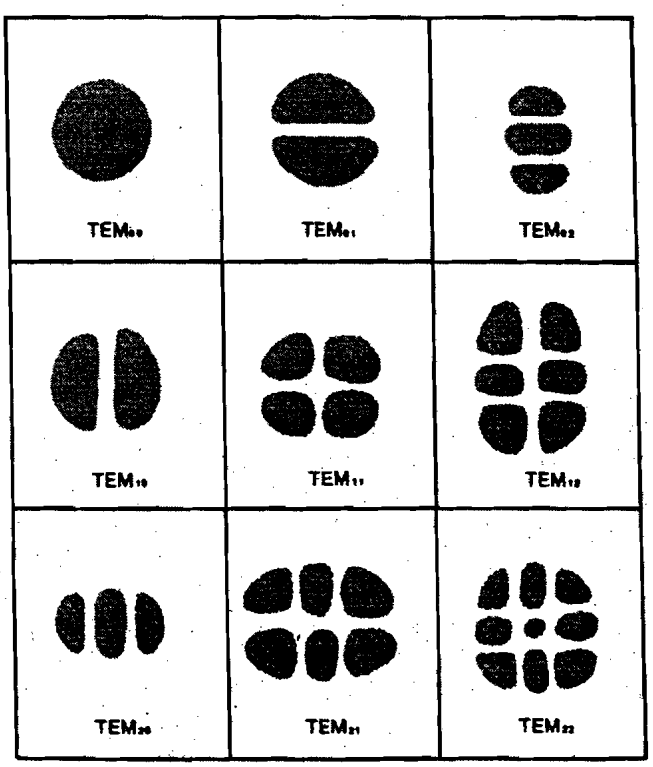

Figure 2.81: Intensity profile of TEM_{lm}-beams, [6], p. 103.

Besides the different mode profiles, the higher order modes experience greater phase advances during propogation, because they are made up of k-vectors with larger transverse components.

Axial Mode Structure

As we have seen for the Fabry-Perot resonator, the longitudinal modes are characterized by a roundtrip phase that is a multiple of 2π. Back then, we did not consider transverse modes. Thus in a resonator with finite transverse beam size, we obtain an extended family of resonances, with distinguishable field patterns. The resonance frequencies ω_{pmn} are determined by the roundtrip phase condition

$$\phi_{pmn} = 2p\pi, \text{ for } p = 0, \pm 1, \pm 2, ... \tag{2.300}$$

For the linear resonator according to Figure 2.74, the roundtrip phase of a Hermite-Gaussian T_{pmn}-beam is

$$\phi_{pmn} = 2kL - 2(m+n+1)\left(\zeta(z_2) - \zeta(z_1)\right), \tag{2.301}$$

where $\zeta(z_2) - \zeta(z_1)$ is the additional Guoy-Phase-Shift, when the beam goes through the focus once on its way from mirror 1 to mirror 2. Then the resonance frequences are

$$\omega_{pmn} = \frac{c}{L}\left[\pi p + (m+n+1)\left(\zeta(z_2) - \zeta(z_1)\right)\right]. \tag{2.302}$$

If the Guoy-Phase-Shift is not a rational number times π, then all resonance frequencies are non degenerate. However, for the special case where the two mirrors have identical radius of curvature R and are spaced a distance $L = R$ apart, which is called a confocal resonator, the Guoy-Phase-shift is $\zeta(z_2) - \zeta(z_1) = \pi/2$, with resonance frequencies

$$\omega_{pmn} = \frac{c}{L}\left[\pi p + (m+n+1)\frac{\pi}{2}\right]. \tag{2.303}$$

In that case all even, i.e. $m+n$, transverse modes are degenerate to the longitudinal or fundamental modes, see Figure 2.82.

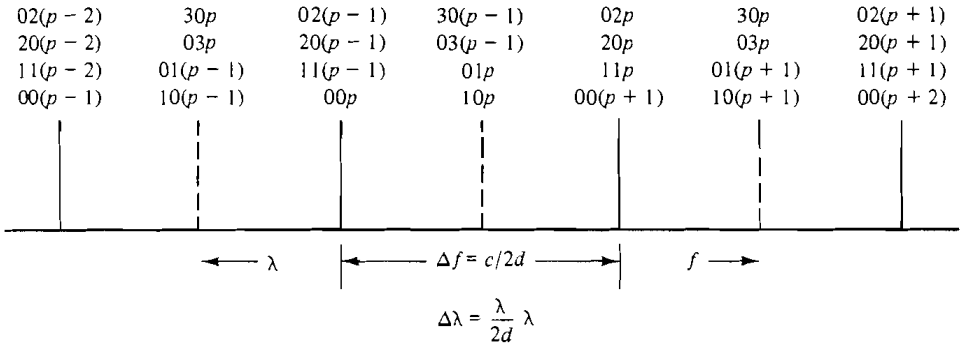

Figure 2.82: Resonance frequencies of the confocal Fabry-Perot resonator, [6], p. 128.

The odd modes are half way inbetween the longitudinal modes. Note, in contrast to the plan parallel Fabry Perot all mode frequencies are shifted by $\pi/2$ due to the Guoy-Phase-Shift.

2.7 Waveguides and Integrated Optics

As with electronics, miniaturization and integration of optics is desired to reduce cost while increasing functionality and reliability. One essential element is the guiding of the optical radiation in waveguides for integrated optical devices and optical fibers for long distance transmission. Waveguides can be as short as a few millimeters. Guiding of light with exceptionally low loss in fiber (0.1dB/km) can be achieved by using total internal reflection. Figure 2.83 shows different optical waveguides with a high index core material and low index cladding. The light will be guided in the high index core. Similar to the Gaussian beam the guided mode is made up of mostly paraxial plane waves that hit the high/low-index interface at grazing incidence and therefore undergo total internal reflections. The concomittant lensing effect overcomes the diffraction of the beam that would happen in free space and leads to stationary mode profiles fof the radiation.

Depending on the index profile and geometry one distinguishes between different waveguide types. Figure 2.83 (a) is a planar slab waveguide, which guides light only in one direction. This case is analyzed in more detail, as it has simple analytical solutions that show all phenomena associated with waveguiding such as cutoff, dispersion, single and multimode operation, coupling of modes and more, which are used later in devices and to achieve certain device properties. The other two cases show complete waveguiding in the transverse direction; (b) planar strip waveguide and (c) optical fiber.

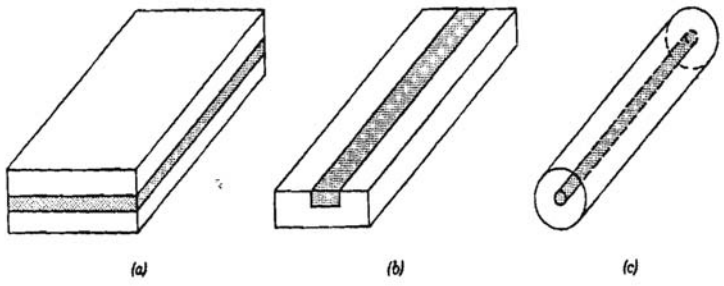

Figure 2.83: Dark shaded area constitute the high index regions. (a) planar slab waveguide; (b) strip waveguide; (c) optical fiber [6], p. 239.

In integrated optics many components are fabricated on a single sub-

strate, see Figure 2.84 with fabrication processes similar to those in microelectronics.

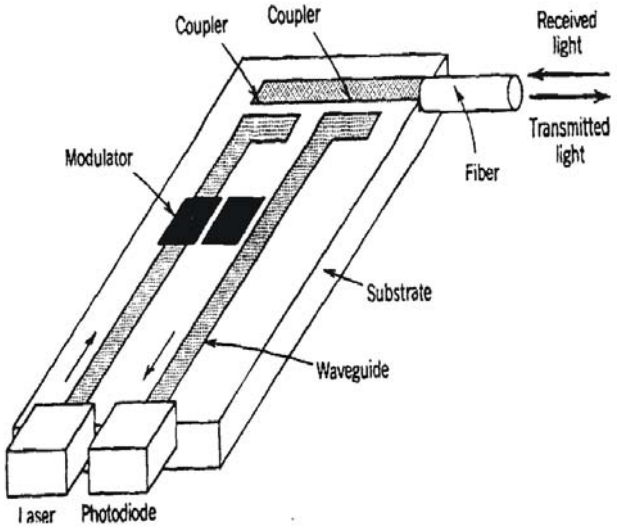

Figure 2.84: Integrated optical device resembling an optical transmitter/receiver, [6], p. 2.83.

As this example shows, the most important passive component to understand in an integrated optical circuit are waveguides and couplers.

2.7.1 Planar Waveguides

To understand the basic physics and phenomena in waveguides, we look at a few examples of guiding in one transverse dimension. These simple cases can be treated analytically.

Planar-Mirror Waveguides

The planar mirror waveguide is composed of two ideal metal mirrors a distance d apart, see Figure 2.85

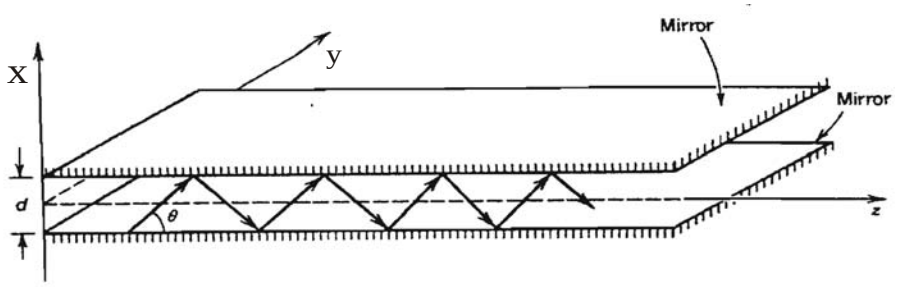

Figure 2.85: Planar mirror waveguide, [6], p. 240.

We consider a TE-wave, whose electric field is polarized in the $y-$direction and that propagates in the $z-$direction. The reflections of the light at the ideal lossless mirrors will guide or confine the light in the $x-$direction. The field will be homogenous in the $y-$direction, i.e. will not depend on y. Therefore, we make the following trial solution for the electric field of a monochromatic complex TE-wave

$$\underline{\vec{E}}(x,z,t) = \underline{E}_y(x,z)\; e^{j\omega t}\; \vec{e}_y. \tag{2.304}$$

Note, this trial solution also satisfies the condition $\nabla \cdot \vec{E} = 0$, see (2.12)

Modes of the planar waveguide Furthermore, we are looking for solutions that do not change their field distribution transverse to the direction of propagation and experience only a phase shift during propagation. We call such solutions modes of the waveguide, because they don't change its transverse field profile. The modes of the above planar waveguide can be expressed as

$$\underline{\vec{E}}_y(x,z) = u(x)\; e^{-j\beta z}\; \vec{e}_y, \tag{2.305}$$

where β is the propagation constant of the mode. This solution has to obey the Helmholtz Eq.(2.18) in the free space section between the mirrors

$$\frac{d^2}{dx^2}u_y(x) = \left(\beta^2 - k^2\right) u_y(x) \text{ with } k^2 = \frac{\omega^2}{c^2}. \tag{2.306}$$

The presence of the metal mirrors requires that the electric fields vanish at the metal mirrors, otherwise infinitely strong currents would start to flow to shorten the electric field.

$$u_y(x = \pm d/2) = 0 \tag{2.307}$$

.Note, that Eq.(2.306) is an eigenvalue problem to the differential operator $\frac{d^2}{dx^2}$

$$\frac{d^2}{dx^2}u(x) = \lambda u(x) \text{ with } u(x = \pm d/2) = 0. \tag{2.308}$$

in a space of functions u, that satisfies the boundary conditions (2.307). The eigenvalues λ are for the moment arbitrary but constant numbers. Depending on the sign of the eigenvalues the solutions can be sine or cosine functions ($\lambda < 0$) or exponentials with real exponents for ($\lambda > 0$). In the latter case, it is impossible to satiesfy the boundary conditions. Therefore, the eigensolutions are

$$u_m(x) = \begin{cases} \sqrt{\frac{2}{d}}\cos\left(k_{x,m}x\right) \text{ with } , k_{x,m} = \frac{\pi}{d}m, \ m = 1, 3, 5, ..., \text{ even modes} \\ \sqrt{\frac{2}{d}}\sin\left(k_{x,m}x\right) \text{ with } , k_{x,m} = \frac{\pi}{d}m, \ m = 2, 4, 6, ..., \text{ odd modes} \end{cases} \tag{2.309}$$

Propagation Constants The propagation constants for these modes follow from comparing (2.306) with (2.308) to be

$$\beta^2 = k^2 - k_{x,m}^2 \tag{2.310}$$

or

$$\beta = \pm\sqrt{\frac{\omega^2}{c^2} - \left(\frac{\pi}{d}m\right)^2} = \pm\sqrt{\left(\frac{2\pi}{\lambda}\right)^2 - \left(\frac{\pi}{d}m\right)^2} \tag{2.311}$$

where $\lambda = \lambda_0/n(\lambda_0)$ is the wavelength in the medium between the mirrors. This relationship is shown in Figure 2.86. The lowest order mode with index $m = 1$ has the smallest k-vector component in x-direction and therefore the largest k-vector component into z-direction. The sum of the squares of both components has to be identical to the magnitude sqaure of the k-vector in the medium k. Higher order modes have increasingly more nodes in the x-direction, i.e. largest k_x-components and the wave vector component in z-direction decreases, until there is no real solution anymore to Eq.(**??**) and the corresponding propagation constants β_m become imaginary. That is, the corresponding waves become evanescent waves, i..e they can not propagate in a waveguide with the given dimensions.

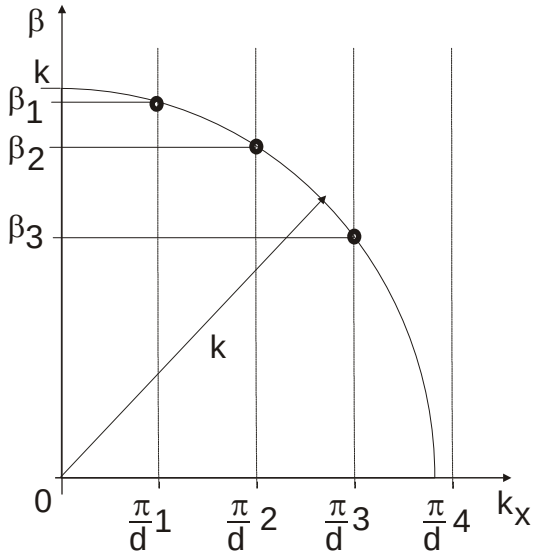

Figure 2.86: Determination of propagation constants for modes

Field Distribution The transverse electric field distributions for the various TE-modes is shown in Figure 2.87

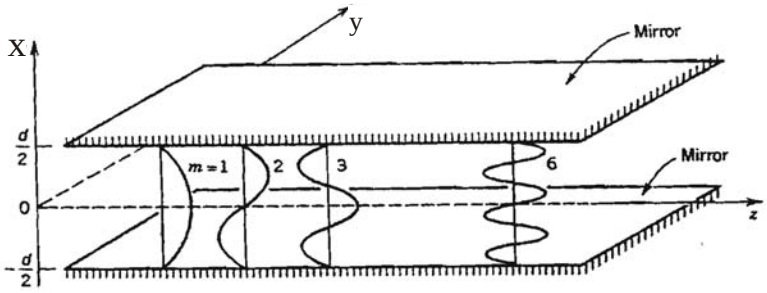

Figure 2.87: Field distributions of the TE-modes of the planar mirror waveguide [6], p. 244.

Cutoff Wavelength/Frequency For a given planar waveguide with separation d, there is a lowest frequency, i.e. longest wavelength, beyond which no propagating mode exists. This wavelenth/frequency is refered to as cutoff

wavelength/frequency which is

$$\lambda_{cutoff} = 2d \tag{2.312}$$

$$f_{cutoff} = \frac{c}{2d}. \tag{2.313}$$

The physical origin for the existence of a cutoff wavelength or frequency is that the guided modes in the mirror waveguide are a superposition of two plane waves, that propagate under a certain angle towards the z-axis, see Figure 2.88

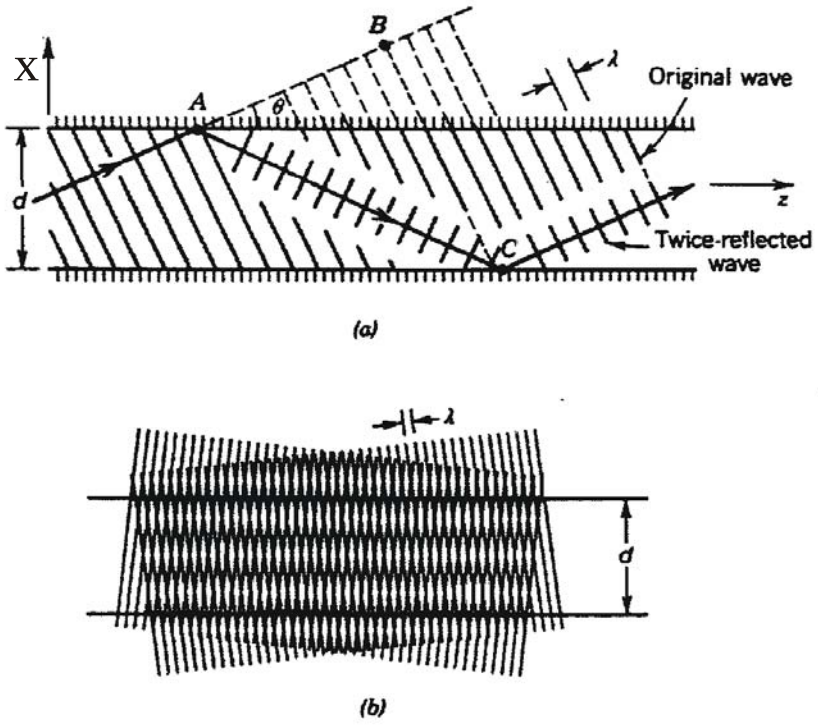

Figure 2.88: (a) Condition for self-consistency: as a wave reflects twice it needs to be in phase with the previous wave. (b) The angles for which self-consistency is achieved determine the x-component of the $\vec{k}$-vectors involved. The corresponding two plane waves setup an interference pattern with an extended node at the position of the metal mirrors satisfying the boundary conditions, [6], p. 241.

In order that the sum of the electric field of the two plane waves fulfills the boundary conditions, the phase of one of the plane waves after reflection on both mirrors needs to be inphase with the other plane wave, i.e. the x-component of the $\vec{k}$-vectors involved, k_x, must be a multiple of 2π

$$2k_x d = \pm 2\pi m.$$

If we superimpose two plane waves with $k_{x,m} = \pm\pi m/d$, we obtain an interference pattern which has nodes along the location of the metal mirrors, which obviously fulfills the boundary conditions. It is clear that the minimum distance between these lines of nodes for waves of a given wavelength λ is $\lambda/2$, hence the separation d must be greater than $\lambda/2$ otherwise no solution is possible.

Single-Mode Operation For a given separation d, there is a wavelength range over which only a single mode can propagate, we call this wavelength range single-mode operation. From Figure 2.86 it follows for the planar mirror waveguide

$$\frac{\pi}{d} < k < \frac{\pi}{d}2 \tag{2.314}$$

or

$$d < \lambda < 2d \tag{2.315}$$

Waveguide Dispersion Due to the waveguiding, the relationship between frequency and propagation constant is no longer linear. This does not imply that the waveguide core, i.e. here the medium between the plan parallel mirrors, has dispersion. For example, even for $n = 1$, we find for phase and group velocity of the m-th mode

$$\frac{1}{v_p} = \frac{\beta(\omega)}{\omega} = \frac{1}{c}\sqrt{1 - \left(\frac{c\pi}{d\omega}m\right)^2} \tag{2.316}$$

$$= \frac{1}{c}\sqrt{1 - \left(\frac{\lambda}{2d}m\right)^2} \tag{2.317}$$

and

$$\frac{1}{v_g} = \frac{d\beta(\omega)}{d\omega} = \frac{1}{2\sqrt{\frac{\omega^2}{c^2} - \left(\frac{\pi}{d}m\right)^2}}2\frac{\omega}{c^2} \tag{2.318}$$

or

$$v_g \cdot v_p = c^2. \tag{2.319}$$

Thus different modes have different group and phase velocities. Figure 2.89 shows group and phase velocity for the different modes as a function of the normalized wave number kd/π.

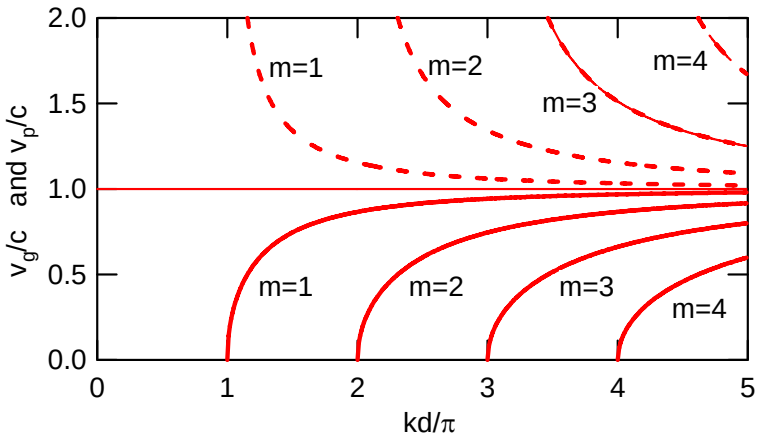

Figure 2.89: Group and phase velocity of propagating modes with index m as a function of normailzed wave number.

TM-Modes The planar mirror waveguide does not only allow for TE-waves to propagate. There are also TM-waves, which have only a magnetic field component transverse to the propagation direction and parallel to the mirrors, i.e. in y-direction

$$\vec{\underline{H}}(x,z,t) = \underline{H}_y(x,z)\; e^{j\omega t}\; \vec{e}_y, \tag{2.320}$$

and now $\underline{H}(x,z)$ has to obey the Helmholtz equation for the magnetic field. The corresponding electric field can be derived from Ampere's law

$$\vec{\underline{E}}(x,z) = \frac{-1}{j\omega\varepsilon}\nabla \times \left(\underline{H}_y(x,z)\; \vec{e}_y\right) \tag{2.321}$$

$$= \frac{1}{j\omega\varepsilon}\frac{\partial \underline{H}_y(x,z)}{\partial z}\vec{e}_x + \frac{-1}{j\omega\varepsilon}\frac{\partial \underline{H}_y(x,z)}{\partial x}\vec{e}_z. \tag{2.322}$$

The electric field tangential to the metal mirrors has to vanish again, which leads to the boundary condition

$$\frac{\partial \underline{H}_y(x,z)}{\partial x}(x = \pm d/2) = 0. \tag{2.323}$$

After an analysis very similar to the discussion of the TE-waves we find for the TM-modes with

$$\underline{H}_y(x,z) = u(x)\ e^{-j\beta z}\ \vec{e}_y, \tag{2.324}$$

the transverse mode shapes

$$u_m(x) = \begin{cases} \sqrt{\frac{2}{d}}\cos\left(k_{x,m}x\right) \text{ with }, k_{x,m} = \frac{\pi}{d}m,\ m = 2,4,6,..., \text{ even modes} \\ \sqrt{\frac{2}{d}}\sin\left(k_{x,m}x\right) \text{ with }, k_{x,m} = \frac{\pi}{d}m,\ m = 1,3,5,..., \text{ odd modes} \end{cases} \tag{2.325}$$

Note, that in contrast to the electric field of the TE-waves being zero at the metal surface, the transverse magnetic field of theTM-waves is at a maximum at the metal surface. We will not consider this case further, because the discussion of cutoff frequencies and dispersion can be worked out very analogous to the case for TE-modes.

Multimode Propagation Depending on the boundary conditions at the input of the waveguide at $z = 0$ many modes may be excited. Eventually there are even excitations with such high transverse wavevectors k_x present, that are below cutoff. Depending on the excitation amplitudes of each mode, the total field in the waveguide will be the superposition of all modes. Lets assume that there are only TE-modes excited, then the total field is

$$\underline{\vec{E}}(x,z,t) = \sum_{m=1}^{\infty}\left(\underline{a}_m\ e^{-j\beta_m z} + \underline{b}_m\ e^{j\beta_m z}\right)\ u_m(x)\ e^{j\omega t}\ \vec{e}_y, \tag{2.326}$$

where the amplitudes $\underline{a}_m$ and $\underline{b}_m$ are the excitations of the m-th mode in forward and backward direction, respectively. It is easy to show that these excitation amplitudes are determined by the transverse electric and magnetic fields at $z = 0$ and $t = 0$. In many cases, the excitation of the waveguide will be such that only the forward propagating modes are excited.

$$\underline{\vec{E}}(x,z,t) = \sum_{m=1}^{\infty}\underline{a}_m\ u_m(x)\ e^{-j\beta_m z}\ e^{j\omega t}\ \vec{e}_y, \tag{2.327}$$

When many modes are excited, the transverse field distribution will change during propagation, see Figure 2.90

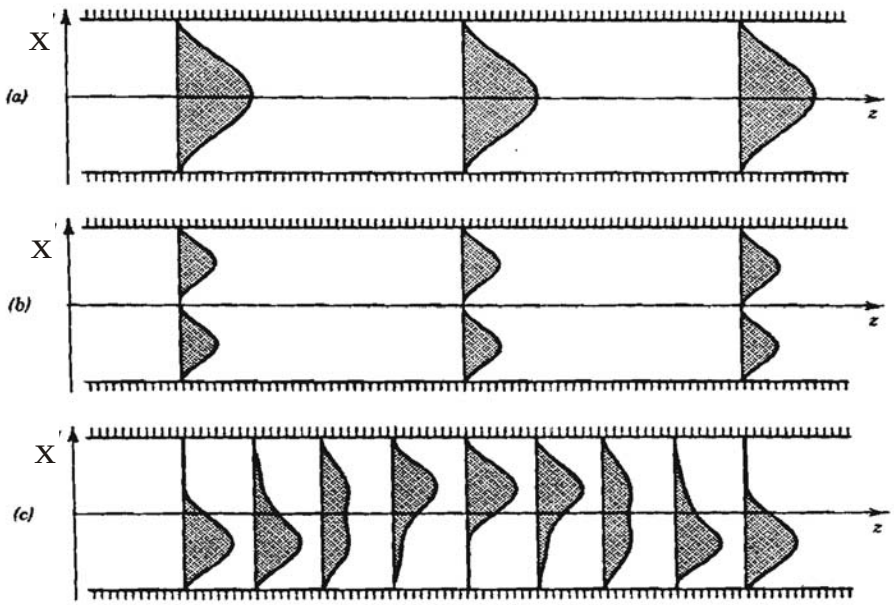

Figure 2.90: Variation of the intensity distribution in the transverse direction x at different distances z. Intensity profile of (a) the fundamental mode $m = 1$, (b) the second mode with $m = 2$ and (c) a linear combination of the fundamental and second mode, [6], p. 247.

Modes which are excited below cutoff will decay rapidly as evanescent waves. The other modes will propagate, but due to the different propagation constants these modes superimpose differently at different propagation distances along the waveguide. This dynamic can be used to build many kinds of important integrated optical devices, such as multimode interference couplers (see problem set 5). Depending on the application, undesired multimode excitation may be very disturbing due to the large group delay difference between the different modes. This effect is called modal dispersion.

Mode Orthogonality

It turns out that the transverse modes determined by the functions $u_m(x)$ build an orthogonal set of basis functions into which any function in a certain function space can be decomposed. This is obvious for the case of the planar-mirror waveguide, where the $u_m(x)$ are a subset of the basis functions for a Fourier series expansion of an arbitrary function $f(x)$ in the interval $[-d/2, 3d/2]$ which is antisymmetric with respect to $x = d/2$ and fullfills the boundary condition $f(x = \pm d/2) = 0$. It is

$$\int_{-d/2}^{d/2} u_m(x)\ u_n(x)\ dx = \delta_{mn}, \tag{2.328}$$

$$f(x) = \sum_m a_m\ \ u_m(x) \tag{2.329}$$

$$\text{with } a_m = \int_{-d/2}^{d/2} u_m(x)\ f(x)\ dx \tag{2.330}$$

From our familiarity with Fourier series expansions of periodic functions, we can accept these relations here without proof. We will return to these equations later in Quantum Mechanics and discuss in which mathematical sense Eqs.(2.328) to (2.329) really hold.

Besides illustrating many important concepts, the planar mirror waveguide is not of much practical use. More in use are dielectric waveguides.

Planar Dielectric Slab Waveguide

In the planar dielectric slab waveguide, waveguiding is not achieved by real reflection on a mirror but rather by total internal reflection at interfaces between two dielectric materials with refractive indices $n_1 > n_2$, see Figure 2.91

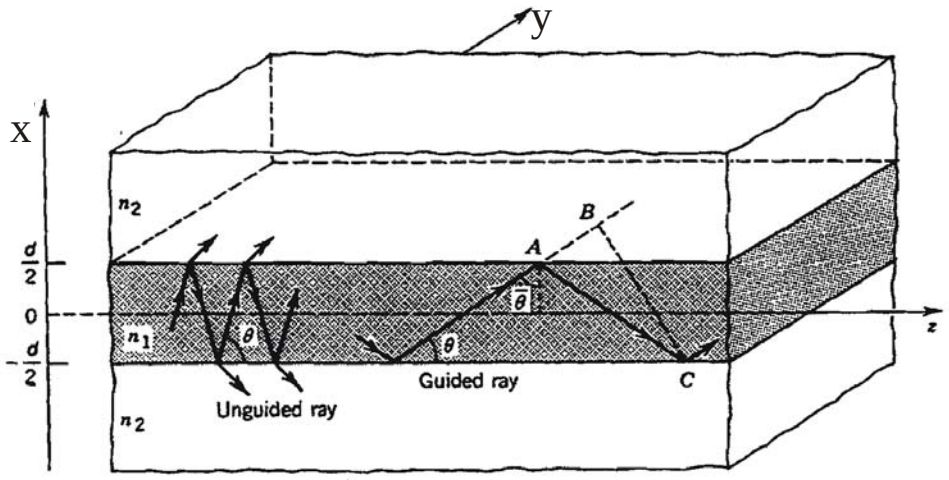

Figure 2.91: Symmetric planar dielectric slab waveguide with $n_1 > n_2$. The light is guided by total internal reflection. The field is evanescent in the cladding material and oscillatory in the core, [6], p. 249.

Waveguide Modes As in the case of the planar mirror waveguide, there are TE and TM-modes and we could find them as a superposition of correspondingly polarized TEM waves propagating with a certain transverse k-vector such that total internal reflection occurs. We do not want to follow this procedure here, but rather use immediately the Helmholtz Equation. We again write the electric field

$$\underline{\vec{E}}_y(x,z) = u(x)\ e^{-j\beta z}\ \vec{e}_y. \tag{2.331}$$

The field has to obey the Helmholtz Eq.(2.18) both in the core and in the cladding

$$\text{core} \quad : \quad \frac{d^2}{dx^2}u(x) = \left(\beta^2 - k_1^2\right) u(x) \text{ with } k_1^2 = \frac{\omega^2}{c_0^2}n_1^2, \tag{2.332}$$

$$\text{cladding} \quad : \quad \frac{d^2}{dx^2}u(x) = \left(\beta^2 - k_2^2\right) u(x) \text{ with } k_2^2 = \frac{\omega^2}{c_0^2}n_2^2 \tag{2.333}$$

The boundary conditions are given by the continuity of electric and magnetic field components tangential to the core/cladding interfaces as in section 2.2.

Since the guided fields must be evanescent in the cladding and oscillatory in the core, we rewrite the Helmholtz Equation as

$$\text{core} \quad : \quad \frac{d^2}{dx^2}u(x) = -k_x^2 u(x) \text{ with } k_x^2 = \left(k_1^2 - \beta^2\right), \tag{2.334}$$

$$\text{cladding} \quad : \quad \frac{d^2}{dx^2}u(x) = \kappa_x^2 u(x) \text{ with } \kappa_x^2 = \left(\beta^2 - k_2^2\right) \tag{2.335}$$

where κ_x is the decay constant of the evanescent waves in the cladding. It is obvious that for obtaining guided modes, the propagation constant of the mode must be between the two propagation constants for core and cladding

$$k_2^2 < \beta^2 < k_1^2. \tag{2.336}$$

Or by defining an effective index for the mode

$$\beta = k_0 n_{eff}, \text{ with } k_0 = \frac{\omega}{c_0} \tag{2.337}$$

we find

$$n_1 > n_{eff} > n_2, \tag{2.338}$$

and Eqs.(2.334), (2.335) can be rewritten as

$$\text{core} \quad : \quad -\frac{d^2}{dx^2}u(x) - k_0^2\left(n_1^2 - n_{eff}^2\right)u(x) = 0 \tag{2.339}$$

$$\text{cladding} \quad : \quad -\frac{d^2}{dx^2}u(x) + \kappa_0^2\left(n_{eff}^2 - n_2^2\right)u(x) = 0 \tag{2.340}$$

For reasons, which will become more obvious later, we draw in Figure 2.92 the negative refractive index profile of the waveguide.

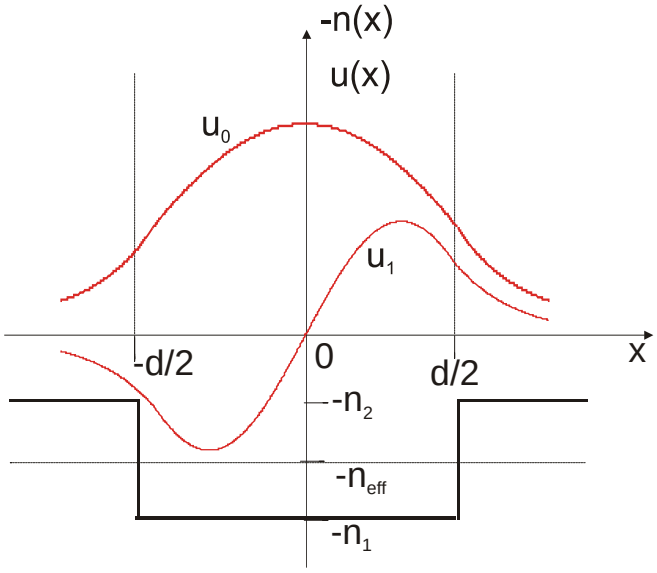

Figure 2.92: Negative refractive index profile and shape of electric field for the fundamental and first higher order transverse TE-mode

From Eq.(2.339) we find that the solution has the general form

$$u(x) = \begin{cases} A\exp(-\kappa_x x) + B\exp(\kappa_x x), & \text{for } x < -d/2 \\ C\cos(k_x x) + D\sin(k_x x), & \text{for } |x| < d/2 \\ E\exp(-\kappa_x x) + F\exp(\kappa_x x), & \text{for } |x| > d/2 \end{cases} \tag{2.341}$$

For a guided wave, i.e. $u_m(x \to \pm\infty) = 0$ the coefficients A and F must be zero. It can be also shown from the symmetry of the problem, that the solutons are either even or odd (proof later)

$$u^{(e)}(x) = \begin{cases} B\exp(\kappa_x x), & \text{for } x < -d/2 \\ C\cos(k_x x), & \text{for } |x| < d/2 \\ E\exp(-\kappa_x x), & \text{for } |x| > d/2 \end{cases}, \tag{2.342}$$

$$u^{(o)}(x) = \begin{cases} B\exp(\kappa_x x), & \text{for } x < -d/2 \\ D\sin(k_x x), & \text{for } |x| < d/2 \\ E\exp(-\kappa_x x), & \text{for } |x| > d/2 \end{cases}. \tag{2.343}$$

The coefficients B and E in each case have to be determined from the boundary conditions. From the continuity of the tangential electric field $\underline{E}_y$, and

the tangential magnetic field $\underline{H}_z$, which follows from Faraday's Law to be

$$\underline{H}_z(x) = \frac{1}{-j\omega\mu_0}\frac{\partial \underline{E}_y}{\partial x} \sim \frac{du}{dx} \tag{2.344}$$

we obtain the boundary conditions for $u(x)$

$$u(x = \pm d/2 + \epsilon) = u(x = \pm d/2 - \epsilon), \tag{2.345}$$

$$\frac{du}{dx}(x = \pm d/2 + \epsilon) = \frac{du}{dx}(x = \pm d/2 - \epsilon). \tag{2.346}$$

Note, these are four conditions determining the coefficients B, D, E and the propagation constant β or refractive index n_{eff}. These conditions solve for the parameters of even and odd modes separately. For the case of the even modes, where $B = E$, we obtain

$$B\exp\left(-\kappa_x\frac{d}{2}\right) = C\cos\left(k_x\frac{d}{2}\right) \tag{2.347}$$

$$B\ \kappa_x\exp\left(-\kappa_x\frac{d}{2}\right) = Ck_x\sin\left(k_x\frac{d}{2}\right) \tag{2.348}$$

or by division of the both equations

$$\kappa_x = k_x\tan\left(k_x\frac{d}{2}\right). \tag{2.349}$$

Eqs.(2.334) and (2.335) can be rewritten as one equation

$$k_x^2 + \kappa_x^2 = \left(k_1^2 - k_2^2\right) = k_0^2\left(n_1^2 - n_2^2\right) \tag{2.350}$$

Eq.(2.349) together with Eq.(2.350) determine the propagation constant β via the two relations.

$$\kappa_x\frac{d}{2} = k_x\frac{d}{2}\tan\left(k_x\frac{d}{2}\right), \text{ and} \tag{2.351}$$

$$\left(k_x\frac{d}{2}\right)^2 + \left(\kappa_x\frac{d}{2}\right)^2 = \left(k_0\frac{d}{2}NA\right)^2 \tag{2.352}$$

where

$$NA = \sqrt{(n_1^2 - n_2^2)} \tag{2.353}$$

is called the numerical apperture of the waveguide. We will discuss the physical significance of the numerical apperture shortly. A graphical solution of these two equations can be found by showing both relations in one plot, see Figure 2.93.

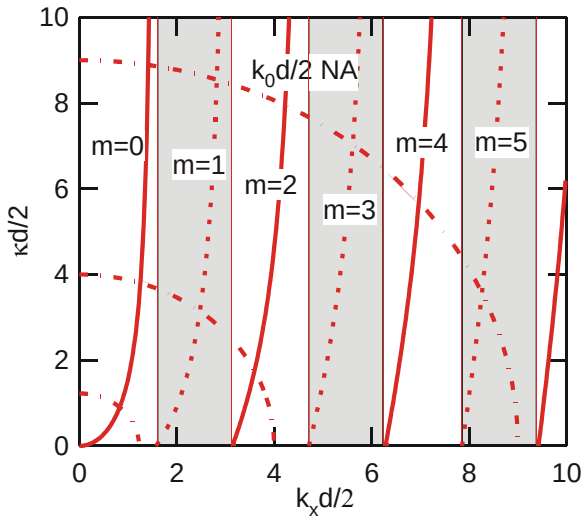

Figure 2.93: Graphical solution of Eqs.(2.351) and (2.352), solid line for even modes and Eq.(2.354) for the odd modes. The dash dotted line shows (2.352) for different values of the product $\left(k_0 \frac{d}{2}\right) NA$

Each crossing in Figure 2.93 of a solid line (2.351) with a circle (2.352) with radius $k_0 \frac{d}{2} NA$ represents an even guided mode. Similarly one finds for the odd modes from the boundary conditions the relation

$$\kappa_x \frac{d}{2} = -k_x \frac{d}{2} \cot\left(k_x \frac{d}{2}\right), \tag{2.354}$$

which is shown in Figure 2.93 as dotted line. The corresponding crossings with the circle indicate the existence of an odd mode.

There are also TM-modes, which we don't want to discuss for the sake of brevity.

Numerical Aperture Figure 2.93 shows that the number of modes guided is determined by he product $k_0 \frac{d}{2} NA$, where NA is the numerical apperture

defined in Eq.(2.353)

$$M = Int\left[k_0\frac{d}{2}NA/(\pi/2)\right]+1, \tag{2.355}$$

$$= Int\left[2\frac{d}{\lambda_0}NA\right]+1, \tag{2.356}$$

where the function $Int[x]$means the largest integer not greater than x. Note, that there is always at least one guided mode no matter how small the sized and the refractive index contrast between core and cladding of the waveguide is. However, for small size and index contrast the mode may extend very far into the cladding and the confinement in the core is low.

The numberical apperture also has an additional physical meaning that becomes obvious from Figure 2.94.

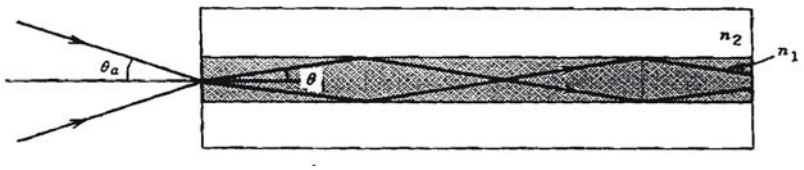

Figure 2.94: Maximum angle of incoming wave guided by a waveguide with numerical apperture NA, [6], p. 262.

The maximum angle of an incoming ray that can still be guided in the waveguide is given by the numerical apperture, because according to Snell's Law

$$n_0 \sin(\theta_a) = n_1 \sin(\theta), \tag{2.357}$$

where n_0 is the refractive index of the medium outside the waveguide. The maximum internal angle θ where light is still guided in the waveguide by total internal reflection is determined by the critical angle for total internal reflection (2.126) , i.e. $\theta_{\max} = \pi/2 - \theta_{tot}$ with

$$\sin(\theta_{tot}) = \frac{n_2}{n_1}. \tag{2.358}$$

Thus for the maximum angle of an incoming ray that can still be guided we find

$$n_0 \sin(\theta_{a,\max}) = n_1 \sin(\theta_{\max}) = n_1 \sqrt{1 - \left(\frac{n_2}{n_1}\right)^2} = NA. \tag{2.359}$$

Most often the external medium is air with $n_0 \approx 1$ and the refractive index contrast is week, so that $\theta_{a,\max} \ll 1$ and we can replace the sinusoid with its argument, which leads to

$$\theta_{a,\max} = NA. \tag{2.360}$$

Field Distributions Figure 2.95 shows the field distribution for the TE guided modes in a dielectric waveguide. Note, these are solutions of the second order differential equations (2.339) and (2.340) for an effective index n_{eff}, that is between the core and cladding index. These guided modes have a oscillatory behavior in those regions in space where the negative effective index is larger than the negative local refractive index, see Figure 2.92 and exponentially decaying solutions where the negative effective index is smaller than than the negative local refractive index.

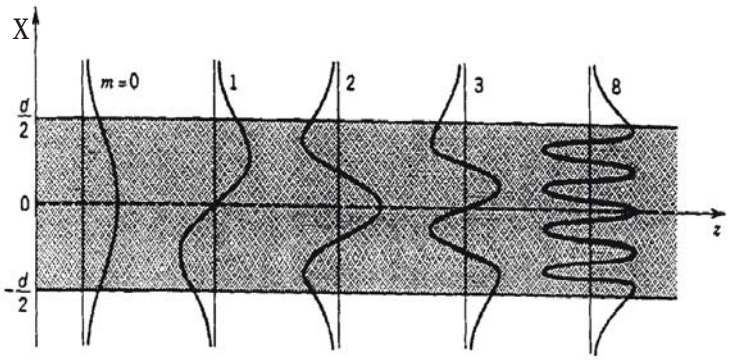

Figure 2.95: Field distributions for TE guided modes in a dielectric waveguide. These results should be compared with those shown in Figure 2.87 for the planar-mirror waveguide [6], p. 254.

Figure 2.96 shows a comparison of the guided modes in a waveguide with a Gaussian beam. In contrast to a the Gaussian beam which diffracts,

in a waveguide diffraction is balanced by the guiding action of the index discontinuity, i.e. total internal reflection. Most importantly the cross section of a waveguide mode stays constant and therefore a waveguide mode can efficiently interact with the medium constituting the core or a medium that is incorporated in the core.

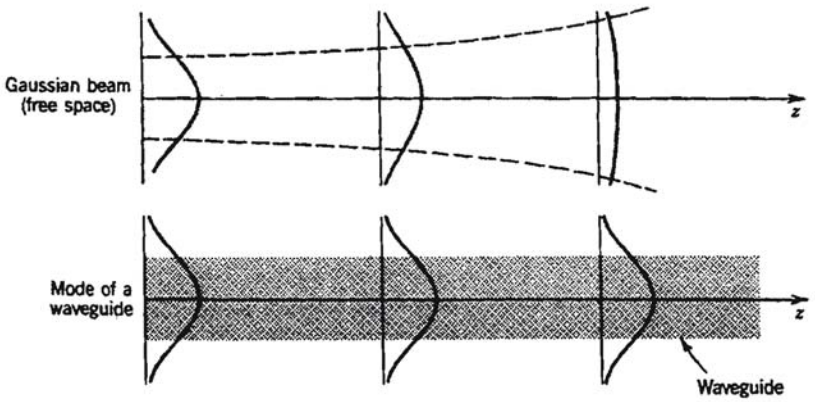

Figure 2.96: Comparison of Gaussian beam in free space and a waveguide mode, [6], p. 255.

Besides integration, this prolong interaction disctance is one of the major reasons for using waveguides. The interaction lenght can be arbitrarily long, only limited by the waveguide loss, in contrast to a Gaussian beam, which stays focused only over the confocal distance or Rayleigh range.

As in the case of a planar-mirror waveguide, one can show that the transverse mode functions are orthogonal to each other. At first, a striking difference here is that we have only a finite number of guided modes and one might worry about the completeness of the transverse mode functions. The answer is that in addition to the guided modes, there are unguided modes or leaky modes, which together with the guided modes from a complete set. Each initial field can be decomposed into these modes. The leaky modes rapidly loose energy because of radiation and after a relatively short propagation distance only the field of guided modes remains in the waveguide. We will not pursue this further in this introductory class. The interested reader should consult with [11].

Confinement Factor

A very important quantity for a waveguide mode is its confinement in the core, which is called the confinement factor

$$\Gamma_m = \frac{\int_0^{d/2} u_m^2(x)\ dx}{\int_0^{\infty} u_m^2(x)\ dx}. \tag{2.361}$$

The confinement factor quantifies the fraction of the mode energy propagating in the core of the waveguide. This is very important for the interaction of the mode with the medium of the core, which may be used to amplify the mode or which may contain nonlinear media for frequency conversion.

Waveguide Dispersion

For the guided modes the effective refractive indices of the modes and therefore the dispersion relations must be between the indices or dispersion relations of core and cladding, see Figure 2.97

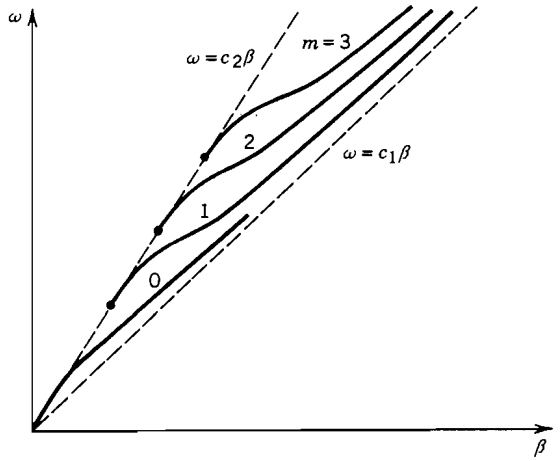

Figure 2.97: Dispersion relations for the different guided TE-Modes in the dielectric slab waveguide.

The different slopes $d\omega/d\beta$ for each mode indicate the difference in group velocity between the modes. Note, that there is at least always one guided mode.

2.7.2 Two-Dimensional Waveguides

Both the planar-mirror waveguide and the planar dielectric slab waveguide confine light only in one direction. It is straight forward to analyze the modes of the two-dimensional planar-mirror waveguide, which you have already done in 6.013. Figure 2.98 shows various waveguides that are used in praxis for various devices. Here, we do not want to analyze them any further, because this is only possible by numerical techniques.

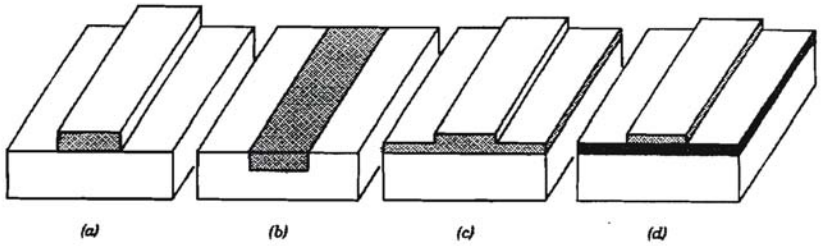

Figure 2.98: Various types of waveguide geometries: (a) strip: (b) embedded strip: (c) rib ro ridge: (d) strip loaded. The darker the shading, the higher the refractive index [6], p. 261.

2.7.3 Waveguide Coupling

The core size of a waveguide can range from a fraction of the free space wavelength to many wavelength for a multimode fiber. For example a typical high-index contrast waveguide with a silicon core and a silica cladding for 1550 nm has a cross section of $0.2\mu m \times 0.4\mu m$, single-mode fiber, which we will discuss in the next section with an index contrast of 0.5-1% between core and cladding has a typical mode-field radius of $6\mu m$.

If the mode cross section is not prohibitively small the simplest approach to couple light into a waveguide is by using a proper lens, see Figure 2.99 (a) or direct butt coupling of the source to the waveguide if the source is a waveguide based device itself.

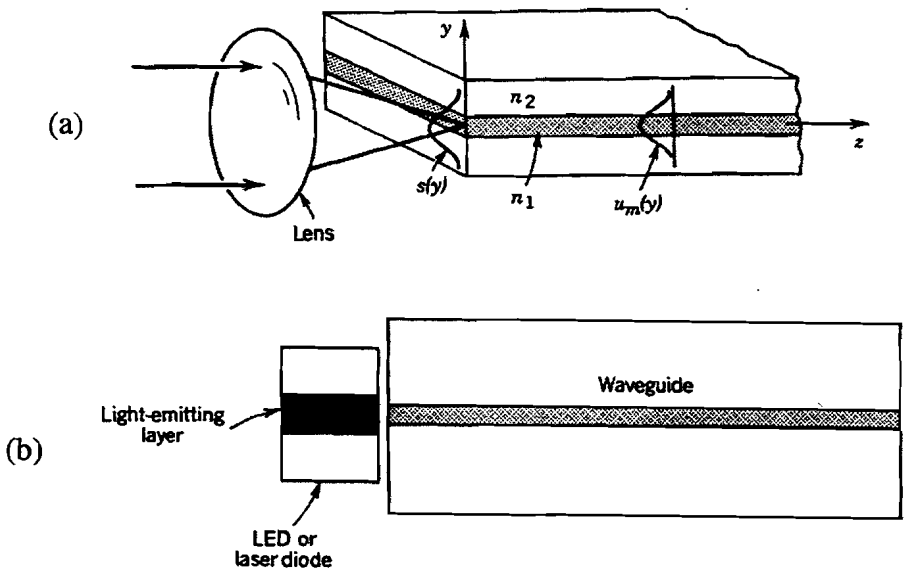

Figure 2.99: Coupling to a waveguide by (a) a lens; (b) direct butt coupling of an LED or laser diode, [6], p. 262

The lens and the beam size in free space must be chosen such that the spot size matches the size of the waveguide mode while the focusing angle in free space is less than the numerical aperture of the waveguide, (see problem set). Other alternatives are coupling to the evanscent field by using a prism coupler, see Figure 2.100

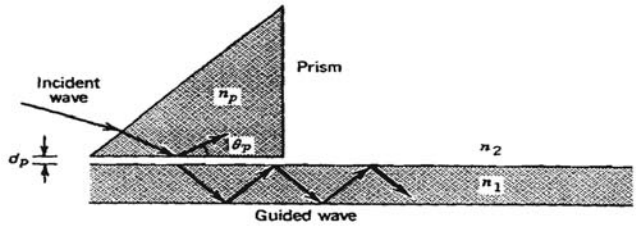

Figure 2.100: Prism coupler, [6], p. 263.

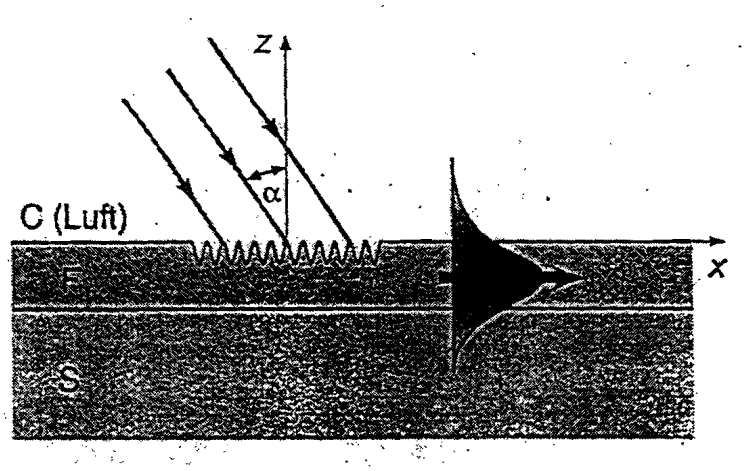

Figure 2.101: Grating Coupler

The coupling with the prism coupler is maximum if the propagation constant of the waveguide mode matches the longitudinal component of the k-vector

$$\beta = kn_p \cos\theta_p,$$

Another way to match the longitudinal component of the k-vector of the incoming light to the propagation constant of the waveguide mode is by a grating coupler, see Figure 2.101

2.7.4 Coupling of Modes

If two dielectric waveguides are placed closely together their fields overlap. This situation is shown in Figure 2.102 at the example of the planar dielectric slab waveguide. Of course this situation can be achieved with any type of two dimensional dielectric waveguide shown in Figure 2.98

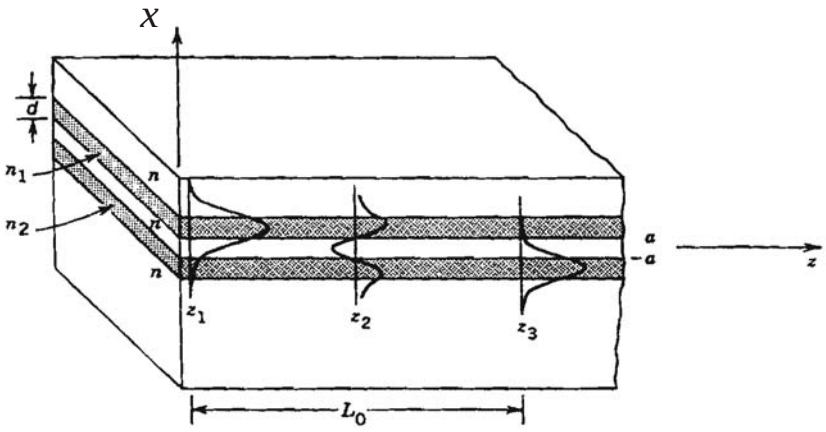

Figure 2.102: Coupling between the two modes of the dielectric slab waveguide, [6], p. 264.

Once the fields significantly overlap the two modes interact. The shape of each mode does not change very much by the interaction. Therefore, we can analyze this situation using perturbation theory. We assume that in zero-th order the mode in each waveguide is independent from the presence of the other waveguide. We consider only the fundamental TE-modes in each of the waveguide which have excitation amplitudes $\underline{a}_1(z)$ and $\underline{a}_2(z)$, respectively. The dynamics of each mode can be understood in terms of this wave amplitude. In the absence of the second waveguide, each waveguide amplitude undergoes only a phase shift during propagation according to its dispersion relations

$$\frac{d\underline{a}_1(z)}{dz} = -j\beta_1\underline{a}_1(z), \tag{2.362}$$

$$\frac{d\underline{a}_2(z)}{dz} = -j\beta_2\underline{a}_2(z). \tag{2.363}$$

The polarization generated by the field of mode 2 in waveguide 1 acts as a source for the field in waveguide 1 and the other way arround. Therefore, the coupling of the modes can be described by adding a source term in each equation proportional to the free propagation of the corresponding wave in

the other guide

$$\frac{d\underline{a}_1(z)}{dz} = -j\beta_1\underline{a}_1(z) - j\kappa_{12}\underline{a}_2(z), \tag{2.364}$$

$$\frac{d\underline{a}_2(z)}{dz} = -j\kappa_{21}\underline{a}_1(z) - j\beta_2\underline{a}_2(z). \tag{2.365}$$

κ_{12} and κ_{21} are the coupling constants of the modes. An expression in terms of waveguide properties is derived in the appendix. These coupled mode equations describe a wealth of phenomena and are of fundamental importance in many areas.

As we will see, there is only a significant interaction of the two modes if the two propagation constants are not much different from each other (phase matching). Therefore, we write the propagation constants in terms of the average β_0 and the phase mismatch $\Delta\beta$

$$\beta_{1/2} = \beta_0 \pm \Delta\beta \text{ with} \tag{2.366}$$

$$\beta_0 = \frac{\beta_1 + \beta_2}{2} \text{ and } \Delta\beta = \frac{\beta_1 - \beta_2}{2}. \tag{2.367}$$

and we take the overall trivial phase shift of both modes out by introducing the slowly varying relative field amplitudes

$$\tilde{\underline{a}}_1(z) = \underline{a}_1(z)e^{j\beta_0 z} \text{ and } \tilde{\underline{a}}_2(z) = \underline{a}_2(z)e^{j\beta_0 z} \tag{2.368}$$

which obey the equation

$$\frac{d}{dz}\tilde{\underline{a}}_1(z) = -j\Delta\beta\tilde{\underline{a}}_1(z) - j\kappa_{12}\tilde{\underline{a}}_2(z), \tag{2.369}$$

$$\frac{d}{dz}\tilde{\underline{a}}_2(z) = -j\kappa_{21}\tilde{\underline{a}}_1(z) + j\Delta\beta\tilde{\underline{a}}_2(z). \tag{2.370}$$

Power conservation during propagation demands

$$\frac{d}{dz}\left(|\tilde{\underline{a}}_1(z)|^2 + |\tilde{\underline{a}}_2(z)|^2\right) = 0 \tag{2.371}$$

which requests that $\kappa_{21} = \kappa_{12}^2$, i.e. the two coupling coefficients are not independent from each other (see problem set).

Note, Eqs.(2.369) and (2.370) are a system of two linear ordinary differential equations with constant coefficients, which is straight forward to solve.

Given the excitation amplitudes $\tilde{\underline{a}}_1(0)$ and $\tilde{\underline{a}}_2(0) = 0$ at the input of the waveguides, i.e. no input in waveguide 2 the solution is

$$\tilde{\underline{a}}_1(z) = \tilde{\underline{a}}_1(0)\left(\cos\gamma z - j\frac{\Delta\beta}{\gamma}\sin\gamma z\right), \tag{2.372}$$

$$\tilde{\underline{a}}_2(z) = -j\tilde{\underline{a}}_1(0)\frac{\kappa_{21}}{\gamma}\sin\gamma z, \tag{2.373}$$

with

$$\gamma = \sqrt{\Delta\beta^2 + |\kappa_{12}|^2}. \tag{2.374}$$

The optical powers after a propagation distance z in both waveguides are then

$$P_1(z) = |\tilde{\underline{a}}_1(z)|^2 = P_1(0)\left(\cos^2\gamma z + \left(\frac{\Delta\beta}{\gamma}\right)^2\sin^2\gamma z\right), \tag{2.375}$$

$$P_2(z) = P_1(0)\left(\frac{|\kappa_{21}|^2}{\gamma}\right)^2\sin^2\gamma z. \tag{2.376}$$

This solution shows, that depending on the difference in phase velocity between the two-waveguides more or less power is coupled back and fourth between the two waveguides, see Figure 2.103.

The period at which the power exchange occurs is

$$L = \frac{2\pi}{\gamma}. \tag{2.377}$$

If both waveguides are identical, i.e. $\Delta\beta = 0$ and $\gamma = |\kappa_{12}|$, the waves are phase matched, Eqs.(2.375) and (2.376) simplify to

$$P_1(z) = P_1(0)\cos^2\gamma z \tag{2.378}$$

$$P_2(z) = P_1(0)\sin^2\gamma z. \tag{2.379}$$

Complete transfer of power occurs between the two waveguides after a distance

$$L_0 = \frac{\pi}{2\gamma}, \tag{2.380}$$

see Figure 2.104

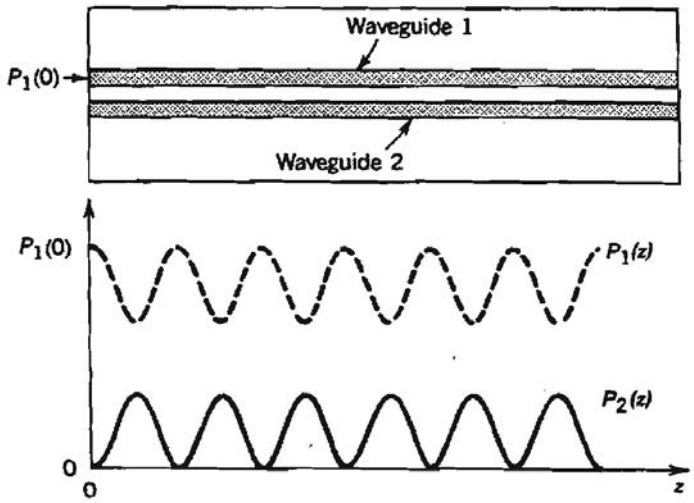

Figure 2.103: Periodic exchange of power between guides 1 and 2 [6], p. 266.

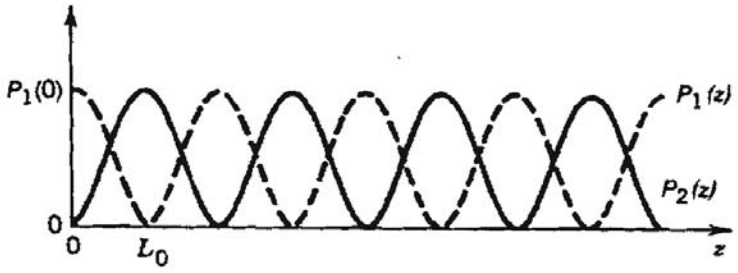

Figure 2.104: Exchange of power between guides 1 and 2 in the phase-matched case, [6], p. 266.

Depending on the length of the coupling region the coupling ratio can be chosen. A device with a distance $L_0/2$ and L_0 achieves 50% and 100% power transfer into waveguide two, respectively, see Figure 2.105

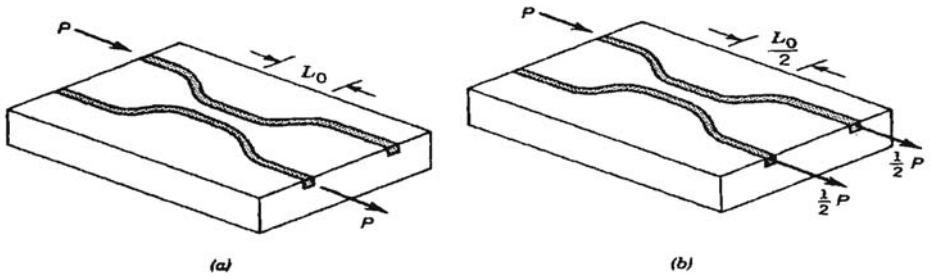

Figure 2.105: Optical couplers: (a) 100% coupler, (b) 3dB coupler, [6], 267.

2.7.5 Switching by Control of Phase Mismatch

If we keep the interaction length of the waveguides fixed at a length L_0, then the power tranfer from waveguide 1 to waveguide 2 depends critically on the phase mismatch $\Delta\beta$

$$T(\Delta\beta) = \frac{P_2}{P_1} = \left(\frac{\pi}{2}\right)^2 \sin c^2 \left(\frac{1}{2}\sqrt{1 + \left(\frac{2\Delta\beta L_0}{\pi}\right)^2}\right), \tag{2.381}$$

where $sinc(x) = \sin(\pi x)/(\pi x)$. Figure 2.106 shows the transfer characteristic as a function of normalized phase mismatch. The phase mismatch between waveguides can be controlled for example by the linear electro-optic or Pockels effect, which we will investigate later.

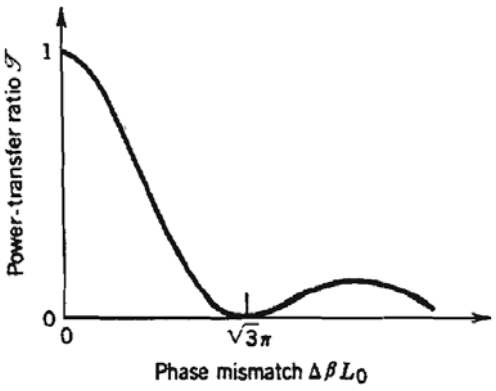

Figure 2.106: Dependence of power transfer from waveguide 1 to waveguide 2 as a function of phase mismatch, [6], p. 267.

The implementation of such a waveguide coupler switch is shown in Figure 2.107.

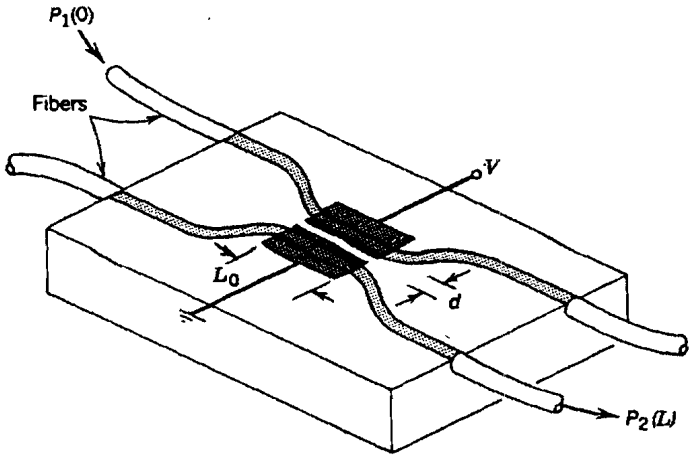

Figure 2.107: Integrated waveguide coupler switch, [6], p. 708

2.7.6 Optical Fibers

Optical fibers are cylindrical waveguides, see Figure 2.108, made of low-loss materials such as silica glass.

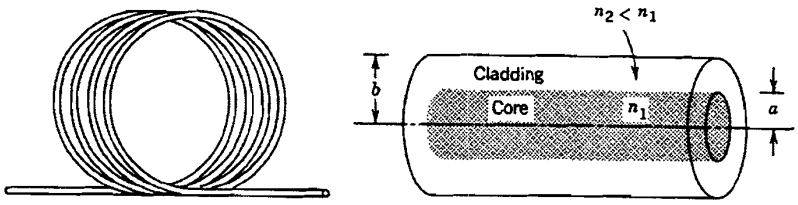

Figure 2.108: Optical fibers are cylindrical dielectric waveguides, [6], p. 273.

Similar to the waveguides studied in the last section the most basic fibers consist of a high index core and a lower index cladding. Today fiber technology is a highly developed art which has pushed many of the physical parameters of a waveguide to values which have been thought to be impossible a few decades ago:

- Fiber with less than 0.16dB/km loss
- Photonic crystal fiber (Nanostructured fiber)
- Hollow core fiber
- Highly nonlinear fiber
- Er-doped fiber for amplifiers
- Yb-doped fiber for efficient lasers and amplifiers
- Raman gain fiber
- Large area single mode fibers for high power (kW) lasers.

Figure 2.109 shows the ranges of attenuation coefficients of silica glass single-mode and multimode fiber.

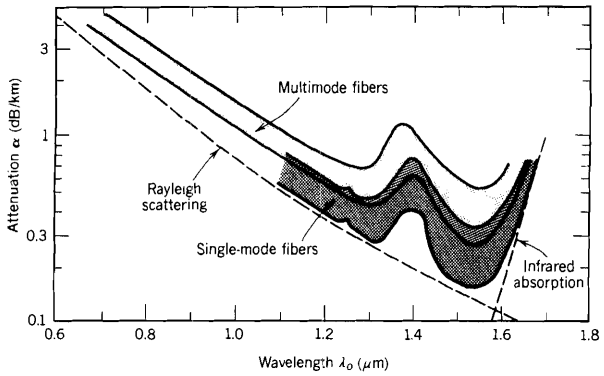

Figure 2.109: Ranges of attenuation coefficients of silica glass single-mode and multimode fiber, [10], p. 298.

For the purpose of this introductory class we only give an overview about the mode structure of the most basic fiber, the step index fiber, see Figure 2.110 (b)

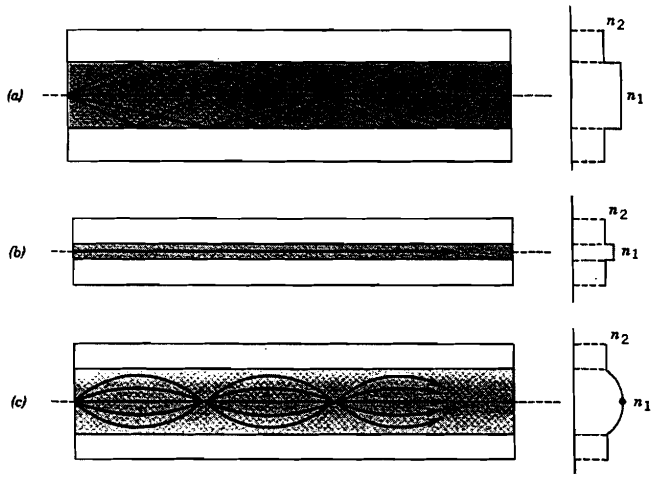

Figure 2.110: Geometry, refractive index profile, and typical rays in: (a) a multimode step-index fiber, (b) a single-mode step-index fiber, (c) a multimode graded-index fiber [6], p. 274

Step-index fiber is a cylindrical dielectric waveguide specified by its core and cladding refractive indices, n_1 and n_2 and the core radius a, see Figure 2.108. Typically the cladding is assumed to be so thick that the finite cladding radius does not need to be taken into account. The guided modes need to be sufficiently decaded before reaching the cladding boundary, which is usually strongly scattering or absorbing. In standard fiber, the cladding indices differ only slightly, so that the relative refractive-index difference

$$\Delta = \frac{n_1 - n_2}{n_1} \tag{2.382}$$

is small, typically $10^{-3} < \Delta < 2 \cdot 10^{-2}$. Most fibers currently used in medium to long optical communication systems are made of fused silica glass (SiO_2) of high chemical purity. The increase in refractive index of the core is achieved by doping with titanium, germanium or boron, among others. The refractive index n_1 ranges from 1.44 to 1.46 depending on the wavelength utilized in the fiber. The acceptance angle of the rays coupling from free space into guided modes of the waveguide is determined by the numerical apperture as already discussed for the dielectric slab waveguide, see Figure 2.111

$$\theta_a \sim \sin(\theta_a) = NA = \sqrt{n_1^2 - n_2^2} \approx n_1\sqrt{2\Delta}. \tag{2.383}$$

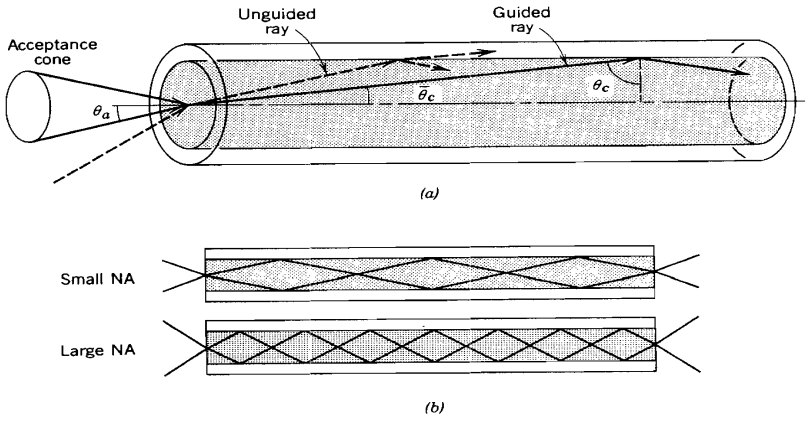

Figure 2.111: The acceptance angle of a fiber and numerical aperture NA [6], p. 276.

Guided Waves

Again the guided waves can be found by looking at solutions of the Helmholtz equations in the core and cladding where the index is homogenous and by additionally requesting the continuity of the tangential electric and magnetic fields at the core-cladding boundary. In general the fiber modes are not anylonger pure TE or TM modes but rather are hybrid modes, i.e. the modes have both transverse and longitudinal electric and magnetic field components. Only the radial symmetric modes are still TE or TM modes. To determine the exact mode solutions of the fiber is beyond the scope of this class and the interested reader may consult reference [2]. However, for weakly guiding fibers, i.e. $\Delta \ll 1$, the modes are actually very much TEM like, i.e. the longitudinal field components are much smaller than the radial field components. The linear in x and y directions polarized modes form orthogonal polarization states. The linearly polarized (l, m) mode is usually denoted as the LP_{lm}-mode.The two polarizations of the mode with indices (l, m) travel with the same propagation constant and have identical intensity distributions.

The generic solutions to the Helmhotz equation in cylindrical coordinates are the ordinary, $J_m(kr)$, and modified, $K_m(kr)$, Bessel functions (analogous to the $\cos(x)/\sin(x)$ and exponential functions $e^{\pm\kappa x}$, that are solutions to the Helmholtz equation in cartesian coordinates). Thus, a generic mode function for a cylinder symmetric fiber has the form

$$u_{l,m}(r,\varphi) = \begin{cases} J_l(k_{l,m}r) \begin{cases} \cos(l\varphi) \\ \sin(l\varphi) \end{cases}, \text{ for } r < a, \text{ core} \\ K_l(k_{l,m}r) \begin{cases} \cos(l\varphi) \\ \sin(l\varphi) \end{cases}, \text{ for } r > a, \text{ cladding} \end{cases} \tag{2.384}$$

For large r, the modified Bessel function approaches an exponential, $K_l(k_{l,m}r) \sim e^{-\kappa_{\ln m} r} \begin{cases} \cos(l\varphi) \\ \sin(l\varphi) \end{cases}$. The propagation constants for this two dimensional waveguide have to fullfil the additional constraints

$$k_{l,m}^2 = \left(n_1^2 k_0^2 - \beta^2\right), \tag{2.385}$$

$$\kappa_{l,m}^2 = \left(\beta^2 - n_2^2 k_0^2\right), \tag{2.386}$$

$$k_{l,m}^2 + \kappa_{l,m}^2 = k_0^2 N A^2. \tag{2.387}$$

Figure 2.112 shows the radial dependence of the mode functions

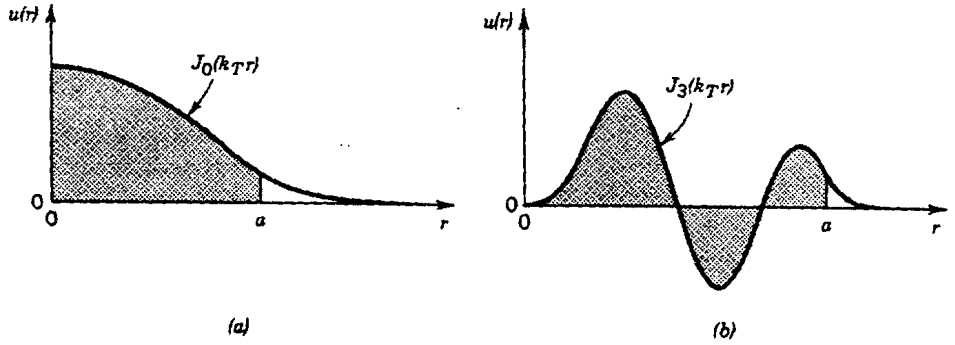

Figure 2.112: Radial dependence of mode functions $u(r)$,[6], p.279.

The transverse intensity distribution of the linearly polarized $LP_{0,1}$ and $LP_{3,4}$ modes are shown in Figure 2.113.

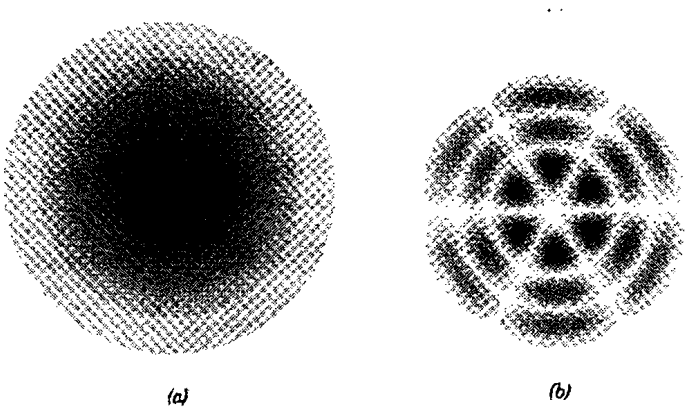

Figure 2.113: Intensity distribtuion of the (a) LP_{01} and (b) $LP_{3,4}$ modes in the transverse plane. The LP_{01} has a intensity distribution similar to the Gaussian beam, [6], p. 283.

Number of Modes

It turns out, that as in the case of the dielectric slab waveguide the number of guided modes critically depends on the numerical aperture or more precisely

on the V-parameter, see Eq.(2.355)

$$V = k_0 \frac{d}{2} NA. \tag{2.388}$$

Without proof the number of modes is

$$M \approx \frac{4}{\pi^2} V^2, \text{ for } V \gg 1. \tag{2.389}$$

which is similar to Eq.(2.355) for the one-dimensional dielectric slab waveguide, but the number of modes here is now related to the square of the V-parameter, because of the two-dimensional transverse confinement of the modes in the fiber. As in the case of the dielectric waveguide, there is always at least one guided mode (two polarizations). However, the smaller the V-parameter the more the mode extends into the cladding and the guiding properties become weak, i.e. small bending of the fiber may already lead to high loss.

2.8 Wave Propagation in Anisotropic Media

So far we have always assumed that the medium in which the electromagnetic wave propagates is isotropic. This causes the induced polarization to be parallel to the applied electric field. In crystaline materials or materials with microscopic fine structure in general, this is no longer the case. Instead of the simple relation

$$\underline{\vec{P}} = \epsilon_0 \underline{\chi} \cdot \underline{\vec{E}}, \tag{2.390}$$

where the susceptibility is a scalar, the induced polarization may have a general lineare dependence on $\vec{E}$ not necessarily parallel to the applied field

$$\underline{P}_x = \epsilon_0 \left(\underline{\chi}_{xx} \underline{E}_x + \underline{\chi}_{xy} \underline{E}_y + \underline{\chi}_{xz} \underline{E}_z \right), \tag{2.391}$$

$$\underline{P}_y = \epsilon_0 \left(\underline{\chi}_{yx} \underline{E}_x + \underline{\chi}_{yy} \underline{E}_y + \underline{\chi}_{yz} \underline{E}_z \right), \tag{2.392}$$

$$\underline{P}_z = \epsilon_0 \left(\underline{\chi}_{zx} \underline{E}_x + \underline{\chi}_{zy} \underline{E}_y + \underline{\chi}_{zz} \underline{E}_z \right). \tag{2.393}$$

The tensor $\underline{\chi}$ is called the electric susceptibility tensor. As shown in Table 2.7 the crytaline structure determines to a large extend the values of the susceptibility tensor elements or in other words the symmetry properties of

isotropic	$\begin{bmatrix} xx & 0 & 0 \\ 0 & xx & 0 \\ 0 & 0 & xx \end{bmatrix}$	cubic
uniaxial	$\begin{bmatrix} xx & 0 & 0 \\ 0 & xx & 0 \\ 0 & 0 & zz \end{bmatrix}$	Tetragonal Trigonal Hexagonal
biaxial	$\begin{bmatrix} xx & 0 & 0 \\ 0 & yy & 0 \\ 0 & 0 & zz \end{bmatrix}$	Orthorhombic
	$\begin{bmatrix} xx & 0 & xz \\ 0 & yy & 0 \\ xz & 0 & zz \end{bmatrix}$	Monoclinic
	$\begin{bmatrix} xx & xy & xz \\ xy & yy & yz \\ xz & yz & zz \end{bmatrix}$	Triclinic

Table 2.7: Form of the electric susceptibility tensor for various crystal systems.

the crystal reflect themselves in the symmetry properties of the susceptibility tensor.

Elementary algebra tells us that we can choose a new coordinate system with axis x', y', z', such that the susceptibility tensor has diagonal form

$$\underline{P}_{x'} = \epsilon_0 \underline{\chi}_{x'x'} \underline{E}_{x'}, \tag{2.394}$$

$$\underline{P}_{y'} = \epsilon_0 \underline{\chi}_{y'y'} \underline{E}_{y'}, \tag{2.395}$$

$$\underline{P}_{z'} = \epsilon_0 \underline{\chi}_{z'z'} \underline{E}_{z'}. \tag{2.396}$$

These directions are called the principle axes of the crystal. In the following, we consider that the crystal axes are aligned with the principle axes. If a TEM-wave is launched along the $z-$axis with the elecric field polarized along one of the prinicple axes, lets say x, the wave will experience a refractive index

$$n_x^2 = 1 + \underline{\chi}_{xx} \tag{2.397}$$

and the wave will have a phase velocity

$$c = c_0/n_x. \tag{2.398}$$

If on the other hand the wave is polarized along the y-axis it will have a different phase velocity corresponding to n_y. If the wave propagates along the $z-$axis with electric field components along both the x- and y-axis, the wave can be decomposed into the two polarization components. During propagation of the wave the will experience a differential phase shift with respect to each other and the state of polarization may change. Later, this phenomenon will be exploited for the construction of modulators and switches.

2.8.1 Birefringence and Index Ellipsoid

If we consider the propagation of a wave into an arbitrary direction of the crystal it is no longer obvious what the plane wave solution and its phase velocity is. We have

$$\vec{D} = \boldsymbol{\varepsilon}\vec{E} \tag{2.399}$$

with

$$\boldsymbol{\varepsilon} = \varepsilon_0 \begin{bmatrix} \boldsymbol{\varepsilon}_x & 0 & 0 \\ 0 & \boldsymbol{\varepsilon}_y & 0 \\ 0 & 0 & \boldsymbol{\varepsilon}_z \end{bmatrix}. \tag{2.400}$$

Let's assume there are plane wave solutions

$$\underline{\vec{E}} = \underline{\vec{E}}_0 \mathrm{e}^{-j\vec{k}\cdot\vec{r}}$$

then Ampere's and Faraday's law give

$$\vec{k} \times \underline{\vec{H}} = -\omega\boldsymbol{\varepsilon}\underline{\vec{E}}, \tag{2.401}$$

$$\vec{k} \times \underline{\vec{E}} = \omega\mu_0\underline{\vec{H}}, \tag{2.402}$$

resulting in the wave equation

$$\vec{k} \times \vec{k} \times \underline{\vec{E}} = -\omega^2\mu_0\boldsymbol{\varepsilon}\underline{\vec{E}}. \tag{2.403}$$

Note, that the wavevector $\vec{k}$ is orthogonal to the dielectric displacement $\vec{D}$ and the magnetic field $\vec{H}$, but not necessarily to the electric field $\vec{E}$. There is

$$\vec{k} \perp \left(\boldsymbol{\varepsilon}\vec{E} = \vec{D}\right) \perp \vec{B}. \tag{2.404}$$

This situation is reflected in Figure 2.114

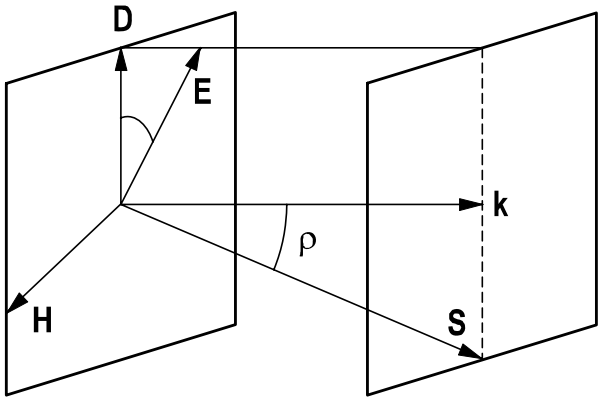

Figure 2.114: Wave propagation in anisotropic media. KDB-system.

One distinguishes between isotropic, uniaxial und biaxial media. We have extensively studied the isotropic case. The most general case is the biaxial case, where the dielectric constants along the three axes are all different. These dielectic constants, or corresponding indices, define an index ellipsoid

$$\frac{x^2}{n_x^2} + \frac{y^2}{n_y^2} + \frac{z^2}{n_z^2} = 1, \tag{2.405}$$

see Figure 2.115.

Here we want to consider the case of an uniaxial crystal, where

$$\boldsymbol{\varepsilon}_{xx} = \boldsymbol{\varepsilon}_{yy} = \boldsymbol{\varepsilon}_1 \neq \boldsymbol{\varepsilon}_{zz} = \boldsymbol{\varepsilon}_3. \tag{2.406}$$

The refractive indices corresponding to these susceptibilities are called ordinary and extraordinary indices

$$n_1 = n_o \neq n_3 = n_e. \tag{2.407}$$

Further, there is a distinction between positive, $n_e > n_o$, and negative, $n_e < n_o$, uniaxial crystals. The uniaxial case corresponds to an index ellipsoid that has rotational symmectry around the z-axis, see Figure 2.115.

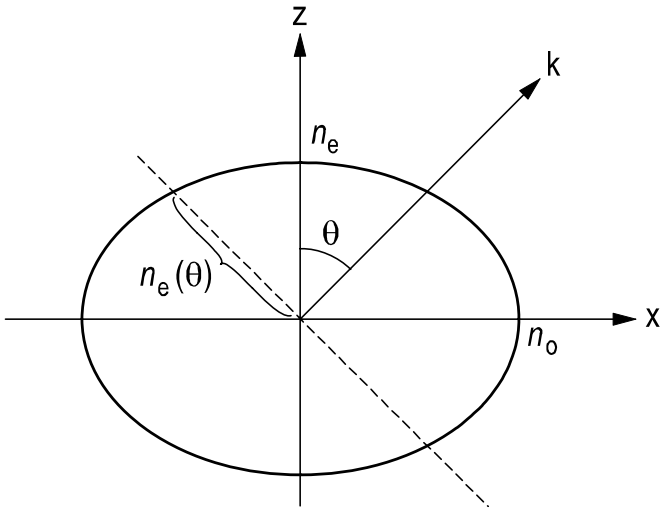

Figure 2.115: Index Ellipsoid

The general case is then a wave with wave vector $\vec{k}$ propagating under an angle θ with respect to the z-axis; the z-axis is also often called the fast axis or c-axis or optical axis. Without restrictions, we assume that the wave vector is in the $x-z-$plane. If the wave vector is aligned with the fast axis, there is no birefringence, because the index experienced by the wave is independent from its polarization. If there is a finite angle, $\theta \neq 0$, then there are two waves with different phase velocity and group velocity as we will show now, see 2.115, and birefringence occurs. With the identity $\vec{A} \times \left(\vec{B} \times \vec{C}\right) = \left(\vec{A} \cdot \vec{C}\right) \vec{B} - \left(\vec{A} \cdot \vec{B}\right) \vec{C}$, when applied to Eq.(2.403), follows

$$\left(\vec{k} \cdot \vec{E}\right) \vec{k} - k^2 \vec{E} + \omega^2 \mu_0 \boldsymbol{\varepsilon} \vec{E} = 0. \tag{2.408}$$

This equation determines the dispersion relation and polarization of the possible waves with wave vector $\vec{k}$. Since the wave vector is in the $x-z-$plane this equation reads

$$\begin{pmatrix} k_0^2 n_o^2 + k_x^2 - k^2 & & k_x k_z \\ & k_0^2 n_o^2 - k^2 & \\ k_z k_x & & k_0^2 n_e^2 + k_z^2 - k^2 \end{pmatrix} \vec{E} = 0 \tag{2.409}$$

This equation clearly shows that a wave polarized along the $y-$axis or in general orthogonal to the plane composed of the wave vector and the fast axis decouples from the other components.

2.8.2 Ordinary Wave

This wave is called the ordinary wave, because it has the dispersion relation

$$k^2 = k_0^2 n_o^2. \tag{2.410}$$

As with the TEM waves in an isotropic medium, the wave vector and the field components build an orthogonal trihedral, $\vec{k} \perp \vec{E} \perp \vec{H}$.

2.8.3 Extraordinary Wave

Eq.(2.409) allows for another wave with a polarization in the $x-z-$plane, and therefore this wave has a longitudinal electric field component. This wave is called extraordinary wave and its dispersion relation follows from

$$\det \begin{vmatrix} k_0^2 n_o^2 + k_x^2 - k^2 & k_x k_z \\ k_z k_x & k_0^2 n_e^2 + k_z^2 - k^2 \end{vmatrix} = \mathbf{0}. \tag{2.411}$$

Calculating the determinant and simplifying we find

$$\frac{k_z^2}{n_o^2} + \frac{k_x^2}{n_e^2} = k_0^2. \tag{2.412}$$

With $k_x = k \sin(\theta)$, $k_z = k \cos(\theta)$ and $k = n(\theta) k_0$ we obtain for the refractive index seen by the extraordinary wave

$$\frac{1}{n(\theta)^2} = \frac{\cos^2(\theta)}{n_o^2} + \frac{\sin^2(\theta)}{n_e^2}. \tag{2.413}$$

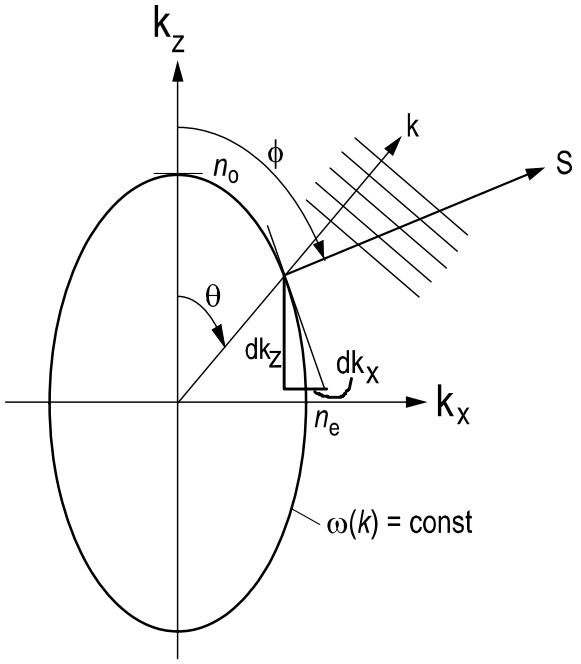

Figure 2.116: Cut through the surface with a constant free space wave number $k_o(k_x, k_y, k_z)$ or frequency, which is also an ellipsoid, but with exchanged principle axis when compared with Figure 2.114

Eqs.(2.412) and (2.413) also describe an ellipse. This ellipse is the location of a constant free space wave number or frequency, $\omega = k_0 c_0$, and therefore determines the refractive index, $n(\theta)$, of the extraordinary wave, see Figure 2.115. The group velocity is found to be parallel to the Poynting vector

$$\boldsymbol{v}_g = \nabla_k \omega(\mathbf{k}) \parallel \mathbf{S}, \tag{2.414}$$

and is orthogonal to the surface. For completeness, we give a derivation of the walk-off angle between the ordinary and extraordinary wave

$$\tan\theta = \frac{k_x}{k_z} \tag{2.415}$$

$$\tan\phi = -\frac{dk_z}{dk_x} \tag{2.416}$$

From Eq.(2.412) we obtain by differentiation along the surface of the ellipsoid

$$\frac{2k_z dk_z}{n_o^2} + \frac{2k_x dk_x}{n_e^2} = 0. \tag{2.417}$$

$$\tan\phi = \frac{n_o^2 k_x}{n_e^2 k_z} = \frac{n_o^2}{n_e^2}\tan\theta$$

Thus, we obtain for the walk-off-angle ϱ between Poynting vector and wave vector

$$\tan\varrho = \tan(\theta - \phi) = \frac{\tan\theta - \tan\phi}{1 + \tan\theta\tan\phi} \tag{2.418}$$

or

$$\tan\varrho = -\frac{\left(\frac{n_0^2}{n_e^2} - 1\right)\tan\theta}{1 + \frac{n_0^2}{n_e^2}\tan^2\theta}. \tag{2.419}$$

2.8.4 Example: Calcite

One example of a birefringent materialis calcite, which is also often used in optical devices, such as polarizers for example. Figure 2.117 and 2.118 show the arrangement of atoms in calcite.

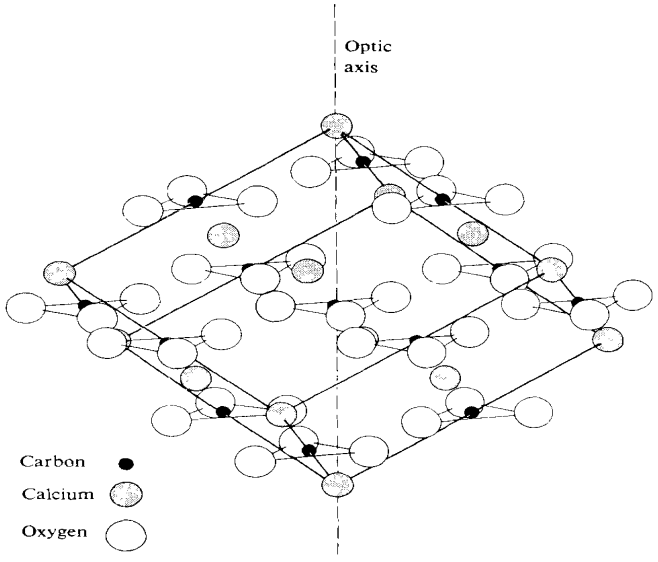

Figure 2.117: Arrangement of atoms in calcite, [1], p. 231.

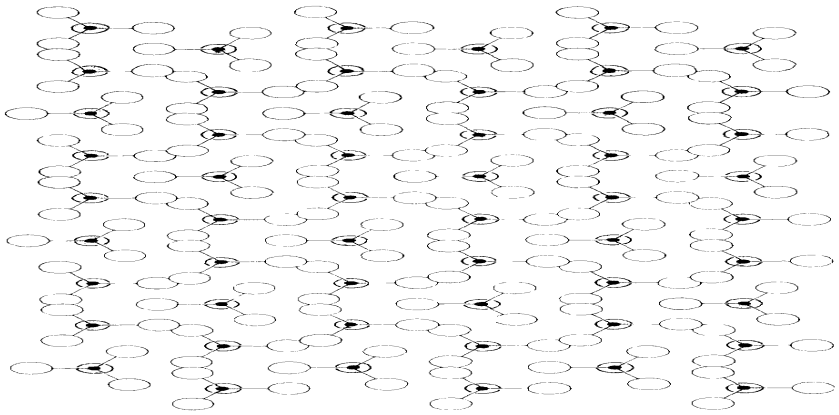

Figure 2.118: Atomic arrangement of calcite looking down the optical axis [1], p. 232.

Figure 2.119 shows a crystal cleaved along the crystal axis (cleavage form).

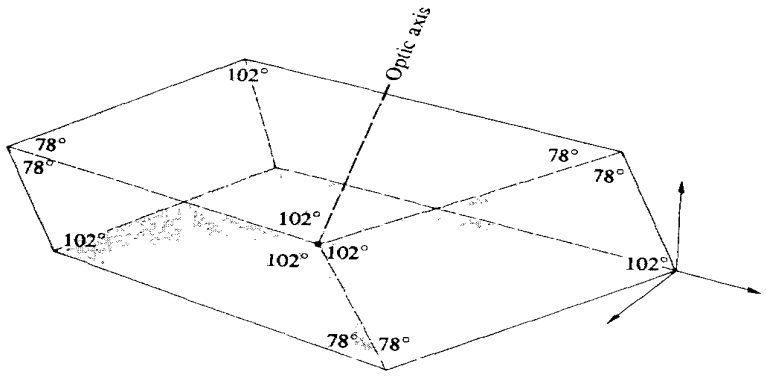

Figure 2.119: Calcite cleavage form [1], p. 232.

Figure 2.120 shows the light path of two orthogonally polarized light beams where one propagates as an ordinary and the other as an extraordinary wave through the crystal. This leads to a double image when an object is viewed through the crystal, see Figure 2.121.

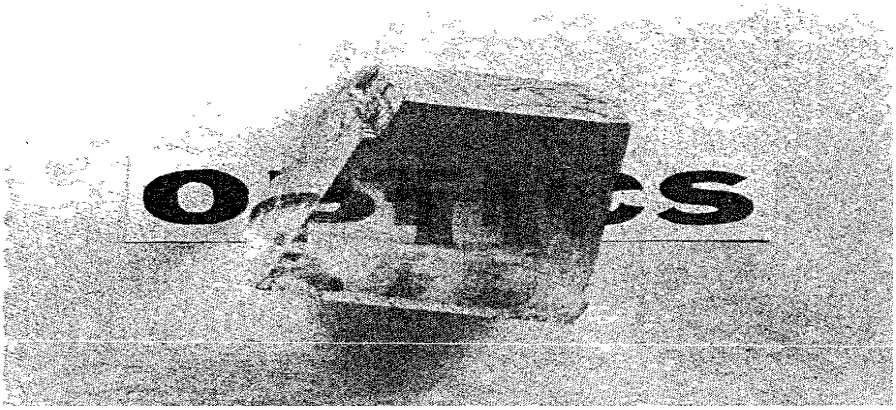

Figure 2.120: A light beam with two orthogonal field components traversing a calcite principal section [1], p. 234.

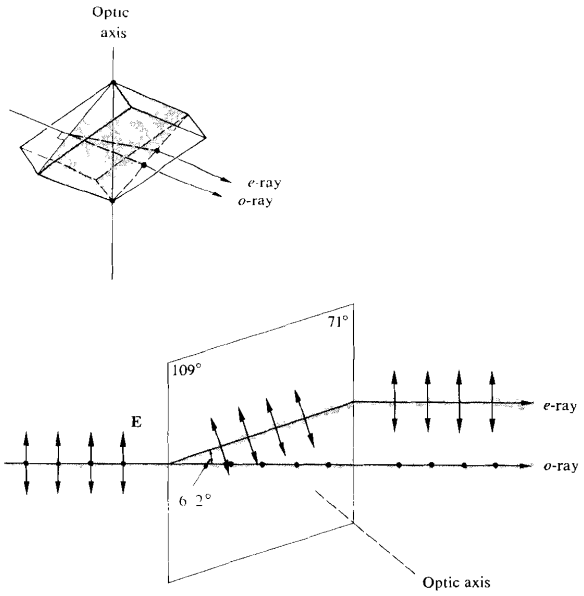

Figure 2.121: Double image formed by a calcite crystal (not cleavage form) [1], p. 233.

Table 2.8 gives the ordinary and extraordinary refractive indices of some

uniaxial crystals. Birefringent materials enable the construction of wave

Crystal	n_o	n_e
Tourmaline	1.669	1.638
Calcite	1.6584	1.4864
Quartz	1.5443	1.5534
Sodim Nitrate	1.5854	1.3369
Ice	1.309	1.313
Rutile (TiO_2)	2.616	1.903

Table 2.8: Refractive indices of some uniaxial birefringent crystals ($\lambda = 589.3$nm) [1], p.236

plates or retardation plates, which enable the manipulation of polarization in a very unique way.

2.9 Polarization and Crystal Optics

So far we have discussed linearly polarized electromagnetic waves, where the electric field of a TEM-wave propagating along the $z-$direction was either polarized along the $x-$ or $y-$axis. The most general TEM-wave has simultaneously electric fields in both polarizations and the direction of the electric field in space, i.e. its polarization, can change during propagation. A description of polarization and polarization evolution in optical systems can be based using Jones vectors and matrices.

2.9.1 Polarization

A general complex TEM-wave propagating along the $z-$direction is given by

$$\vec{\underline{E}}(z,t) = \begin{pmatrix} \underline{E}_{0x} \\ \underline{E}_{0y} \\ 0 \end{pmatrix} e^{\mathrm{j}(\omega t - kz)}, \tag{2.420}$$

where $\underline{E}_{0x} = E_{0x}e^{j\varphi_x}$ and $\underline{E}_{0y} = E_{0y}e^{j\varphi_y}$ are the complex field amplitudes of the $x-$ and $y-$ polarized components of the wave. The real electric field is

given by

$$\vec{E}(z,t) = \begin{pmatrix} E_{0x}\cos(\omega t - kz + \varphi_x) \\ E_{0y}\cos(\omega t - kz + \varphi_y) \\ 0 \end{pmatrix}, \tag{2.421}$$

Both components are periodic functions in $\omega t - kz = \omega(t - z/c)$.

Linear Polarization

If the phases of the complex field amplitudes along the $x-$ and $y-$axis are equal, i.e.

$$\underline{E}_{0x} = |\underline{E}_{0x}|\, e^{j\varphi} \text{ and } \underline{E}_{0y} = \left|\underline{E}_{0y}\right| e^{j\varphi}$$

then the real electric field

$$\vec{E}(z,t) = \begin{pmatrix} \underline{E}_{0x} \\ \underline{E}_{0y} \\ 0 \end{pmatrix} \cos(\omega t - kz + \varphi) \tag{2.422}$$

always oscillates along a fixed direction in the x-y-plane, see Figure 2.122

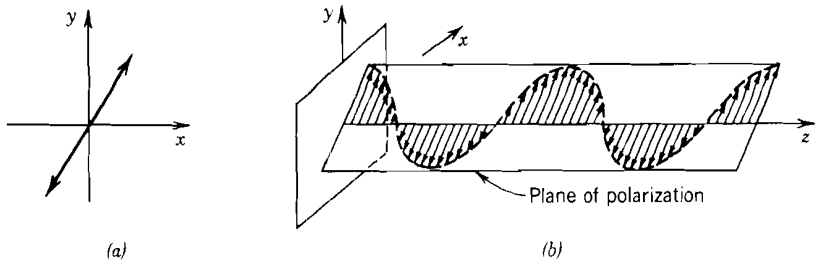

Figure 2.122: Linearly polarized light. (a) Time course at a fixed position z. (b) A snapshot at a fixed time t, [6], p. 197.

The angle between the polarization direction and the x-axis, α, is given by $\alpha = \arctan(E_{0y}/E_{0x})$. If there is a phase difference of the complex field amplitudes along the $x-$ and $y-$axis, the direction and magnitude of the electric field amplitude changes periodically in time at a given position z.

Circular Polarization

Special cases occur when the magnitude of the fields in both linear polarizations are equal $E_{0x} = E_{0y} = E_0$, but there is a phase difference $\Delta\varphi = \pm\frac{\pi}{2}$ in both components. Then we obtain

$$\vec{E}(z,t) = E_0 \,\mathrm{Re}\left\{ \begin{pmatrix} e^{j\varphi} \\ e^{j(\varphi-\Delta\varphi)} \\ 0 \end{pmatrix} e^{\mathrm{j}(\omega t - kz)} \right\} \tag{2.423}$$

$$= E_0 \begin{pmatrix} \cos(\omega t - kz + \varphi) \\ \sin(\omega t - kz + \varphi) \\ 0 \end{pmatrix}. \tag{2.424}$$

For this case, the tip of the electric field vector describes a circle in the $x-y-$plane, as

$$|E_x(z,t)|^2 + |E_y(z,t)|^2 = E_0^2 \text{ for all } z,t, \tag{2.425}$$

see Figure 2.123.

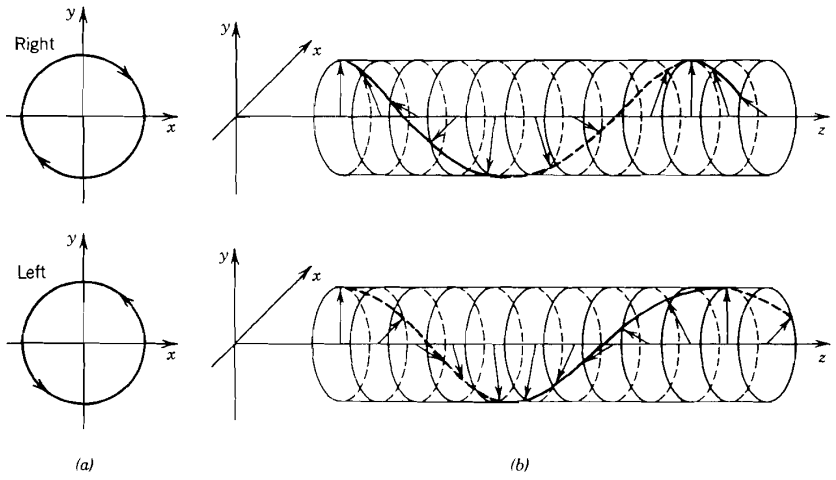

Figure 2.123: Trajectories of the tip of the electric field vector of a right and left circularly polarized plane wave. (a) Time course at a fixed position z. (b) A snapshot at a fixed time t. Note, the sense of rotation in (a) is opposite to that in (b) [6], p. 197.

Right Circular Polarization If the tip of the electric field at a given time, t, rotates counter clockwise with respect to the phase fronts of the wave, here in the positive $z-$direction, then the wave is called right circularly polarized light, i.e.

$$\vec{E}_{rc}(z,t) = E_0 \,\mathrm{Re}\left\{ \begin{pmatrix} 1 \\ j \\ 0 \end{pmatrix} e^{\mathrm{j}(\omega t - kz + \varphi)} \right\} = E_0 \begin{pmatrix} \cos(\omega t - kz + \varphi) \\ -\sin(\omega t - kz + \varphi) \\ 0 \end{pmatrix}. \tag{2.426}$$

A snapshot of the lines traced by the end points of the electric-field vectors at different positions is a right-handed helix, like a right-handed screw pointing in the direction of the phase fronts of the wave, i.e. $k-$vector see Figure 2.123 (b).

Left Circular Polarization If the tip of the electric field at a given fixed time, t, rotates clockwise with respect to the phase fronts of the wave, here in the again in the positive $z-$direction, then the wave is called left circularly polarized light, i.e.

$$\vec{E}_{lc}(z,t) = E_0 \,\mathrm{Re}\left\{ \begin{pmatrix} 1 \\ -j \\ 0 \end{pmatrix} e^{\mathrm{j}(\omega t - kz + \varphi)} \right\} = E_0 \begin{pmatrix} \cos(\omega t - kz + \varphi) \\ \sin(\omega t - kz + \varphi) \\ 0 \end{pmatrix}. \tag{2.427}$$

Eliptical Polarization The general polarization case is called eliptical polarization, as for arbitrary $\underline{E}_{0x} = E_{0x} e^{j\varphi_x}$ and $\underline{E}_{0y} = E_{0y} e^{j\varphi_y}$, we obtain for the locus of the tip of the electric field vector from

$$\vec{E}(z,t) = \begin{pmatrix} E_{0\times} \cos(\omega t - kz + \varphi_x) \\ E_{0y} \cos\left(\omega t - kz + \varphi_y\right) \\ 0 \end{pmatrix}. \tag{2.428}$$

the relations

$$\frac{E_y}{E_{0y}} = \cos\left(\omega t - kz + \varphi_y\right) \tag{2.429}$$

$$= \cos(\omega t - kz + \varphi_x)\cos\left(\varphi_y - \varphi_x\right) - \sin(\omega t - kz + \varphi_x)\sin\left(\varphi_y - \varphi_x\right). \tag{2.430}$$

and

$$\frac{E_x}{E_{0x}} = \cos\left(\omega t - kz + \varphi_x\right). \tag{2.431}$$

These relations can be combined to

$$\frac{E_y}{E_{0y}} - \frac{E_x}{E_{0x}}\cos\left(\varphi_y - \varphi_x\right) = -\sin\left(\omega t - kz + \varphi_x\right)\sin\left(\varphi_y - \varphi_x\right) \tag{2.432}$$

$$\sin\left(\omega t - kz + \varphi_x\right) = \sqrt{1 - \left(\frac{E_x}{E_{0x}}\right)^2} \tag{2.433}$$

Substituting Eq.(2.433) in Eq.(2.432) and building the square results in

$$\left(\frac{E_y}{E_{0y}} - \frac{E_x}{E_{0x}}\cos\left(\varphi_y - \varphi_x\right)\right)^2 = \left(1 - \left(\frac{E_x}{E_{0x}}\right)^2\right)\sin^2\left(\varphi_y - \varphi_x\right). \tag{2.434}$$

After reordering of the terms we obtain

$$\left(\frac{E_x}{E_{0x}}\right)^2 + \left(\frac{E_y}{E_{0y}}\right)^2 - 2\frac{E_x}{E_{0x}}\frac{E_y}{E_{0y}}\cos\left(\varphi_y - \varphi_x\right) = \sin^2\left(\varphi_y - \varphi_x\right). \tag{2.435}$$

This is the equation of an ellipse making an angle α with respect to the x-axis given by

$$\tan 2\alpha = \frac{2E_{0x}E_{0y}\cos\left(\varphi_y - \varphi_x\right)}{E_{0x}^2 - E_{0y}^2}. \tag{2.436}$$

see Figure 2.124.

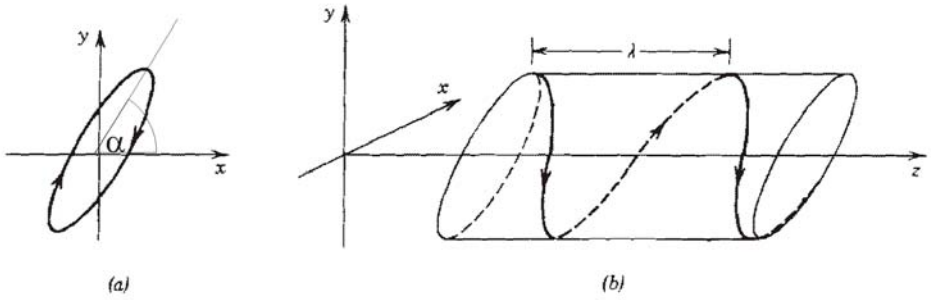

Figure 2.124: (a) Rotation of the endpoint of the electric field vector in the x-y-plane at a fixed position z. (b) A snapshot at a fixed time t [6], p. 197.

Elliptically polarized light can also be understood as a superposition of a right and left cicular polarized light, see Figure 2.125.

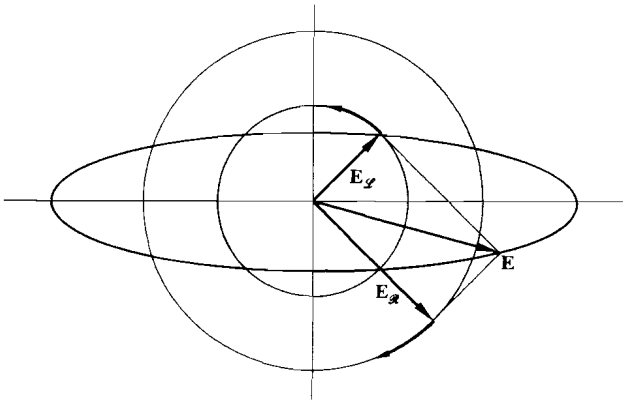

Figure 2.125: Elliptically polarized light as a superposition of right and left circularly polarized light [1], p. 223.

2.9.2 Jones Calculus

As seen in the last section, the information about polarization of a TEM-wave can be tracked by a vector that is proportional to the complex electric-field vector. This vector is called the Jones vector

$$\begin{pmatrix} \underline{E}_{0x} \\ \underline{E}_{0y} \end{pmatrix} \sim \vec{\underline{V}} = \begin{pmatrix} \underline{V}_x \\ \underline{V}_y \end{pmatrix} : \text{Jones Vector} \tag{2.437}$$

Jones Matrix

Figure 2.126 shows a light beam that is normally incident on a retardation plate along the $z-$axis with a polarization state described by a Jones vector

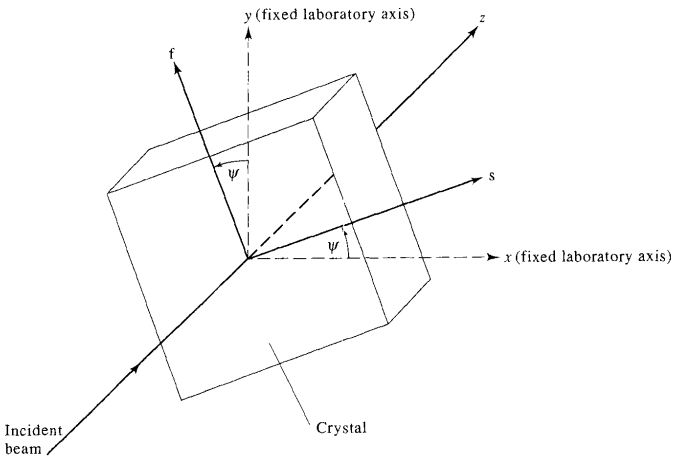

Figure 2.126: A retardation plate rotated at an angle ψ about the z-axis. f("fast") and s("slow") are the two principal dielectric axes of the crystal for light propagating along the $z-$axis [2], p. 17.

The principle axis ($s-$ for slow and $f-$ for fast axis) of the retardation plate are rotated by an angle ψ with respect to the $x-$ and $y-$axis. Let n_s and n_f be the refractive index of the slow and fast principle axis, respectively. The polarization state of the emerging beam in the crystal coordinate system is thus given by

$$\begin{pmatrix} V_s' \\ V_f' \end{pmatrix} = \begin{pmatrix} e^{-jk_o n_s L} & 0 \\ 0 & e^{-jk_o n_f L} \end{pmatrix} \begin{pmatrix} V_s \\ V_f \end{pmatrix}, \tag{2.438}$$

The phase retardation is defined as the phase difference between the two components

$$\Gamma = (n_s - n_f)\, k_o L. \tag{2.439}$$

In birefringent crystals the difference in refractive index is much smaller than the index itself, $|n_s - n_f| \ll n_s, n_f$, therefore parallel to the evolving differential phase a large absolute phase shift occurs. Taking the mean phase shift

$$\phi = \frac{1}{2}(n_s + n_f)\, k_o L, \tag{2.440}$$

out, we can rewrite (2.438) as

$$\begin{pmatrix} V_s' \\ V_f' \end{pmatrix} = e^{-j\phi} \begin{pmatrix} e^{-j\Gamma/2} & 0 \\ 0 & e^{j\Gamma/2} \end{pmatrix} \begin{pmatrix} V_s \\ V_f \end{pmatrix}. \tag{2.441}$$

The matrix connecting the Jones vector at the input of an optical component with the Jones vector at the output is called a Jones matrix.

If no coherent additon with another field is planned at the output of the system, the average phase ϕ can be dropped. With the rotation matrix, R, connecting the (x, y) coordinate system with the (s, f) coordinate system

$$R(\psi) = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix}, \tag{2.442}$$

we find the Jones matrix W describing the propagation of the field components through the retardation plate as

$$\begin{pmatrix} V_x' \\ V_y' \end{pmatrix} = W \begin{pmatrix} V_x \\ V_y \end{pmatrix}. \tag{2.443}$$

with

$$W = R(-\psi) W_0 R(\psi). \tag{2.444}$$

and

$$W_0 = \begin{pmatrix} e^{-j\Gamma/2} & 0 \\ 0 & e^{j\Gamma/2} \end{pmatrix}. \tag{2.445}$$

Carrying out the matix multiplications leads to

$$W = \begin{pmatrix} e^{-j\Gamma/2}\cos^2(\psi) + e^{j\Gamma/2}\sin^2(\psi) & -j\sin\frac{\Gamma}{2}\sin(2\psi) \\ -j\sin\frac{\Gamma}{2}\sin(2\psi) & e^{-j\Gamma/2}\sin^2(\psi) + e^{j\Gamma/2}\cos^2(\psi) \end{pmatrix}. \tag{2.446}$$

Note that the Jones matrix of a wave plate is a unitary matrix, that is

$$W^\dagger W = 1.$$

Unitary matrices have the property that they transform orthogonal vectors into another pair of orthogonal vectors. Thus two orthogonal polarization states remain orthogonal when propagating through wave plates.

Polarizer

A polarizer is a device that absorbs one component of the polarization vector. The Jones matrix of polarizer along the x-axis or y-axis is

$$P_x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \text{ and } P_y = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.447}$$

Half-Wave Plate

A half-wave plate has a phase retardation of $\Gamma = \pi$, i.e. its thickness is $t = \lambda/2(n_e - n_o)$. The corresponding Jones matrix follows from Eq.(2.446)

$$W = -j \begin{pmatrix} \cos(2\psi) & \sin(2\psi) \\ \sin(2\psi) & -\cos(2\psi) \end{pmatrix}. \tag{2.448}$$

For the special case of $\psi = 45^o$, see Figure 2.127, the half-wave plate rotates a linearly polarized beam exactly by 90^0, i.e. it exchanges the polarization axis. It can be shown, that for a general azimuth angle ψ, the half-wave plate will rotate the polarization by an angle 2ψ, see problem set. When the incident light is circularly polarized a half-wave plate will convert right-hand circularly polarized light into left-hand circularly polarized light and vice versa, regardless of the azimuth angle ψ.

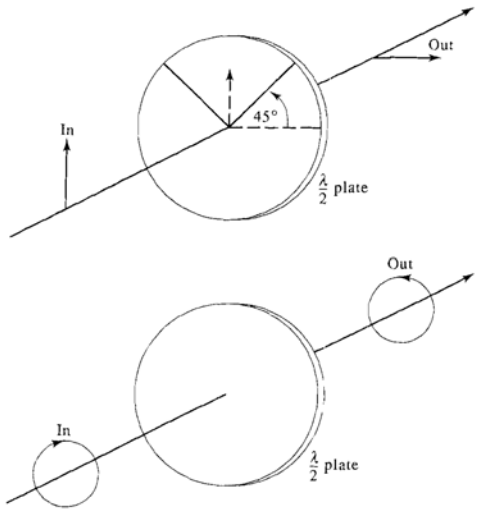

Figure 2.127: The effect of a half-wave plate on the polarziation state of a beam, [2], p.21.

Quarter-Wave Plate

A quarter-wave plate has a phase retardation of $\Gamma = \pi/2$, i.e. its thickness is $t = \lambda/4(n_e - n_o)$. The corresponding Jones matrix follows again from Eq.(2.446)

$$W = \begin{pmatrix} \frac{1}{\sqrt{2}}\left[1 - j\cos(2\psi)\right] & -j\frac{1}{\sqrt{2}}\sin(2\psi) \\ -j\frac{1}{\sqrt{2}}\sin(2\psi) & \frac{1}{\sqrt{2}}\left[1 + j\cos(2\psi)\right] \end{pmatrix}. \tag{2.449}$$

and for the special case of $\psi = 45^o$, see Figure 2.127 we obtain

$$W = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -j \\ -j & 1 \end{pmatrix}, \tag{2.450}$$

see Figure 2.128.

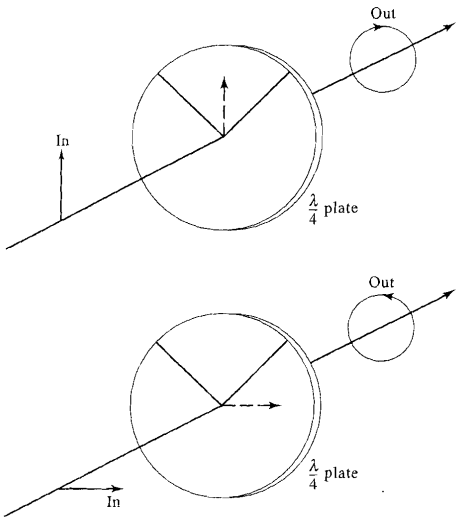

Figure 2.128: The effect of a quarter wave plate on the polarization state of a linearly polarized input wave [2], p.22.

If the incident beam is vertically polarized, i.e.

$$\begin{pmatrix} V_x \\ V_y \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{2.451}$$

the effect of a 45^o -oriented quarter-wave plate is to convert vertically polarized light into left-handed circularly polarized light. If the incident beam is horizontally polarized the outgoing beam is a right-handed circularly polarized, see Figure 2.128.

$$\begin{pmatrix} V'_x \\ V'_y \end{pmatrix} = \frac{-j}{\sqrt{2}} \begin{pmatrix} 1 \\ j \end{pmatrix}. \tag{2.452}$$

Bibliography

[1] Hecht and Zajac, "Optics," Addison and Wesley, Publishing Co., 1979.

[2] B.E.A. Saleh and M.C. Teich, "Fundamentals of Photonics," John Wiley and Sons, Inc., 1991.

[3] Bergmann and Schaefer, "Lehrbuch der Experimentalphysik: Opitk," 1993.

[4] H. Kogelnik and T. Li, "Laser Beams and Resonators," Appl. Opt. **5**, pp. 1550 – 1566 (1966).

[5] H. Kogelnik, E. P. Ippen, A. Dienes and C. V. Shank, "Astigmatically Compensated Cavities for CW Dye Lasers," IEEE J. Quantum Electron. **QE-8**, pp. 373 – 379 (1972).

[6] H. A. Haus, "Fields and Waves in Optoelectronics", Prentice Hall 1984.

[7] F. K. Kneubühl and M. W. Sigrist, "Laser," 3rd Edition, Teubner Verlag, Stuttgart (1991).

[8] A. E. Siegman, "Lasers," University Science Books, Mill Valley, California (1986).

[9] Optical Electronics, A. Yariv, Holt, Rinehart & Winston, New York, 1991.

[10] Photonic Devices, Jia, Ming-Liu , Cambridge University Press, 2005.

[11] T. Tamir, "Guided-Wave Optoelectronics," Springer, 1990.

Chapter 3

Quantum Nature of Light and Matter

We understand classical mechanical motion of particles governed by Newton's law. In the last chapter we examined in some detail the wave nature of electromagnetic fields. We understand the occurance of guided traveling modes and of resonator modes. There are characteristic dispersion relations or resonance frequencies associated with that. In this chapter, we want to summarize some experimental findings at the turn of the 19th century that ultimately lead to the discover of quantum mechanics, which is that matter has in addition to its particle like properties wave properties and electromagnetic waves have in addition to its wave properties particle like properties. As turns out the final theory, which will be developed in subsequent chapters is much more than just that because the quantum mechanical wave function has a different physical interpretation than a electromagnetic wave only the mathematical concepts used is in many cases very similar. However, this is a tremendous help and guideance in doing and finally understanding quantum mechanics.

3.1 Black Body Radiation

In 1900 the physicist Max Planck found the law that governs the emission of electromagnetic radiation from a black body in thermal equilibrium. More specifically Planck's law gives the energy stored in the electromagntic field in a unit volume and unit frequency range, $[f, f+df]$ with $df = 1Hz$, when the

electromagnetic field is in thermal equilibrium with its surrounding that is at temperature T. A black body is simply defined as an object that absorbs all light. The best implementation of a black body is the Ulbricht sphere, see Figure 3.1.

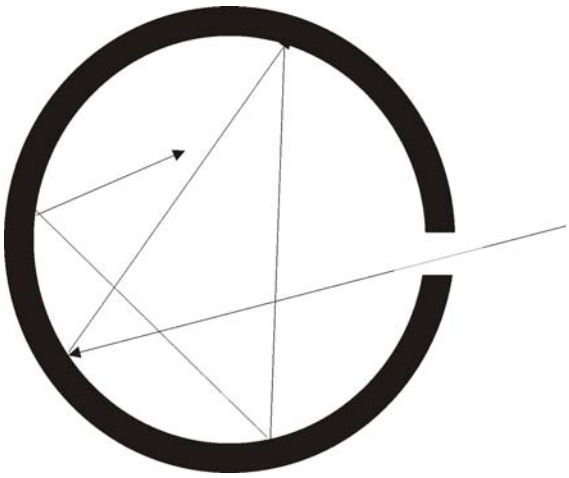

Figure 3.1: The Ulbricht sphere, is a sphere with a small opening, where only a small amount of radiation can escape, so that the interior of the sphere is in thermal equilibrum with the walls, which are kept at a constant tremperature. The inside walls are typically made of diffuse material, so that after multiple scattering of the walls any incoming ray is absorbed, i.e. the wall opening is black.

Figure 3.2 shows the energy density $w(f)$ of electromagnetic radiation in a black body at temperature T. Around the turn of the 19th century $w(f)$ was measured with high precision and one was able to distinguish between various approximations that were presented by other researchers earlier, like the Rayleigh-Jeans law and Wien's law, which turned out to be asymptotic approximations to Planck's Law for low and high frequencies.

In order to find the formula describing the graphs shown in Figure 3.2 Planck had to introduce the hypothesis that harmonic oscillators with frequency f can not exchange arbitrary amounts of energy but rather only in discrete portions, so called quanta. Planck modelled atoms as classical oscillators with frequency f . Therefore, the energy of an oscillator must be quantized in energy levels corresponding to these energy quanta, which he

found to be equal to hf, where h is Planck's constant

$$h = 6.62620 \pm 5 \cdot 10^{-34} Js. \tag{3.1}$$

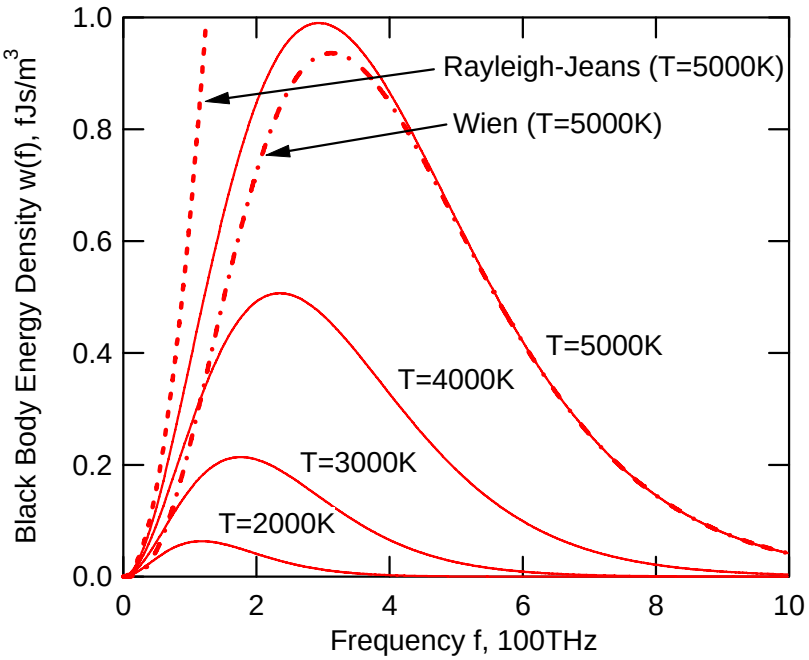

Figure 3.2: Spectral energy density of the black body radiation according to Planck's Law.

As a model for a black body we use now a cavity with perfectly reflecting walls, somewhat different from the Ulbricht sphere. In order to tap of a small but negligible amount of radiation from the inside, a small opening is in the wall. We can make this opening so small that it does essentially not change the internal radiation field. Then the radiation in the cavity is the sum over all possible resonator modes in the cavity. If the cavity is at temperature T all the modes are thermally excited by emission and absorption of energy quanta from the atoms of the wall.

For the derivation of Planck's law we consider a cavity with perfectly conducting walls, see Figure 3.3.

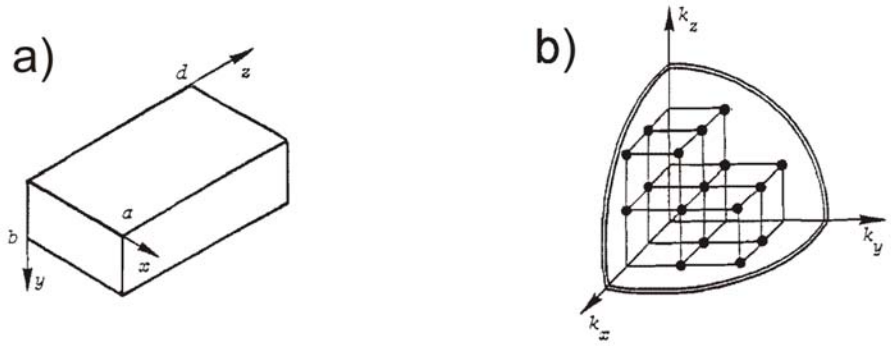

Figure 3.3: (a) Cavity resonator with metallic walls. (b) Resonator modes characterized by a certain k-vector.

If we extend the analysis of the plan parallel mirror waveguide to find the TE and TM modes of a three-dimensional metalic resonator, the resonator modes are $\mathrm{TE}_{mnp}-$ and $\mathrm{TH}_{mnp}-$modes characterized by its wave vector components in $x-, y-$, and $z-$direction. The resonances are standing waves in three dimensions

$$k_x = \frac{m\pi}{L_x},\ k_y = \frac{n\pi}{L_y}, k_z = \frac{p\pi}{L_z},\ \text{for } m, n, p = 0, 1, 2, ... \tag{3.2}$$

An expression for the number of modes in a frequency interval $[f, f + df]$ can be found by recognizing that this is identical to the number of points in Figure 3.3(b) that are in the first octant of a spherical shell with thickness dk at $k = 2\pi f/c$.The volume occupied by one mode in the space of wave numbers k is $\Delta V = \frac{\pi}{L_x} \cdot \frac{\pi}{L_y} \cdot \frac{\pi}{L_z} = \frac{\pi^3}{V}$ with the volume $V = L_x L_y L_z$. Then the number of modes dN in the frequency interval $[f, f + df]$ in volume V are

$$dN = 2 \cdot \frac{4\pi k^2 dk}{8\frac{\pi^3}{V}} = V\frac{k^2 dk}{\pi^2}, \tag{3.3}$$

where the factor of 2 in front accounts for the two polarizations or TE and TH-modes of the resonator and the 8 in the denominator accounts for the fact that only one eighth of the sphere, an octand, is occupied by the positive wave vectors. With $k = 2\pi f/c$ and $dk = 2\pi df/c$, we obtain for the number of modes finally

$$dN = V\frac{8\pi}{c^3}f^2 df \tag{3.4}$$

Note, that the same density of states is obtained using periodic boundary conditions in all three dimensions, i.e. then we can represent all fields in terms of a three dimensional Fourier series. The possible wave vectors would range from negative to positive values

$$k_x = \frac{2m\pi}{L_x},\ k_y = \frac{2n\pi}{L_y},\ k_z = \frac{2p\pi}{L_z} \quad \text{for} \quad m,\, n,\, p = 0,\ \pm 1,\ \pm 2\ldots. \tag{3.5}$$

However, these wavevectors fill the whole sphere and not just one 8-th, which compensates for the 8-times larger volume occupied by one mode. If we imply periodic boundary conditions, we have forward and backward running waves that are independent from each other. If we use the boundary conditions of the resonator, the forward and backward running waves are connected and not independent and form standing waves. One should not be disturbed by this fact as all volume properties, such as the energy density, only depends on the density of states, and not on surface effects, as long as the volume is reasonably large.

3.1.1 Rayleigh-Jeans-Law

The excitation amplitude of each mode obeys the equation of motion of a harmonic oscillator. Therefore, classically one expects that each of mode is in thermal equilibrium excute with a thermal energy kT according to the equipartition theorem, where k is Boltzmann's constant with

$$k = 1.38062 \pm 6 \cdot 10^{-23} J/K. \tag{3.6}$$

If that is the case the spectral energy density is given by the Rayleigh-Jeans-Law, see Figure 3.2.

$$w(f) = \frac{1}{V}\frac{dN}{df}kT = \frac{8\pi}{c^3}f^2 kT. \tag{3.7}$$

As can be seen from Figure 3.2, this law describes very well the black body radiation for frequencies $hf \ll kT$ but there is an arbitrary large deviation for high frequencies. This formula can not be correct, because it predicts infinite energy density for the high frequency modes resulting in an "ultraviolet catastrophy", i.e. the electromagnetic field contains an infinite amount of energy at thermal equilibrium.

3.1.2 Wien's Law

The high frequency or short wavelength region of the black body radiation was first empirically described by Wien's Law

$$w(f) = \frac{8\pi h f^3}{c^3} e^{-hf/kT}. \tag{3.8}$$

Wien's law is surprisingly close to Planck's law, however it slightly fails to correctly predicts the asympthotic behaviour at low frequencies or long wavelengths.

3.1.3 Planck's Law

In the winter of 1900, Max Planck found the correct law for the black body radiation by assuming that each oscillator can only exchange energy in discrete portions or quanta. We rederive it by assuming that each mode can only have the discrete energie values

$$E_s = s \cdot hf, \ \ for \ \ s = 0, 1, 2, ... \tag{3.9}$$

Thus s is the number of energy quanta stored in the oscillator. If the oscillator is a mode of the electromagnetic field we call s the number of photons. For the probability p_s, that the oscillator has the energy E_s we assume a Boltzmann-distribution

$$p_s = \frac{1}{Z} \ \exp\left(-\frac{E_s}{kT}\right) = \frac{1}{Z} \ \exp\left(-\frac{hf}{kT}s\right), \tag{3.10}$$

where Z is a normalization factor such that the total propability of the oscillator to have any of the allowed energy values is

$$\sum_{s=0}^{\infty} p_s = 1. \tag{3.11}$$

Note, due to the fact th$\angle t$ the oscillator energy is proportional to the number of photons, the statistics are exponential statistics. From Eqs.(3.10) and (3.11) we obtain for the normalization factor

$$Z = \sum_{s=0}^{\infty} \exp\left(-\frac{hf}{kT}s\right) = \frac{1}{1 - \ \exp\left(-\frac{hf}{kT}\right)} \ \ , \tag{3.12}$$

which is also called the partition function. The photon statistics are then given by

$$p_s = \exp\left(-\frac{E_s}{kT}\right) \quad \left[1 - \exp\left(-\frac{hf}{kT}\right)\right]^{-1} \tag{3.13}$$

or with $\beta = \frac{hf}{kT}$

$$p_s = \frac{1}{Z(\beta)}\, e^{-\beta s}\ , \text{ with } Z(\beta) = \sum_{s=0}^{\infty} e^{-\beta s} = \frac{1}{1 - e^{-\beta}}. \tag{3.14}$$

Given the statistics of the photon number, we can compute moments of the probability distribution, such as the average number of photons in the mode

$$\left\langle s^1 \right\rangle = \sum_{s=0}^{\infty} s^1 p_s. \tag{3.15}$$

This first moment of the photon statistics can be computed from the partition function, using the "trick"

$$\left\langle s^1 \right\rangle = \frac{1}{Z(\beta)} \frac{\partial^1}{\partial\left(-\beta\right)^1} Z(\beta) \ = \ Z(\beta)\ e^{-\beta}\,, \tag{3.16}$$

which is

$$\langle s \rangle = \frac{1}{\exp\frac{hf}{kT} - 1} \quad . \tag{3.17}$$

With the average photon number $\langle s \rangle$, we obtain for the average energy stored in the mode

$$\langle E_s \rangle = \langle s \rangle\, hf, \tag{3.18}$$

and the energy density in the frequency intervall $[f, f + df]$ is then given by

$$w\left(f\right) = \langle E_s \rangle \frac{dN}{V\, df} \quad . \tag{3.19}$$

With the density of modes from Eq.(3.4) we find Planck's law for the black body radiation

$$w\left(f\right) = \frac{8\pi \ \ f^{\ 2}}{c^3} \quad \frac{hf}{\exp\ \frac{hf}{kT} - 1} \quad , \tag{3.20}$$

which was used to make the plots shown in Figure 3.2. In the limits of low and high frequencies, i.e. $hf \ll kT$ and $hf \gg kT$, respectively Planck's law asympthotically approaches the Rayleigh-Jeans law and Wien's law.

3.1.4 Thermal Photon Statistics

It is interesting to further investigate the intensity fluctuations of the thermal radiation emitted from a black body. If the wall opening in the Ulbricht sphere, see Figure 3.1, is small enough very little radiation escapes through it. If the Ulbricht sphere is kept at constant temperature the radiation inside the Ulbricht sphere stays in thermal equilibrium and the intensity of the radiation emitted from the wall opening in a frequency interval $[f, f + df]$ is

$$I(f) = c \cdot w\,(f)\,. \tag{3.21}$$

Thus the intensity fluctuations of the emitted black body radiation is directly related to the photon statistics or quantum statistics of the radiation modes at freuqency f, i.e. related to the stochastic variable s : the number of photons in a mode with frequency f. This gives us direclty experimental access to the photon statistics of an ensemble of modes or even a single mode when proper spatial and spectral filtering is applied.

Using the expectation value of the photon number 3.17, we can rewrite the photon statistics for a thermally excited mode in terms of its average photon number in the mode as

$$p_s = \frac{\langle s \rangle^s}{(\langle s \rangle + 1)^{s+1}} = \frac{1}{(\langle s \rangle + 1)} \left(\frac{\langle s \rangle}{(\langle s \rangle + 1)} \right)^s , \tag{3.22}$$

The thermal photon statistics display an exponential distribution, see Figure 3.4. Before we move on, lets see how the average photon number in a given mode depends on temperature and the frequency range considered. Figure 3.5 shows the relationship between average number of photons in a mode with frequency f or wavelength λ and temperature T.

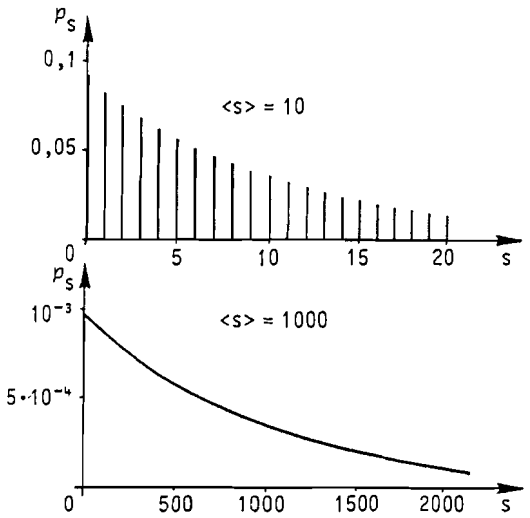

Figure 3.4: Photon statistics of a mode in thermal equilibrium with a mean photon number $< s >= 10$ (a) and $< s >= 1000$ (b).

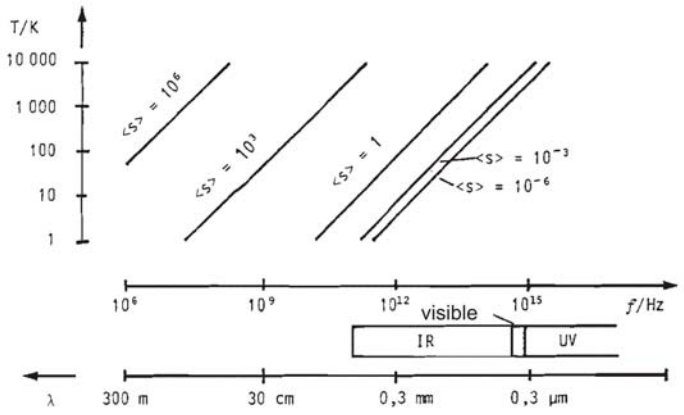

Figure 3.5: Average photon number in a mode at frequency f or wavelength λ and temperature.

Figure 3.5 shows that at room temperature and micorwave frequencies

large numbers of photons are present due to the thermal excitation of the mode. This is the reason that at room temperature the thermal noise overwhelms eventual quantum fluctuations. However, quantum fluctuations are important at high frequencies, which start for room temperature in the far to mid infrared range, where on average much less than one photon is thermally excited.
The variance of the photon number distribution is

$$\sigma_s^2 = \left\langle s^2 \right\rangle - \left\langle s \right\rangle^2 . \tag{3.23}$$

By generalizing Eq.(3.12) to the m-th moment by replacing the exponent 1 by m

$$\left\langle s^m \right\rangle = \sum_{s=0}^{\infty} s^m p_s \tag{3.24}$$

$$= \frac{1}{Z} \frac{\partial^m}{\partial \left(-\beta\right)^m} Z\left(\beta\right), \tag{3.25}$$

we obtain for the second moment

$$\left\langle s^2 \right\rangle = 2Z\left(\beta\right)\ ^2 e^{-2\beta} - Z\left(\beta\right)\ ^2 e^{-2\beta} = 2\left\langle s^2 \right\rangle + \left\langle s \right\rangle . \tag{3.26}$$

and therefore for the variance of the photon number using Eq.(3.23) is

$$\sigma_s^2 = \left\langle s \right\rangle^2 + \left\langle s \right\rangle . \tag{3.27}$$

As expected from the wide distribution of photon numbers the variance is larger than the square of the expectation value. This means that if we look at the light intensity of a single mode the intensity is subject to extremly strong fluctuations as large as the mean value. So why don't we see this rapid thermal fluctuations when we look at the black body radiation coming, for example, from the surface of the sun? Well we don't look at a single mode but rather at a whole multitude of modes. Even when we restrict us to a certain narrow frequency range and spatial direction, there is a multitude of transverse modes presence. We obtain for the average total number of photons in a group of modes and its variance

$$\left\langle s_{tot} \right\rangle = \sum_{i=1}^{\mathbb{N}} \left\langle s_i \right\rangle , \tag{3.28}$$

$$\sigma_{tot}^2 = \sum_{i=1}^{\mathbb{N}} \sigma_i^2 \ . \tag{3.29}$$

Since these modes are independent identical systems, we have

$$\begin{aligned} \langle s_{tot} \rangle &= N \cdot \langle s \rangle \,. \\ \sigma_{tot}^2 &= N \left(\langle s \rangle^2 + \langle s \rangle \right) = \frac{1}{N} \langle s_{tot} \rangle^2 + \langle s_{tot} \rangle \,. \end{aligned} \tag{3.30}$$

Due to the averaging over many modes, the photon number fluctuations in a large number of modes is reduced compared to its mean value

$$SNR = \frac{\sigma_{tot}^2}{\langle s_{tot} \rangle^2} = \frac{1}{N} + \frac{1}{\langle s_{tot} \rangle}. \tag{3.31}$$

Thus if one averages over many modes and has many photons in these modes the intensity fluctuations become small.

3.1.5 Mode Counting

It is interested to estimate the number of modes one is averaging over given a certain emitting surface and a certain measurement time, see Figure 3.6.

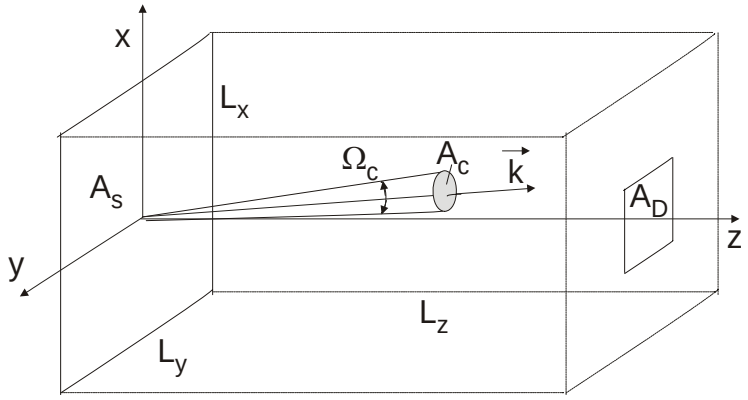

Figure 3.6: Counting of longitudinal and transverse modes excited from a radiating surface of size A_s.

If the area A_s is emitting light, it will couple to the modes of the free field. To count the modes we put a large box (universe) over the experimental arrangement under consideration. The emitting surface is one side of the box. The light from this surface, i.e. specifying the transverse electric and magnetic fields, couples to the modes of the universe with wave vectors according to Eq.(3.5).

Longitudinal Modes

The number of longitudinal modes, that propagate along the positive z-direction in the frequency interval Δf can be derived from $\Delta k = (2\pi/L_z)\,\Delta N$ and $\Delta k = (2\pi/c_0)\,\Delta f$

$$\Delta N = \frac{L_z}{c_0}\Delta f, \tag{3.32}$$

or using the propagation or measurement time over which the experiment extends

$$\tau = L_z/c_0, \tag{3.33}$$

we obtain for the number of longitudinal modes that are involved in the measurement that is carried out over a time intervall τ and a frequency range Δf

$$\Delta N = \tau \Delta f \quad . \tag{3.34}$$

Transverse Modes

The free space modes that arrive at the detector area A_D will not only have wave vectors with a $z-$component, but also transverse components. Lets assume that the detector area is far from the emitting surface, and we consider only the paraxial plane waves. The wave vectors of these waves at a given frequency or free space wave number k_0 can be approximated by

$$\vec{k}_{mn} = \left(\frac{2\pi m}{L_x},\, \frac{2\pi n}{L_y},\, k_0\right) \quad \text{with } m, n = 0,\, 1,\, 2,\, ... \tag{3.35}$$

where m and n are transverse mode indices. Then one mode occupies the volume angle

$$\begin{aligned} \Omega_c &= \frac{4\pi^2}{L_x L_y\, k_0^2}, \\ &= \lambda_0^2/A_s\,. \end{aligned} \tag{3.36}$$

If the modes are thermally excited, the radiation in individual modes is uncorrelated. Therefore, if there is a detector at a distance r then only the field within an area

$$A_c = r^2 \Omega_c \quad , \tag{3.37}$$

is correlated. If the photodetector has an area A_d, then the number of transverse modes detected is

$$N_t = A_d/A_{c.} \tag{3.38}$$

The total number of modes detected is

$$N_{tot} = \frac{A_d}{A_c}\,\tau\Delta f \;\; = \frac{A_d A_s}{r^2\lambda_0^2}\,\tau\Delta f\,. \tag{3.39}$$

Note, that there is perfect symmetry between the area of the emitting and receiving surface. The emitter and the receiver could both be black bodies. If one of them is at a higher temperature than the other, there is a net flow of energy from the warmer body to the colder body until equilibrium is reached. This would not be possible without interaction over the same number of modes. Thus the formula which is completely unrelated to thermodynamics is necessary to fulfill one of the main theorems of thermodynamcis, that is that energy flows from warmer to colder bodies.

3.2 Photo-electric Effect

Another strong indication for the quantum nature of light was the photoelectric effect by Lenard in 1903. He discovered that when ultra violet light is radiated on a photo cathode electrons are emitted, see Figure 3.7.

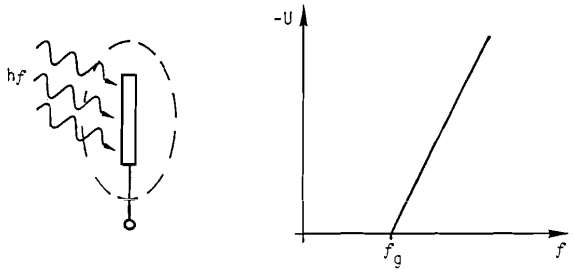

Figure 3.7: Photo-electric effect: (a) Schematic setup and (b) dependence of the necessary grid voltage to supress the electron current as a funtion of light frequency.

Lenard surrounded the photo cathode by a grid, which is charged by the emitted electrons up to a voltage U, which blocks the emission of further electrons. Figure 3.7 shows the blocking voltage as a function of the frequency of the incoming light. Depending on the cathode material there is a cutoff frequency. For lower frequencies no electrons are emitted at all. This frequency as well as the blocking voltage does not depend on the intensity of the light. In 1905, this effect was explained by Einstein introducing the quantum hypothesis for radiation. According to him, each electron emission is caused by a light quantum, now called photon. This photon has an energy hf and this quantum energy must be larger than the work function W_e of the material. The remainder of the energy $m_e v^2/2$ is transfered to the electron in form of kinetic energy. The resulting energy balance is

$$hf = W_e + \frac{1}{2} m_e v^2 \tag{3.40}$$

The kinetic energy of the electron can be used to reach the grid surrounding the photo cathode until the charging energy due to the grid potential is equal to the kinetic energy of the electrons

$$eU = \frac{1}{2} m_e v^2 hf = W_e + \frac{1}{2} m_e v^2 \tag{3.41}$$

or

$$-U = \frac{1}{e}(hf - W_e), \text{ for } hf > W_e. \tag{3.42}$$

This equation explains the empirically found law by Lenard explaining the cutoff frequency and the charge buildup as a function of light frequency. Einstein was first to introduce the idea that the electromagnetic field contains light quanta or photons.

3.3 Spontaneous and Induced Emission

The number of photons in a radiation mode may change via emission of photon into the mode or absorption of a photon from the mode by atoms, molecules or a solid state material. Einstein introduced a phenomenological theory of these processes in order to explain how matter may get into thermal equilibrium by interaction with the modes of the radiation field. He considered the interaction of a mode with atoms modeled by two energy levels E_1 and E_2, see Figure 3.8.

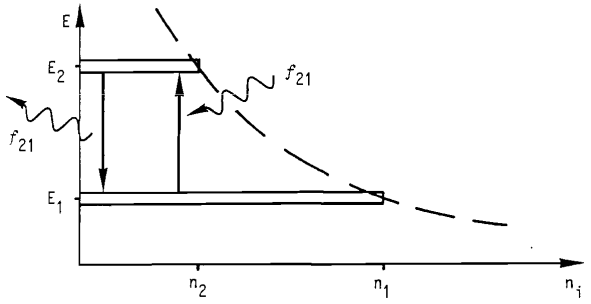

Figure 3.8: Energy levels of a two level atom and populations.

n_1 and n_2 are the population densities of these two levels considering a whole ensemble of these atoms. Transitions are possible in the atom between the two energy levels by emission of a photon at a frequency

$$f = \frac{E_2 - E_1}{h} \tag{3.43}$$

Absorption of a photon is only possible if there is energy present in the radiation field. Einstein wrote for the corresponding transition rates, which should be proportional to the population densities and the photon density at the transition frequency

$$-\frac{dn_1}{dt}\bigg|_{Abs} = \frac{dn_2}{dt}\bigg|_{Abs} = B_{12} n_1 w(f_{21}). \tag{3.44}$$

The coefficient B_{12} characterizes the absorption properties of the transition. Einstein had to allow for two different kind of processes for reasons that become clear a little later. Transitions induced by the already present photons or radiation energy as well as spontaneous transitions

$$\frac{dn_1}{dt}\bigg|_{Em} = -\frac{dn_2}{dt}\bigg|_{Em} = B_{21}\, n_2\, w\,(f_{21}) + A_{21}\, n_2\,. \tag{3.45}$$

The coefficient B_{21}describes the induced and A_{21} the spontaneous emissions. The latter transitions occur even in the absence of any radiation and the corresponding coefficient determines the lifetime of the excited state

$$\tau_{sp} = A_{21}^{-1}, \tag{3.46}$$

in the absence of the radiation field. The total change in the population densities is due to both absorption and emission processes

$$\frac{dn_i}{dt} = \left.\frac{dn_i}{dt}\right|_{Em} + \left.\frac{dn_i}{dt}\right|_{Abs}, \text{ for } i = 1, 2 \tag{3.47}$$

Using Eqs.(3.44) and (3.45) we find

$$-\frac{dn_1}{dt} = \frac{dn_2}{dt} = (B_{12}\, n_1 - B_{21}\, n_2)\ w\,(f_{21}) - A_{21}\, n_2. \tag{3.48}$$

In thermal equilibrium the energy density of the radiation field must fulfill the condition

$$w\,(f_{21}) = \frac{A_{21}/B_{12}}{n_1/n_2 - B_{21}/B_{12}}, \tag{3.49}$$

while the atomic ensemble itself should also be in thermal equilibrium which again should be described by the Boltzmann statistics, i.e. the ratio between the population densities are determined by the Boltzmann factor

$$n_2/n_1 = \exp\left(-\frac{E_2 - E_1}{kT}\right) \quad . \tag{3.50}$$

And with it the energy density of the radiation field must be

$$w\,(f_{21}) = \frac{A_{21}/B_{12}}{\exp\left(\frac{hf_{21}}{kT}\right) - B_{21}/B_{12}} \quad . \tag{3.51}$$

A comparison with Planck's law, Eq.(3.20), gives

$$B_{21} = B_{12}, \tag{3.52}$$

and

$$A_{21} = \frac{8\pi\ hf_{21}^3}{c^3} B_{12}. \tag{3.53}$$

Clearly, without the spontaneous emission process it is impossible to arrive at Planck's Law in equilibrium. The spectral energy density of the radiation field can be rewritten with the average photon number in the modes at the transition frequency f_{21} as

$$w\,(f_{21}) = \frac{8\pi\ f_{21}^2}{c^3} hf_{21}\ \langle s\rangle, \quad \text{with } \ \langle s\rangle = \langle s\rangle = \frac{1}{\exp\frac{hf_{21}}{kT} - 1}. \tag{3.54}$$

Or we can write

$$w\left(f_{21}\right)=\frac{A_{21}}{B_{12}}\left\langle s\right\rangle \ .$$

With that relationship Eq.(3.45) can be rewritten as

$$\left.\frac{dn_1}{dt}\right|_{Em}=-\left.\frac{dn_2}{dt}\right|_{Em}=A_{21}\,n_2\,\left(\left\langle s\right\rangle+1\right)\ , \tag{3.55}$$

which indicates that the number of spontaneous emissions is equvalent to induced emissions caused by the presence of a single photon per mode. Having identified the coefficients describing the transition rates in the atom interacting with the field from equilibrium considerations, we can rewrite the rate equations also for the non equilibrium situation, because the coefficients are constants depending only on the transition considered

$$\frac{dn_1}{dt}=-\frac{dn_2}{dt}=\frac{1}{\tau_{sp}}\,\left[\left(n_2-n_1\right)\left\langle s\right\rangle+n_2\right]\ . \tag{3.56}$$

With each transition from the excited state of the atom to the ground state an emission of a photon goes along with it. From this, we obtain a change in the average photon number of the modes

$$\frac{d\left\langle s\right\rangle}{dt}=V\frac{dn_1}{dt}, \tag{3.57}$$

which is

$$\frac{d\left\langle s\right\rangle}{dt}=\frac{V}{\tau_{sp}}\left[\left(n_2-n_1\right)\left\langle s\right\rangle+n_2\right]\ . \tag{3.58}$$

Again the first term describes the stimulated or induced processes and the second term the spontaneous processes. As we will see later, the stimulated emission processes are coherent with the already present radiation field that is inducing the transitions. This is not so for the spontaneous emissions, which add noise to the already present field. For $n_1>n_2$ the stimulated processes lead to a decrease in the photon number and the medium is absorbing. In the case of inversion, $n_2>n_1$, the photon number increases exponentially. According to Eq.(3.50) inversion corresponds to a negative temperature, which is an indication for a non equilibrium situation that can only be maintained by additional means. It is impossible to achieve inversion by simple irradiation of the atoms with intense radiation. As we see from Eq.(3.58) in steady

state the ratio between excited state and ground state population is

$$\frac{n_2}{n_1} = \frac{\langle s \rangle}{\langle s \rangle + 1}, \tag{3.59}$$

which at most approaches equal population for very large photon number. However, such a process can be exploited in a three or four level system, see Figure 3.9, to achieve inversion.

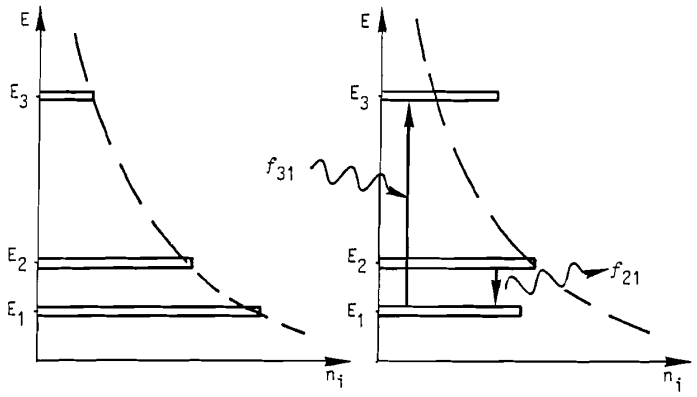

Figure 3.9: Three level system: (a) in thermal equilibrium and (b) under optical pumping at the transition frequency f_{31}.

By optical pumping population from the ground state can be transfered to the excited level with energy E_3. If there is a fast relaxation from this level to level E_2, where level two in contrast has a long lifetime, it is conceivable that an inversion between level E_2 and E_1 can build up. If inversion is achieved radiation at the frequency f_{21} is amplified.

3.4 Matter Waves and Bohr's Model of an Atom

By systematic scattering experiments Ernest Rutherford showed in 1911, that the negative charges in an atom are homogenously distributed in contrast to

the positive charge which is concentrated in a small nucleus about 10,000 times smaller than the atom itself. The nucleus also carries almost all of the atomic mass. Rutherford proposed a model of an atom where the electrons circle the nucleas similar to the planets circling the sun where the gravitational force is replaced by the Coulomb force between the electrons and the nucleus.

This model had many short comings. How was it possible that the electrons, which undergo acceleration on their trajectory around the nucleus, do not radiate according to classical electromagnetism, loose energy and finally fall into the nucleus? Due to advances in optical instrumentation the light emitted from thermally excited atomic vapors was known to be in the form of discrete lines. Balmer found in 1885 that these lines could be expressed by the rule

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right), \text{ with } n = 3, 4, 5, ... \tag{3.60}$$

where λ is the wavelength of light and $R_H = 10.968 \cdot \mu m^{-1}$ is the Rydberg constant for hydrogen. For $n = 3$ this corresponds to the red H_α-line at $\lambda = 656.3$nm, for $n \to \infty$ one obtains the wavelength of the limiting line in this series at $\lambda = 364.6$nm, see Figure 3.10.

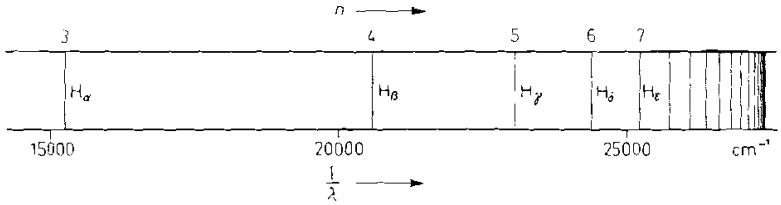

Figure 3.10: Balmer series on a wave number scale.

.

In the subsequent spectroscopy work further sequences where found:
1. Lyman Series:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{1^2} - \frac{1}{n^2} \right), \text{ with } n = 2, 4, 5, ... \tag{3.61}$$

2. Balmer Series:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right), \text{ with } n = 3, 4, 5, ... \tag{3.62}$$

3. Paschen Series:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{3^2} - \frac{1}{n^2} \right), \text{ with } n = 4, 5, 6, ... \tag{3.63}$$

4. Brackett Series:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{4^2} - \frac{1}{n^2} \right), \text{ with } n = 5, 6, 7, ... \tag{3.64}$$

5. Pfund Series:

$$\frac{1}{\lambda} = R_H \left(\frac{1}{5^2} - \frac{1}{n^2} \right), \text{ with } n = 6, 7, 8, ... \tag{3.65}$$

The Lyman series in the UV-region of the spectrum, whereas the Pfund series is in the far infrared. These sequences can be represented as transitions between energy levels as shown in Figure 3.11.

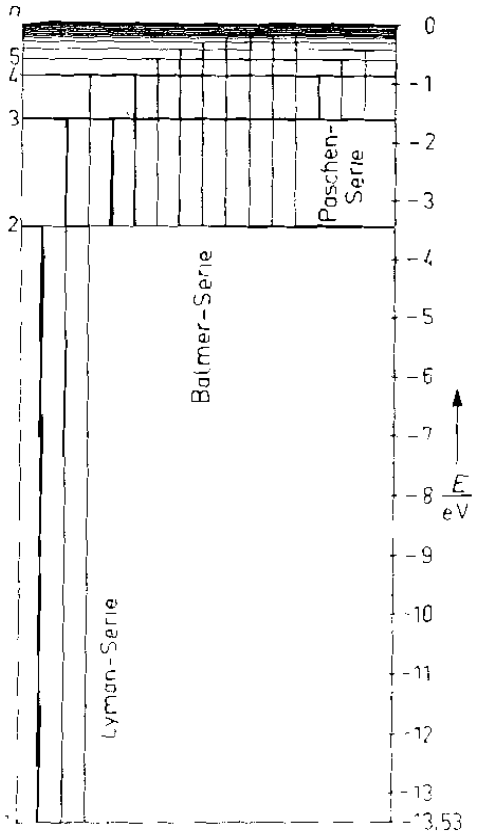

Figure 3.11: Energy level diagram for the hydrogen atom.

.

In 1913, Niels Bohr found the quantization condition for the electron trajectories in the Hydrogen atom and he was able to derive from that the spectral series discussed above. He postulated that only those electron trajectories are allowed that within one rountrip around the nucleus have an action equal to a multiple of Planck's quantum of action h.

$$\oint p \cdot ds = n\,h, \quad \text{with} \qquad n = 1, 2, 3.... \tag{3.66}$$

Second, he postulated that the electron can switch from one energy level or trajectory to another one by the emission or absorption of a photon with an

energy equivalent to the energy difference between the two energy levels, see Figure 3.12.

$$hf = \Delta E. \tag{3.67}$$

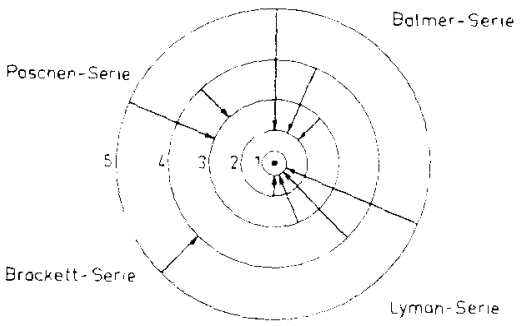

Figure 3.12: Transition between different energy levels in the hydrogen atom.

Assuming a circular trajectory of the electron with radius r_n around the nucleus, the quantization condition for the electron trajectory (3.66) leads to

$$2\pi r_n m v_n = nh, \qquad \text{with} \qquad n = 1, 2, 3... \tag{3.68}$$

The other condition for radius and velocity of the electron around the nucleus is given by the equality of Coulomb and centrifugal force at radius r_n, which leads to

$$\frac{e^2}{4\pi\varepsilon_0 r_n^2} = \frac{m v_n^2}{r_n}, \tag{3.69}$$

or

$$v_n^2 = \frac{e^2}{4\pi\varepsilon_0 r_n m}. \tag{3.70}$$

Substituting this value for the electron velocity in the squared quantization condition (3.68), we find the radius of the electron trajectories

$$r_n = \frac{\varepsilon_0 h^2}{\pi e^2 m} n^2. \tag{3.71}$$

The radius of the first trajectory, called Bohr radius is $r_1 = 0.529 \cdot 10^{-10} m$. The velocities on the individual trajectories are

$$v_n = \frac{e^2}{2\varepsilon_0 h} \frac{1}{n}. \tag{3.72}$$

The highest velocity is found for the tightest trajectory around the nucleus, i.e. for $n = 1$, which can be expressed in terms of the velocity of light as

$$v_1 = \frac{e^2}{2\varepsilon_0 hc} \cdot c = \frac{1}{137} \cdot c, \tag{3.73}$$

where $\frac{e^2}{2\varepsilon_0 hc} = \frac{1}{137}$ is the fine structure constant.

The energy of the electrons on these trajectories with the quantum number n is due to both potential and kinetic energy

$$E_{kin} = \frac{1}{2}mv_n^2 = \frac{me^4}{8\varepsilon_0^2 h^2 n^2}, \tag{3.74}$$

$$E_{pot} = -\frac{e^2}{4\pi\varepsilon_0 r_n} = -\frac{me^4}{4\varepsilon_0^2 h^2 n^2}. \tag{3.75}$$

or

$$E_n = E_{kin} + E_{pot} \tag{3.76}$$

$$E_{pot} = -\frac{me^4}{8\varepsilon_0^2 h^2 n^2}. \tag{3.77}$$

Note, the energy of a bound electron is negative. For $n \to \infty$, $E_n = 0$. The electron becomes detached from the atom, i.e. the atom becomes ionized. The lowest and most stable energy state of the electron is for $n = 1$

$$E_n = -\frac{me^4}{8\varepsilon_0^2 h^2} = -13.53\text{eV}, \tag{3.78}$$

with correspondes to the ground state in hydrogen. When a transition between two of this energy eigenstates occurs a photon with the corresponding energy is released

$$hf = E_k - E_n, \tag{3.79}$$

$$= -\frac{me^4}{8\varepsilon_0^2 h^2}\left(\frac{1}{k^2} - \frac{1}{n^2}\right). \tag{3.80}$$

3.5 Wave Particle Duality

Bohr's postulates were not able to explain all the intricacies of the observed spectra and they couldn't explain satisfactory the structure of the more complex atoms. This was only achieved with the introduction of wave mechanics. In 1923, de Broglie was the first to argue that matter might also have wave properties. Starting from the equivalence principle of mass and energy by Einstein

$$E_0 = m_0\ c_0^2 \tag{3.81}$$

he associated a frequency with this energy accordingly

$$f_0 = m_0\ c_0^2/h. \tag{3.82}$$

Since energy and frequency are not relativistically invariant quantities but rather components of a four-vector which has the particle momentum as its other components $(E_0/c_0, p_x, p_y, p_z)$ or $(\omega_0/c_0, k_x, k_y, k_z)$, it was a necessity that with the energy frequency relationship

$$E = hf = \hbar\omega, \tag{3.83}$$

there must also be a wave number associated with the momentum of a particle according to

$$p = \hbar k. \tag{3.84}$$

In 1927, C. J. Davisson and L. H. Germer experimentally confirmed this prediction by finding strong diffraction peaks when an electron beam penetrated a thin metal film. The pictures were close to the observations of Laue in 1912 and Bragg in 1913, who studied the structure of crystaline and poly crystaline materials with x-ray diffraction.

With that finding the duality between waves and particles for both light and matter was established. Duality means that both light and matter have simultaneous wave and particle properties and it depends on the experimental arrangement whether one or the other property manifests itself strongly in the experimental outcome.

Bibliography

[1] Fundamentals of Photonics, B.E.A. Saleh and M.C. Teich,

[2] Optical Electronics, A. Yariv, Holt, Rinehart & Winston, New York, 1991.

[3] Introduction to Quantum Mechanics, Griffiths, David J., Prentice Hall, 1995.

[4] Quantum Mechanics I, C. Cohen-Tannoudji, B. Diu, F. Laloe, John Wiley and Sons, Inc., 1978.

Chapter 4

Schroedinger Equation

Einstein's relation between particle energy and frequency Eq.(3.83) and de Broglie's relation between particle momentum and wave number of a corresponding matter wave Eq.(3.84) suggest a wave equation for matter waves. This search for an equation describing matter waves was carried out by Erwin Schroedinger. He was successful in the year 1926.

The energy of a classical, nonrelativistic particle with momentum $\vec{p}$ that is subject to a conservative force derived from a potential $V(\vec{r})$ is

$$E = \frac{\vec{p}^{\,2}}{2m} + V(\vec{r})\,. \tag{4.1}$$

For simplicity lets begin first with a constant potential $V(\vec{r}) = V_0 = \text{const}$. This is the force free case. According to Einstein and de Broglie, the dispersion relation between ω and $\vec{k}$ for waves describing the particle motion should be

$$\hbar\omega = \frac{\hbar^2\vec{k}^2}{2m} + V_0. \tag{4.2}$$

Note, so far we had a dispersion relation for waves in one dimension, where the wavenumber $k(\omega)$, was a function of frequency. For waves in three dimensions the frequency of the wave is rather a function of the three components of the wave vector. Each wave with a given wave vector $\vec{k}$ has the following time dependence

$$e^{\mathrm{j}\left(\vec{k}\cdot\vec{r}-\omega t\right)},\ \text{with}\ \omega = \frac{\hbar\vec{k}^2}{2m} + \frac{V_0}{\hbar} \tag{4.3}$$

Note, this is a wave with phase fronts traveling to the right. In contrast to our notation used in chapter 2 for rightward traveling electromagnetic waves, we

switched the sign in the exponent. This notation conforms with the physics oriented literature. A superposition of such waves in $\vec{k}-$space enables us to construct wave packets in real space

$$\Psi\left(\vec{r},t\right)=\int\phi_{\omega}\left(\vec{k},\omega\right)\ e^{\mathrm{j}\left(\vec{k}\cdot\vec{r}-\omega t\right)}d^{3}k\ \ d\omega \tag{4.4}$$

The inverse transform of the above expression is

$$\phi_{\omega}\left(\vec{k},\omega\right)=\frac{1}{\left(2\pi\right)^{4}}\int\Psi\left(\vec{r},t\right)e^{-\mathrm{j}\left(\vec{k}\cdot\vec{r}-\omega t\right)}d^{3}r\ \ dt, \tag{4.5}$$

with

$$\phi_{\omega}\left(\vec{k},\omega\right)=\phi\left(k\right)\ \ \delta\left(\omega-\frac{\hbar\vec{k}^{2}}{2m}-\frac{V_{0}}{\hbar}\right). \tag{4.6}$$

Or we can rewrite the wave function in Eq.(4.4) by carrying out the trivial frequency integration over ω

$$\Psi\left(\vec{r},t\right)=\int\phi\left(k\right)\ \ \exp\left(\mathrm{j}\left[\vec{k}\cdot\vec{r}-\left(\frac{\hbar\vec{k}^{2}}{2m}+\frac{V_{0}}{\hbar}\right)t\right]\right)\ \ d^{3}k. \tag{4.7}$$

Due to the Fourier relationship between the wave function in space and time coordinates and the wave function in wave vector and frequency coordinates

$$\phi_{\omega}\left(\vec{k},\omega\right)\ \ \leftrightarrow\ \ \Psi\left(\vec{r},t\right) \tag{4.8}$$

we have

$$\omega\ \phi_{\omega}\left(k,\omega\right)\ \ \leftrightarrow\ \ \mathrm{j}\frac{\partial\Psi\left(\vec{r},t\right)}{\partial t}\ \ , \tag{4.9}$$

$$\vec{k}\ \phi_{\omega}\left(k,\omega\right)\ \ \leftrightarrow\ \ -\mathrm{j}\nabla\ \Psi\left(\vec{r},t\right)\ \ , \tag{4.10}$$

$$\vec{k}^{2}\ \phi_{\omega}\left(k,\omega\right)\ \ \leftrightarrow\ \ -\Delta\ \Psi\left(\vec{r},t\right)\ \ . \tag{4.11}$$

where

$$\nabla=\vec{e}_{\mathrm{x}}\frac{\partial}{\partial x}+\vec{e}_{\mathrm{y}}\frac{\partial}{\partial y}+\vec{e}_{\mathrm{z}}\frac{\partial}{\partial z}\ \ , \tag{4.12}$$

$$\Delta=\nabla\cdot\nabla\equiv\nabla^{2}=\frac{\partial^{2}}{\partial x^{2}}+\frac{\partial^{2}}{\partial y^{2}}+\frac{\partial^{2}}{\partial z^{2}}. \tag{4.13}$$

From the dispersion relation follows by multiplication with the wave function in the wave vector and frequency domain

$$\hbar\,\omega\,\phi_\omega\left(k,\omega\right)=\frac{\hbar^2k^2}{2m}\phi_\omega\left(k,\omega\right)+V_0\,\phi_\omega\left(k,\omega\right)\quad . \tag{4.14}$$

With the inverse transformation the corresponding equation in the space and tieme domain is

$$\mathrm{j}\,\hbar\frac{\partial\Psi\left(\vec{r},t\right)}{\partial t}=-\frac{\hbar^2}{2m}\Delta\ \Psi\left(\vec{r},t\right)+V_0\ \Psi\left(\vec{r},t\right)\ . \tag{4.15}$$

Generalization of the above equation for a constant potential to the instance of an arbitrary potential in space leads finally to the Schroedinger equation

$$\mathrm{j}\,\hbar\frac{\partial\Psi\left(\vec{r},t\right)}{\partial t}=-\frac{\hbar^2}{2m}\Delta\ \Psi\left(\vec{r},t\right)+V\left(\vec{r}\right)\ \Psi\left(\vec{r},t\right)\ . \tag{4.16}$$

Note, the last few pages ar not a derivation of the Schroedinger Equation but rather a motivation for it based on the findings of Einstein and deBroglie. The Schroedinger Equation can not be derived from classical mechanics. But classical mechanics can be rederived from the Schroedinger Equation in some limit. It is the success of this equation in describing the experimentally observed quantum mechanical phenomena correctly, that justifies this equation.

The wave function $\Psi\left(\vec{r},t\right)$ is complex. Note, we will no longer underline complex quantities. Which quantities are complex will be determined from the context.

Initially the magnitude square of the wave function $\left|\Psi\left(\vec{r},t\right)\right|^2$ was interpreted as the particle density. However, Eq.(4.15) in one spatial dimension is mathematical equivalent to the dispersive wave motion Eq.(2.72), where the space and time variables have been exchanged. The dispersion leads to spreading of the wave function. This would mean that any initially compact particle, which has a well localized particle density, would decay, which does not agree with observations. In the framwork of the "Kopenhagen Interpretation" of Quantum Mechanics, whose meaning we will define later in detail, $\left|\Psi\left(\vec{r},t\right)\right|^2\,dV$ is the probability to find a particle in the volume dV at position $\vec{r}$, if an optimum measurement of the particle position is carried out at time t. The particle is assumed to be point like. $\Psi\left(\vec{r},t\right)$ itself is then considered to be the probability amplitude to find the particle at position $\vec{r}$ at time t.

Note, that the measurement of physical observables like the position of a particle plays a central role in quantum theory. In contrast to classical

mechanics where the state of a particle is precisely described by its position and momentum in quantum theory the full information about a particle is represented by its wave function $\Psi(\vec{r},t)$. $\Psi(\vec{r},t)$ enables to compute the outcome of a measurement of any possible observable related to the particle, like its position or momentum.

Before, we discuss this issue in more detail lets look at a few examples to get familiar with the mathematics of quantum mechanics.

4.1 Free Motion

Eq.(4.15) describes the motion of a free particle. For simplicity, we consider only a one-dimensional motion along the x-axis. Initially, we might only know the position of the particle with finite precision and therefore we use a Gaussian wave packet with finite width as the initial wave function

$$\Psi(x,t=0) = A\ \exp\left(-\frac{x^2}{4\sigma_0^2} + \mathrm{j}k_0 x\right). \tag{4.17}$$

The probability density to find the particle at position x is a Gaussian distribution

$$|\Psi(x,t=0)|^2 = |A|^2 \exp\left(-\frac{x^2}{2\sigma_0^2}\right), \tag{4.18}$$

σ_0^2 is the variance of the initial particle position. Since the probability to find the particle somewhere must be one, we can determine the amplitude of the wave function by requireing

$$\int_{-\infty}^{\infty} |\Psi(x,t=0)|^2\, dx = 1 \rightarrow A\ = \frac{1}{\sqrt[4]{2\pi}\sqrt{\sigma_0}} \tag{4.19}$$

The meaning of the wave number k_0 in the wave function (4.17) becomes obvious by expressing the solution to the wave equation by its Fourier transform

$$\Psi(x,t) = \int_{-\infty}^{+\infty} \phi(k) \exp\quad \mathrm{j}\left(kx - \omega(k)\, t\right)\ dk \tag{4.20}$$

or specifically for $t = 0$

$$\Psi\left(x,0\right) = \int_{-\infty}^{+\infty} \phi\left(k\right) e^{\mathrm{j}kx}\, dk\,, \tag{4.21}$$

or

$$\phi\left(k\right) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Psi(x,0)\, e^{-\mathrm{j}kx}\, dx \quad . \tag{4.22}$$

For the initial Gaussian wavepacket of

$$\Psi\left(x,0\right) = A\ \exp\left(-\frac{x^2}{4\sigma_0^2} + \mathrm{j}k_0 x\right) \tag{4.23}$$

we obtain

$$\phi\left(k\right) = \frac{A\sigma_0}{\sqrt{\pi}} \exp\left[-\sigma_0^2\left(k-k_0\right)^2\right] \quad . \tag{4.24}$$

This is a Gaussian distribution for the wave number, and therefore momentum, of the particle with its center at k_0. With the dispersion relation

$$\omega = \frac{\hbar\ k^2}{2\ m}\,, \tag{4.25}$$

with the constant potential V_0 set to zero, the wave function at any later time is

$$\Psi\left(x,t\right) = \frac{A\sigma_0}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp\left[-\sigma_0^2\left(k-k_0\right)^2 - \mathrm{j}\frac{\hbar k^2}{2m}t + \mathrm{j}kx\right] dk. \tag{4.26}$$

This is exactly the same Gaussian integral we were studying for dispersive pulse propagation or the diffraction of a Gaussian beam in chapter 2 which results in

$$\Psi\left(x,t\right) = \frac{A}{\sqrt{1+\mathrm{j}\frac{\hbar\, t}{m2\sigma_0^2}}}\ \exp\ \left[-\ \frac{x^2 - 4\,\mathrm{j}\sigma_0^2 k_0\ x + \mathrm{j}\,\frac{\hbar 2\sigma_0^2 k_0^2}{m}\ t}{4\sigma_0^2\left(1+\ \mathrm{j}\frac{\hbar}{m2\sigma_0^2}\ t\right)}\right] \quad . \tag{4.27}$$

As expected the wave packet stays Gaussian. The probability density is

$$\left|\Psi(x,t)\right|^2 = \frac{\left|A\right|^2}{\sqrt{1+(\frac{\hbar\, t}{2m\sigma_0^2}))}}\ \exp\ \left[-\ \frac{\left(x-\frac{\hbar\, k_0}{m}\ t\right)^2}{2\sigma_0^2\left[1+\left(\frac{\hbar\, t}{2m\sigma_0^2}\right)^2\right]}\right] \quad . \tag{4.28}$$

With the value for the amplitude A according to Eq.(4.19), the wave packet remains normalized

$$\int_{-\infty}^{+\infty} |\Psi(x,t)|^2 \, dx = 1. \tag{4.29}$$

Using the probability distribution for the particle position, we obtain for its expected value

$$\langle x \rangle = \int_{-\infty}^{+\infty} x \ |\Psi(x,t)|^2 \, dx \tag{4.30}$$

or

$$\langle x \rangle = \frac{\hbar \ k_0}{m} t \quad . \tag{4.31}$$

Thus the center of the wave packet moves with the velocity of the classical particle

$$v_0 = \frac{\hbar \ k_0}{m}, \tag{4.32}$$

which is the group velocity derived from the dispersion relation (4.2)

$$v_0 = \left. \frac{\partial \omega(k)}{\partial k} \right|_{k=k_0} = \frac{1}{\hbar} \left. \frac{\partial E(k)}{\partial k} \right|_{k=k_0} . \tag{4.33}$$

As we will see later, the expected value for the center of mass of the particle follows Newton's law, which is called Ehrenfest's Theorem. For the uncertainty in the particle position

$$\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} \tag{4.34}$$

follows for the freely moving particle

$$\Delta x = \sigma_0 \sqrt{1 + \left(\frac{\hbar \, t}{2m\sigma_0^2} \right)^2} \quad . \tag{4.35}$$

The probability density for the particle position disperses over time. Asymptotically one finds

$$\Delta x \doteq \frac{\hbar \, t}{2m\sigma_0^2} \quad \text{for} \quad \frac{\hbar \, t}{2m\sigma_0^2} \gg 1 \quad . \tag{4.36}$$

Figure 4.1 (a) is a sketch of the complex wave packet and (b) indicates the temporal evolution of the average and variance of the particle center of mass

motion described by the complex wave packet. The wave packet disperses faster, if it is initially stronger localised.

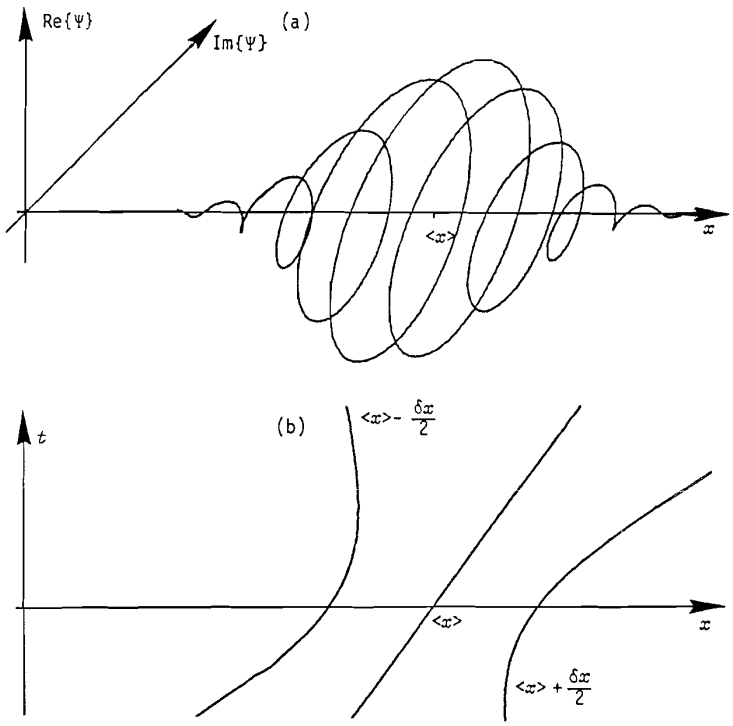

Figure 4.1: Gaussian wave packet: (a) Real and Imaginary part of the complex wave packet. (b) width and center of mass of the wave packet.

Example:

Using this one dimensional model, we can estimate how rapidly an electron moves in a hydrogen atom. If we localize an electron in a box with a size similar to that of a hydrogen atom, i.e. $\sigma_0 = a_0 = 0.5 \cdot 10^{-10} m$, without the presence of the proton that holds the electron back from escaping, it will only take $t = 2m\sigma_0^2/\hbar = 2 \cdot 9.81 \cdot 10^{-31}\text{kg} \cdot (0.5 \cdot 10^{-10})^2\text{m}^2/6.626 \cdot 10^{-34}\text{Js} = 46.5\text{as}$ (attoseconds=10^{-18} sec) until its wave function disperses significantly. This result indicates that electronic motion in atoms occurs on a attosecond time scale. Note, these time scales quickly become very long if macroscopic objects are described quantum mechanically. For example, for a particle with

a mass of 1μg localized in a box with dimensions $1\mu m$, the equivalent time for significant dispersion of the wave function is $t = 2 \cdot 10^{19}$s $= 2$Million years. This result gives us a first indication why we are far, far away from encountering quantum mechanical effects in our everyday life and why the mechanics of the micro cosmos, on an atomic or molecular level, is so different from our macroscopic experience. The reason is the smallness of the quantum of action h.

The reason for this behaviour is that a well localized particle has a wider momentum or wave number distribution. This is in one to one analogy that an otpical pulse disperses faster in a medium with a given dispersion if it is shorter because of larger spectral width. The wave number spread is

$$(\Delta k)^2 = \frac{\int_{-\infty}^{\infty} (k-k_0)^2 \left|\phi(k)\right|^2 dk}{\int_{-\infty}^{\infty} \left|\phi(k)\right|^2 dk} \quad . \tag{4.37}$$

Here, we have

$$\Delta k = \frac{1}{2\sigma_0}, \tag{4.38}$$

or for the momentum spread

$$\Delta p = \frac{\hbar}{2\sigma_0}. \tag{4.39}$$

The position-momentum uncertainty product is then

$$\Delta p \; \Delta x = \frac{\hbar}{2}\sqrt{1 + \left(\frac{\hbar\, t}{2m_0^2}\right)^2} \quad . \tag{4.40}$$

The position-momentum uncertainty product is a minimum at $t = 0$ and steadily increases from this initial value. As we will show later it is in generally true that the position-momentum uncertainty product satisfies the condition

$$\Delta x_{\rm i} \, \Delta p_{\rm i} \geqslant \frac{\hbar}{2}. \tag{4.41}$$

Note, that the index i indicates the coordinate. This is Heisenberg's uncertainy relation between particle position and moment, which holds for each

component individually. Later, we will find other pairs of physical observavles, which are called conjugate observables and which satiesfy similar uncertainty relations. The product of such quantities is always an action. This is for example also true for the product of energy and frequency and the resulting energy-time uncertainty relation is

$$\Delta E \, \Delta t \geqslant \frac{\hbar}{2}. \tag{4.42}$$

Note, whereas the position-momentum uncertainy is related to the choice of the particle state described by the wave function, the energy-time uncertainty relation is related to the dynamics of a quantum process. What it means is that a quantum system can only change its state significantly within a time span Δt, if the state, the quantum system is in, has an energy uncertainty larger than $\delta E \geqslant \frac{\hbar}{2\delta t}$.

Position and momentum variables that do not belong to the same degree of freedom, such as y, and p_x are not subject to an uncertainty relation.

4.2 Probability Conservation and Propability Currents

Max Born was the first to introduce the propabilistic interpretation of the wave function found by Schroedinger, that is the propability to find the center of mass of a particle at position $\vec{r}$ in a volume element dV is given by the magnitude square of the wave function multiplied by dV

$$p\left(\vec{r},t\right) = \left|\Psi\left(\vec{r},t\right)\right|^2 \mathrm{d}\, V \quad . \tag{4.43}$$

If this interpretation makes sense, then the total propability that the particle can be found somewhere should by 1 and this normalization should not change during the dynamics. We found that this is true for the Gaussian wave packet undergoing free motion. Here, we want to show that this is true under the most general circumstances. We look at the rate of change of the

probability to find the particle in an arbitrary but fixed volume $V = Vol$

$$\begin{aligned} &\frac{d}{dt}\int_{Vol} p\left(\vec{r},t\right)\ d^3r = \\ &= \int_{Vol} \frac{\partial}{dt}\left|\Psi\left(\vec{r},t\right)\right|^2\ d^3r \\ &= \int_{Vol}\left[\left(\frac{\partial}{\partial t}\Psi^*\left(\vec{r},t\right)\right)\Psi\left(\vec{r},t\right) + \Psi^*\left(\vec{r},t\right)\left(\frac{\partial}{\partial t}\Psi\left(\vec{r},t\right)\right)\ \right] d^3r \end{aligned} \tag{4.44}$$

Using the Schroedinger Equation (4.16) for the temporal change of the wave function we obtain

$$\begin{aligned} &\frac{d}{dt}\int_{Vol} p\left(\vec{r},t\right)\ d^3r = \\ = &\int_{Vol}\left[\left(\frac{\hbar}{j2m}\nabla\cdot\nabla\ \Psi^*\left(\vec{r},t\right) - \frac{j}{\hbar}V\left(\vec{r}\right)\ {}^*\Psi^*\left(\vec{r},t\right)\right)\Psi\left(\vec{r},t\right)\right] d^3r \\ &+\int_{Vol}\left[\Psi^*\left(\vec{r},t\right)\left(-\frac{\hbar}{j2m}\nabla\cdot\nabla\ \ \Psi\left(\vec{r},t\right) + \frac{j}{\hbar}V\left(\vec{r}\right)\ \ \Psi\left(\vec{r},t\right)\right)\right] d^3r \end{aligned} \tag{4.45}$$

Since the potential $V\left(\vec{r}\right)$ is real the terms related to it cancel. The other two terms can be written of the divergence of a current density

$$\int_{Vol}\frac{\partial}{\partial t}p\left(\vec{r},t\right)\ d^3r = -\int_{Vol}\nabla\cdot\vec{J}\left(\vec{r},t\right)d^3r, \tag{4.46}$$

with

$$\vec{J}\left(\vec{r},t\right) = \frac{\hbar}{j2m}\left(\Psi^*\left(\vec{r},t\right)\ \left(\nabla\Psi\left(\vec{r},t\right)\right) - \Psi\left(\vec{r},t\right)\left(\nabla\Psi^*\left(\vec{r},t\right)\right)\right). \tag{4.47}$$

Eq.(4.46) is true for any volume, i.e.

$$\int_{Vol}\left[\frac{\partial}{\partial t}p\left(\vec{r},t\right) + \nabla\cdot\vec{J}\left(\vec{r},t\right)\right]\ d^3r = 0, \tag{4.48}$$

which is only possible if the integrand vanishes

$$\frac{\partial}{\partial t}p\left(\vec{r},t\right) = -\ \nabla\cdot\vec{J}\left(\vec{r},t\right). \tag{4.49}$$

Clearly, $\vec{J}\left(\vec{r},t\right)$ has the physical meaning of a probability current. The probability in a volume element changes because of probablity flowing out of

this volume element. Note, this is the same local law that we have for the conservation of charge. In fact, if the particle is a charged particle, like an electron is, multiplication of $\vec{J}$ with $-e_0$ would result in the electrical current associated with the wave function $\Psi(\vec{r},t)$.

Gauss's theorem states

$$\int_{Vol} \nabla \cdot \vec{J}(\vec{r},t)\ d^3r = \int_S \vec{J}(\vec{r},t)\ d\vec{S}, \tag{4.50}$$

where Vol is the volume over which the integration is carried out and S is the surface that encloses the volume with $d\vec{S}$ an outward pointing surface normal vector. With Gauss's theorem the local conservation of probability can be transfered to a global result, since

$$\frac{d}{dt}\int_{Vol} p(\vec{r},t)\ d^3r = -\int_{Vol} \nabla \cdot \vec{J}(\vec{r},t)\, d^3r = -\int_S \vec{J}(\vec{r},t)\ d\vec{S}. \tag{4.51}$$

If we choose as the volume the whole space and if $\Psi(\vec{r},t)$ and $\frac{\partial}{\partial t}\Psi(\vec{r},t)$ vanish rapidly enough for $\vec{r}\to\infty$ such that the probability current vanishes at infinity, the total probability is conserved. These findings proove that the probabilty interpretation of the wave function is a valid interpretation not contradicting basic laws of probability. If the wave function properly normalized at the beginning it will stay normalized.

Example The Gaussian wave packet satisfies the condition that the probability current decays rapidly enough at the surface of a large enough chosen volume so that the normalization is preserved. A monochromatic plane wave does not satisfy this condition. However, the probability current density gives a physical meaning to it. The wave function corresponding to a plan wave

$$\Psi(\vec{r},t) = e^{\mathrm{j}\left(\vec{k}\cdot\vec{r}-\omega t\right)},\ \text{with}\ \omega = \frac{\hbar\vec{k}^2}{2m} + \frac{V_0}{\hbar} \tag{4.52}$$

which is not normalizable, results in a homogenous probability current

$$\begin{aligned}\vec{J}(\vec{r},t) &= \frac{\hbar}{j2m}\left(\Psi^*(\vec{r},t)\ \left(\nabla\Psi(\vec{r},t)\right) - \Psi(\vec{r},t)\left(\nabla\Psi^*(\vec{r},t)\right)\right) \\ &= \frac{\hbar\vec{k}}{m}\left|\Psi(\vec{r},t)\right|^2 = \frac{\vec{p}}{m} = \vec{v},\end{aligned} \tag{4.53}$$

that is identical to the classical velocity of the particle. Thus a plane wave describes a particle with a precise velocity or momentum but completely unknown position, therefore the related probability current density is completely homogenous but directed into the direction of $\vec{v}$. Such waves describe the initial state in a scattering experiment, where we shoot particles with a precisely defined velocity $\vec{v}$ or momentum $\vec{p}$ or energy $E = \frac{m\vec{v}^2}{2} = \frac{\hbar\vec{p}^2}{2m} = \frac{\hbar\vec{k}^2}{2m}$ onto another object described by a scattering potential, see problem set. The position of these particles is completely unspecified, i.e. $|\Psi(\vec{r},t)|^2$ =const.

4.3 Measureability of Physical Quantities (Observables)

The reason for the more intricate description necessary for microscopic processes in comparison with macroscopic processes is simply the fact that these systems are so small that the interaction of the system with an eventual measurement apparatus can no longer be neglected. It turns out this fact is not to overcome by choosing more and more sophisticated measurement apparati but rather is a principle limitation. If this is so, then it eventually doesn't make sense or it becomes even inconsistent to attribute to a system more precisely defined physical quantities than actually can be retrieved by measurements. This is the physical reason behind the introduction of the wave function in stead of the precisely defined position and momentum of the particle that we used to deterministically predict the trajectory of a particle in an external field.

It is impossible to assign to a microscopic particle a precise position and momentum at the same time. To demonstrate this, we consider the following (Heisenberg) microscope to measure the exact position of a particle. We use light with wavelength λ and focus it strongly with a lense of some focal distance d, see Figure 4.2.

From our construction of the Gaussian beam in section 2.4.2, we found that if we generate a focused beam with a waist w_o having a Rayleigh range $z_R = \frac{\pi w_o^2}{\lambda}$, the beam is composed of plane waves which have a Gaussian distribution in its transverse k-vector, which has a variance $k_T^2/2$, see Eq.(2.220). The Rayleigh range of the beam is related to the transverse wave number spread of the beam by $z_R = k_0/k_T^2$, with $k_0 = 2\pi/\lambda$, see (2.221) and thereafter. Note, the intensity profile of the beam has a variance $w_o^2/4$. If a particle

crosses the focus of the beam and scatters a single photon, which we detect with the surrounding photo detector arrangement, then it is reasonable to assume that we know the position of the particle in the x-direction, with an uncertainty equal to the uncertainty in the transverse photon or intensity distribution of the beam, i.e. $\Delta x = w_o\ /2$.

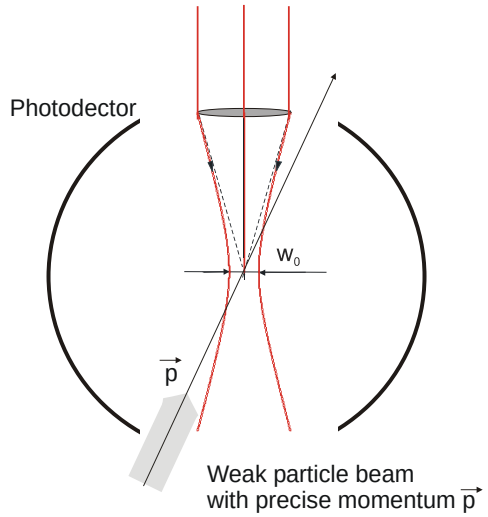

Figure 4.2: Determination of particle position with an optical microscope. A weak particle beam with precisely defined moment $\vec{p}$ of the particles is directed towards the focus of the Gaussian beam. In the focus the particle scatters at least one photon. Detection of the scattered photon with the surrounding photodetector signals, that the position of the particle in x-direction has been determined within the beam waist of the Gaussian beam. However, due to the scattering of the photon a momentum uncertainty has been introduced to the particle state.

During the measurement, the photon recoil induces a momentum kick with an uncertainty $\Delta p_x = \hbar k_T/\sqrt{2}$. So even if the momentum of the particle was perfectly know before the measurement, after the additional determination of its position with a precision Δx it has at least aquired an uncertainty in its momentum of magnitude Δp_x. The product of the uncertainties in postion and momentum after the measurement is

$$\Delta p_x \cdot \Delta x = \hbar k_T w_o / \left(2\sqrt{2}\right) = \frac{\hbar}{2}. \tag{4.54}$$

Note, this result is exact and is independent of focusing. Tighter focusing will enable us to more precisely determine the position of the particle, but we will introduce more momentum uncertainty due to the photon recoil; the opposite is true for less focusing. Since we can not determine, and therefore, prepare a particle in a state with its position and momentum more precisely determined than this uncertainty product allows, there is nowsuch state and (4.54) is the minimum uncertainty product achievable.

The experimental setup can easily be extended to measure the momentum and position of a particle in all three dimensions. For example one can use three focused laser beams at different wavelength, which are orthogonal to each other. Once a particle will fly through the focus and scatters three photons, each of different color. If we knew its momentum initially precisely, we would know afterwards its 3-dimensional position with a position and momentum spread as described by Eq.(4.54).

4.4 Stationary States

One of the great mysteries before the advent of quantum mechanics was the orgin of the discrete energy spectra observed in spectroscopic investigations and empirically described by the Bohr-Sommerfeld model of the atom. This mystery is easily explained by the Schroedinger Equation (4.16)

$$\mathrm{j}\,\hbar\frac{\partial\Psi\left(\vec{r},t\right)}{\partial t}=-\frac{\hbar^{2}}{2m}\Delta\ \Psi\left(\vec{r},t\right)+V\left(\vec{r}\right)\ \ \Psi\left(\vec{r},t\right)\ . \tag{4.55}$$

It allows for solutions

$$\Psi\left(\vec{r},t\right)=\psi\left(\vec{r}\right)\ e^{\mathrm{j}\,\omega t}, \tag{4.56}$$

which have a time independent probability density, i.e.

$$\left|\Psi\left(\vec{r},t\right)\right|^{2}=\left|\psi\left(\vec{r}\right)\right|^{2}=\text{const.}, \tag{4.57}$$

which is the reason for calling these states stationary states. Since the right side of the Schroedinger Equation is equal to the total energy of the system, these states correspond to energy eigenstates of the system with energy eigenvalues

$$E=\hbar\omega. \tag{4.58}$$

These energy eigenstates $\psi(\vec{r})$ are eigen solutions to the stationary or time independent Schroedinger Equation

$$-\frac{\hbar^2}{2m}\Delta\ \psi(\vec{r}) + V(\vec{r})\ \psi(\vec{r}) = E\ \psi(\vec{r})\ . \tag{4.59}$$

We get familiar with this equation by considering a few one-dimensional examples, before we apply it to the Hydrogen atom.

4.4.1 The One-dimensional Infinite Box Potential

A simple example for a quantum mechanical system is an electron that can freely move in one dimension x but only over a finite distance a. Such a situation closely describes an electron that is strongly bound to a molecule with a cigar like shape with length a. The potential describing this situation is the one-dimensional box potential

$$V(x) = \begin{cases} 0, \text{ for } |x| < a/2 \\ \infty, \text{ for } |x| \geq a/2 \end{cases}\ , \tag{4.60}$$

see Figure 4.3.

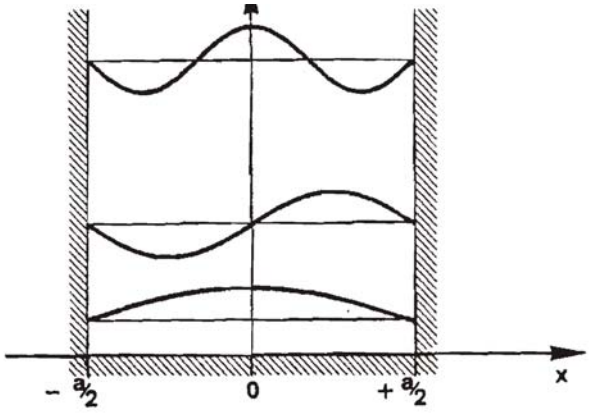

Figure 4.3: One dimensional box potential with infinite barriers.

In the interval $[-a/2, a/2]$ the stationary Schroedinger equation is

$$-\frac{\hbar^2\ d^2\psi(x)}{2m\ dx^2} = E\ \psi(x)\,. \tag{4.61}$$

For $|x| \geq a/2$ the wave function must vanish, otherwise the energy eigenvalue can not be finite, i.e. $\psi(x = \pm a/2) = 0$. This is analogous to the electric field solutions for the TE-modes for a planar mirror waveguide and we find

$$\psi_n(x) = \sqrt{\frac{2}{a}} \cos \frac{n\pi x}{a} \quad \text{for} \quad n = 1, 3, 5 \ldots, \tag{4.62}$$

$$\psi_n(x) = \sqrt{\frac{2}{a}} \sin \frac{n\pi x}{a} \quad \text{for} \quad n = 2, 4, 6 \ldots. \tag{4.63}$$

The corresponding energy eigenvalues are

$$E_n = \frac{n^2 \pi^2 \hbar^2}{2ma^2}. \tag{4.64}$$

We also find that the stationary states constitute an orthogonal system of functions

$$\int_{-\infty}^{+\infty} \psi_m(x)^* \psi_n(x)\, dx = \delta_{mn}. \tag{4.65}$$

In fact this system is complete. Any function in the interval $[-a/2, a/2]$ can be expanded in a superposition of the basis functions $\psi_n(x)$, which is a Fourier series

$$f(x) = \sum_{n=0}^{\infty} c_n \psi_n(x) \tag{4.66}$$

with

$$c_m = \int_{-a/2}^{a/2} \psi_m(x)^* f(x)\, dx, \tag{4.67}$$

which is a consequence of the orthogonality relation (4.65).

Example: If we approximate the binding potential of a hydrogen atom by a one-dimensional box potential with a width equal to twice the Bohr radius $a = 2a_0 = 10^{-10}$m, the energy eigenvalues are $E_n = n^2 \cdot 35eV$. Clearly, the spacing of the energy eigenvalues does not conform with what has been observed experimentaly, compare with section 3.4, however the energy scale is within an order of magnitude. The ionization potential of the hydrogen atom is $13.5eV$.

4.4.2 The One-dimensional Harmonic Oscillator

The most important example of a quantum system is the one-dimensional harmonic oscillator. It is the most basic mechanical and electrical system and it describes the dynamics of a mode of the radiation field, see Figure 4.4.

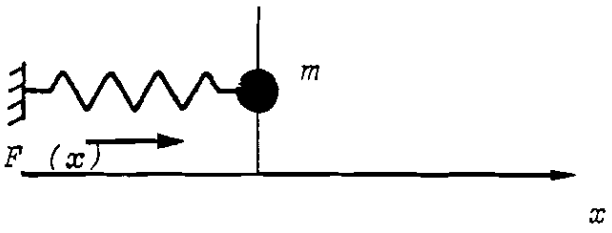

Figure 4.4: Elastically bound particle

Mechanically, a harmonic oscillation comes about by the elastic force obeying Hook's law

$$F(x) = -Kx, \tag{4.68}$$

that pulls back a particle with mass m in its equilibrium position. This force is conservative and can be derived from a potential by

$$F(x) = -\frac{\mathrm{d}\, V(x)}{\mathrm{d}x}, \tag{4.69}$$

with

$$V(x) = \frac{1}{2}\, Kx^2\,. \tag{4.70}$$

Newton's law results in the classical equation of motion

$$m\ddot{x} = F(x)\,, \tag{4.71}$$

or

$$\ddot{x} + \omega_0^2 x = 0, \tag{4.72}$$

with the oscillation frequency

$$\omega_0 = \sqrt{\frac{K}{m}} \tag{4.73}$$

The corresponding stationary Schroedinger Equation is

$$\frac{d^2\psi(x)}{dx^2} + \frac{2m}{\hbar^2}\left(E - \frac{1}{2}Kx^2\right)\ \psi(x) = 0. \tag{4.74}$$

This equation is well known in mathematical physics and we want to bring it into standardized form by the scale transformation, i.e. introducing a normalized distance

$$\xi = ax, \tag{4.75}$$

with the scale factor

$$a = \left(\frac{mK}{\hbar^2}\right)^{\frac{1}{4}} = \sqrt{\frac{\omega_0 m}{\hbar}} = \sqrt{\frac{K}{\hbar\omega_0}}. \tag{4.76}$$

In addition we introduce the energy scale factor

$$\gamma = \frac{2E}{\hbar\omega_0}. \tag{4.77}$$

Then the stationary Schroedinger Equation for the harmonic oscillator is

$$\frac{d^2\psi(\xi)}{d\xi^2} + \left(\gamma - \xi^2\right)\ \psi(\xi) = 0. \tag{4.78}$$

It turns out [4][6], that this equation has only solutions that are bounded, i.e. $\psi(\xi \to \pm\infty) = 0$, if the normalized energies are

$$\gamma_n = 2n + 1. \tag{4.79}$$

And the corresponding eigensolutions are the Hermite Gaussians,

$$\psi_n(\xi) = \text{const. } H_n(\xi)\ e^{-\frac{1}{2}\xi^2}, \tag{4.80}$$

which we discovered already as solutions of the paraxial wave equation, see Eqs.(2.298) and (2.299), i.e.

$$H_n(\xi) = (-1)^n e^{\xi^2}\ \frac{d^n}{d\xi^n}\ e^{-\xi^2} \tag{4.81}$$

$$\begin{array}{ll} H_0(\xi) = 1 \quad , & H_3(\xi) = 8\ \xi^3 - 12\ \xi\ , \\ H_1(\xi) = 2\ \xi\ , & H_4(\xi) = 16\ \xi^4 - 48\ \xi^2 + 12\ , \\ H_2(\xi) = 4\ \xi^2 - 2\ , & H_5(\xi) = 32\ \xi^5 - 160\ \xi^3 + 120\ \xi\ . \end{array} \tag{4.82}$$

After denormalization and normalization the stationary wave functions are

$$\psi_n(x) = \sqrt{\frac{a}{2^n\sqrt{\pi}\ n!}}\ H_n(ax)\ e^{-\frac{1}{2}a^2x^2}. \tag{4.83}$$

Again, we find that the Hermite Gaussians constitute an orthogonal system of functions such that

$$\int_{-\infty}^{+\infty} \psi_m(x)^*\,\psi_n(x)\,dx = \delta_{mn}. \tag{4.84}$$

Figure 4.5 shows the first six stationary states or energy eigenstates of the harmonic oscillator.

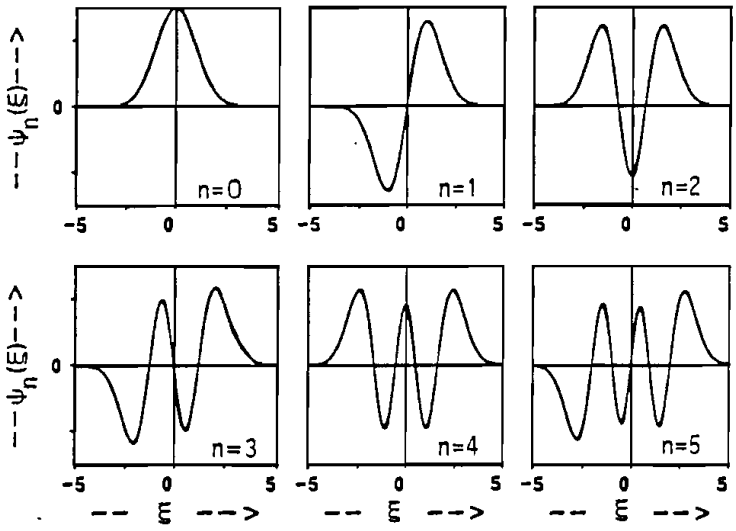

Figure 4.5: First six stationary states of the harmonic oscillators.

The energy eigenvalues of the stationary states are

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega_0\,. \tag{4.85}$$

Note, that the energy eigenvalues are equidistant and the difference between two energy eigenstates follows the findings of Planck. An oscillator has discrete energy levels which differ by energy quanta of size $\hbar\omega_0$, see Figure

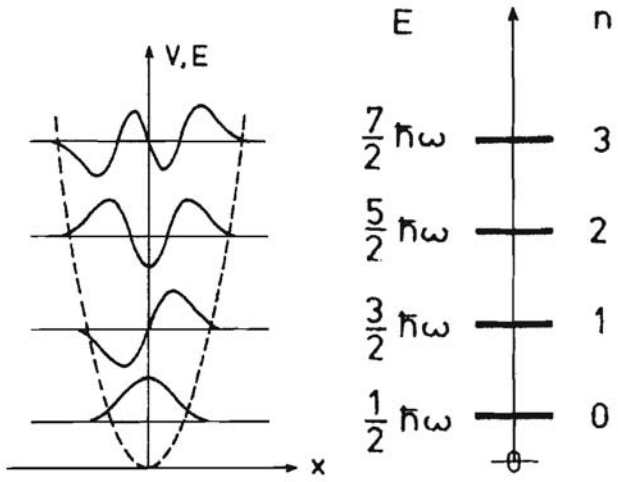

Figure 4.6: Lowest order wavefunctions of the harmonic oscillator and the corresponding energy eigenvalues [3].

.The only difference is, that the whole energy scale is shifted by the energy of half a quantum, which is the lowest energy eigenvalue. Thus the minimum energy, or ground state energy, of a harmonic oscillator is not zero but $E_0 = \frac{1}{2}\hbar\omega_0$.It is obvious, that an oscillator can not have zero energy because its energy is made up of kinetic and potential energy

$$E = \frac{p^2}{2m} + \frac{1}{2} Kx^2. \tag{4.86}$$

Since every state has to fulfill Heisenberg's uncertainty relation $\Delta p \cdot \Delta p \geq \frac{\hbar}{2}$, one can show that the state with minimum energy possible has an energy $E_0 = \frac{1}{2}\hbar\omega_0$, which is true for the ground state $\psi_0(x)$ according to Eq.(4.83). The stationary states of the harmonic oscillator correspond to states with precisely definied energy but completely undefined phase. If we assume a classical harmonic oscillator with a well defined energy $E = \frac{1}{2} Kx_0^2$. Note, that during a harmonic oscillation the energy is periodically converted from potential energy to kinetic energy. Then the oscillator oscillates with a fixed ampltiude x_0

$$x(t) = x_0 \cos(\omega_0 t + \varphi). \tag{4.87}$$

If the phase is assumed to be random in the interval [-π, π], one finds for the

probability density of the position x to be

$$p(x) = \frac{1}{\pi\sqrt{x_0^2 - x^2}}.$$

Figure 4.7 shows this probability density corresponding to an energy eigenstate $\psi_n(x)$ with quantum large quantum number $n = 10$.

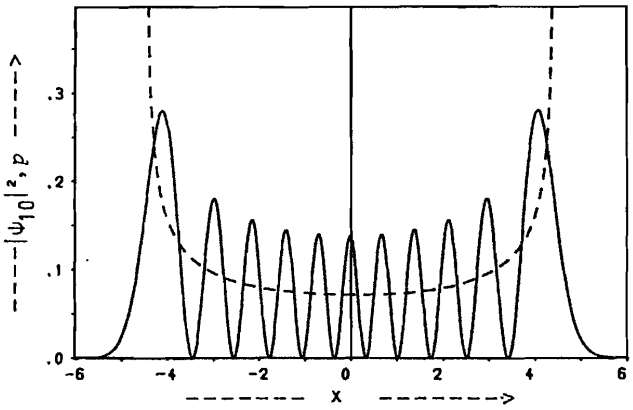

Figure 4.7: Probability density $|\psi_{10}|^2$ of the harmonic oscillator containing exactly 10 energy quanta.

On average, the quantum mechanical probability density agrees with the classical probability density, which is some form of the correspondence principle, which says that for large quantum numbers n the wave functions resume classical properties.

4.5 The Hydrogen Atom

The simplest of all atoms is the Hydrogen atom, which is made up of a positively charged proton with rest mass $m_p = 1.6726231 \times 10^{-27}\,\mathrm{kg}$, and a negatively charged electron with rest mass $m_e = 9.1093897 \times 10^{-31}\,\mathrm{kg}$. Therefore, the hydrogen atom is the only atom which consists of only two particles. This makes an analytical solution of both the classical as well as the quantum mechanical dynamics of the hydrogen atom possible. All other atomes are composed of a nucleus and more than one electron. According

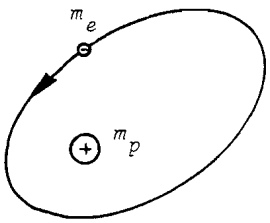

Figure 4.8: Bohr Sommerfeld model of the Hydrogen atom.

to the Bohr-Somerfeld model of hydrogen, the electron circles the proton on a planetary like orbit, see Figure 4.8.The stationary Schroedinger Equation for the Hydrogen atom is

$$\Delta\psi\left(\vec{r}\right)+\frac{2m_0}{\hbar^2}\left(E-V\left(\vec{r}\right)\right)\ \psi\left(\vec{r}\right)=0 \tag{4.88}$$

The potential is a Coulomb potential between the proton and the electron such that

$$V\left(\vec{r}\right)=-\frac{e_0^2}{4\,\pi\,\varepsilon_0\ \left|\vec{r}\right|} \tag{4.89}$$

and the mass is actually the reduced mass

$$m_0=\frac{m_p\ \cdot\ me}{m_p+m_e} \tag{4.90}$$

that arises when we transform the two body problem between electron and proton into a problem for the center of mass and relative coordinate motion. Due to the large, but finite, mass of the proton, i.e. the proton mass is 1836 times the electron mass, both bodies circle around a common center of mass. The center of mass is very close to the position of the proton and the reduced mass is almost identical to the proton mass. Due to the spherical symmetry of the potential the use of spherical coordinates is advantageous

$$\Delta\psi=\frac{\partial^2\psi}{\partial r^2}+\frac{2}{r}\ \frac{\partial\psi}{\partial r}+\frac{1}{r^2}\left[\frac{1}{\sin\vartheta}\ \frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial\psi}{\partial\vartheta}\right)+\frac{1}{\sin^2\vartheta}\ \frac{\partial^2\psi}{\partial\ \varphi^2}\right] \tag{4.91}$$

We will derive separate equations for the radial and angular coordinates by assuming trial solutions which are products of functions only depending on

one of the coordinates $r\,,\vartheta\,,$ or φ

$$\psi\left(r,\vartheta,\varphi\right)=R\left(r\right)\;\;\theta\left(\vartheta\right)\;\;\phi\left(\varphi\right)\;. \tag{4.92}$$

Substituting this trial solution into the stationary Schroedinger Eq.(4.91) and separating variables leads to radial equation

$$\frac{d^2R}{dr^2}+\frac{2}{r}\;\frac{dR}{dr}+\left(\frac{2m_0E}{\hbar^2}+\frac{m_0e_0^2}{2\pi\varepsilon_0\hbar^2r}-\frac{\alpha}{r^2}\right)R=0\quad, \tag{4.93}$$

the azimuthal equation

$$\frac{1}{\sin\vartheta}\;\frac{d}{d\vartheta}\left(\sin\vartheta\frac{d\theta}{d\vartheta}\right)+\left(\alpha-\frac{m^2}{\sin^2\vartheta}\right)\theta=0\quad, \tag{4.94}$$

and the polar equation

$$\frac{d^2\phi}{d\varphi^2}+m^2\phi=0\quad, \tag{4.95}$$

where α and m are constants yet to be determined. The polar equation has the complex solutions

$$\phi\left(\varphi\right)=\text{const. }e^{\mathrm{j}m\varphi},\text{ with }m=\ldots-2,-1,0,1,2\ldots \tag{4.96}$$

because of the symmetry of the problem in the polar angle φ, i.e. the wave-function must be periodic in φ with period 2π.

4.5.1 Spherical Harmonics

The azimuthal equation is transformed by the substitution

$$\xi=\cos\vartheta \tag{4.97}$$

into

$$\left(1-\xi^2\right)\frac{d^2\theta}{d\xi^2}-2\xi\frac{d\theta}{d\xi}+\left(\alpha-\frac{m^2}{1-\xi^2}\right)\theta=0\quad. \tag{4.98}$$

It turns out, that this equation has only bounded solutions on the interval $\xi\epsilon[-1,1]$, if the constant α is a whole number

$$\alpha=l\left(l+1\right)\quad\text{with,}\quad l=0,1,2\ldots \tag{4.99}$$

and

$$m = -l, -l+1, \ldots \quad -1, 0, 1 \ldots l-1, l \tag{4.100}$$

For $m = 0$, Eq.(4.98) is Legendre's Differential Equation and the solutions are the Legendre-Polynomialsm [5]

$$\begin{array}{ll} P_0(\xi) = 1\,, & P_3(\xi) = \frac{5}{2}\xi^3 - \frac{3}{2}\xi\,, \\ P_1(\xi) = \xi\,, & P_4(\xi) = \frac{35}{8}\xi^4 - \frac{15}{4}\xi^2 + \frac{3}{8}\,, \\ P_2(\xi) = \frac{3}{2}\xi^2 - \frac{1}{2}\,, & P_5(\xi) = \frac{63}{8}\xi^5 - \frac{35}{4}\xi^3 + \frac{15}{8}\xi\,. \end{array} \tag{4.101}$$

For $m \neq 0$, Eq.(4.98) is the associated Legendre's Differential Equation and the solutions are the associated Legendre-Polynomials, which can be generated from the Legendre-Polynomials by

$$P_1^m(\xi) = \left(1-\xi^2\right)^{m/2} \frac{d^m P_1(\xi)}{d\xi^m}. \tag{4.102}$$

Overall the angular functions can be combined to form the spherical harmonics

$$Y_1^m(\vartheta,\varphi) = (-1)^m \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}}\; P_1^m(\cos\vartheta)\, e^{\mathrm{j}m\,\varphi}\,, \tag{4.103}$$

which play an important role whenever a partial differential equation that contains the Laplace operator is solved in spherical coordinates. The spherical harmonics form a system of orthogonal functions on the full volume angle 4π, i.e. $\vartheta\epsilon[0,\pi]$ and $\varphi\epsilon[-\pi,\pi]$

$$\int_0^{\pi}\int_0^{2\pi} Y_l^{m*}(\vartheta,\varphi)\, Y_{l'}^{m'}(\vartheta,\varphi)\sin\vartheta\; d\vartheta\; d\varphi = \delta_{ll'}, \delta_{mm'}. \tag{4.104}$$

Therefore, a function of the angular variable (ϑ,φ) can be expanded in spherical harmonics. The spherical harmonics with negative azimuthal number -m can be expressed in terms of those with positive azimuthal number m.

$$Y_1^{-m}(\vartheta,\varphi) = (-1)^m \left(Y_l^m(\vartheta,\varphi)\right)^* \quad . \tag{4.105}$$

The lowest order spherical harmonics are listed in Table 4.1. Figure 4.9 shows a cut through the spherical harmonics $Y_1^m(\vartheta,\varphi)$ along the meridional plane.

$$\mathrm{Y}_0^0(\vartheta,\varphi)=\tfrac{1}{\sqrt{4\pi}},\qquad Y_1^0(\vartheta,\varphi)=\sqrt{\tfrac{3}{4\pi}}\cos\vartheta\,,\qquad Y_1^1(\vartheta,\varphi)=-\sqrt{\tfrac{3}{8\pi}}\sin\vartheta\;\mathrm{e}^{\mathrm{j}\varphi}\;,$$

$$\mathrm{Y}_2^0(\vartheta,\varphi)=\sqrt{\tfrac{5}{16\pi}}\left(3\cos^2\vartheta-1\right),\ \mathrm{Y}_2^1(\vartheta,\varphi)=\text{-}\sqrt{\tfrac{15}{8\pi}}\sin\vartheta\cos\varphi\,\mathrm{e}^{\mathrm{j}\varphi},\ \mathrm{Y}_2^2(\vartheta,\varphi)=\sqrt{\tfrac{15}{32\pi}}\sin^2\vartheta\;\mathrm{e}^{2\mathrm{j}\varphi},$$

$$\mathrm{Y}_3^0(\vartheta,\varphi)=\sqrt{\tfrac{7}{16\pi}}\left(5\;\cos^3\vartheta-3\;\cos\vartheta\right),\qquad \mathrm{Y}_3^1(\vartheta,\varphi)=\text{-}\sqrt{\tfrac{21}{64\pi}}\sin\vartheta\left(5\;\cos^2\vartheta-1\right)\mathrm{e}^{\mathrm{j}\varphi}\,,$$

$$\mathrm{Y}_3^2(\vartheta,\varphi)=\sqrt{\tfrac{105}{32\pi}}\sin^2\vartheta\;\cos\vartheta\;\mathrm{e}^{\mathrm{j}2\varphi}\,,\qquad \mathrm{Y}_3^3(\vartheta,\varphi)=\text{-}\sqrt{\tfrac{35}{64\pi}}\sin^3\vartheta\;\mathrm{e}^{\mathrm{j}3\varphi}\quad .$$

Table 4.1: Lowest order spherical harmonics

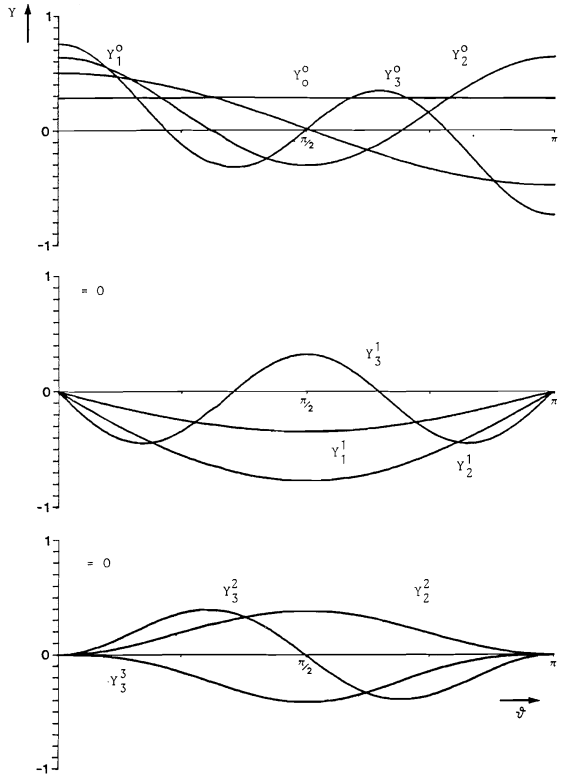

Figure 4.9: Lowest order spherical harmonics $Y_1^m(\vartheta,\varphi)$, along the meridional plane, i.e. $\varphi=0$.

4.5.2 Radial Wave Functions

Obviously, the spherical harmonics are related to the angular momentum L of the particle, because after choosing the spherical harmonic with indices l, m the radial Equation (4.93) is

$$\frac{d^2R}{dr^2} + \frac{2}{r}\,\frac{dR}{d\,r} + \left(\frac{2m_0E}{\hbar^2} + \frac{m_0e_0^2}{2\pi\varepsilon_0\hbar^2 r} - \frac{l\,(l+1)}{r^2}\right)\,R = 0. \tag{4.106}$$

The radial equation has in addition to the $1/r$ Coulomb potential the centrifugal potential

$$E_{rot} = \frac{\hbar^2}{2m_0}\frac{l\,(l+1)}{r^2} = \frac{\vec{L}^2}{2m_0r^2}, \tag{4.107}$$

which is the rotation energy of a particle with angular momentum $\left|\vec{L}\right| = \sqrt{l\,(l+1)}\hbar$ and moment of inertia m_0r^2. Thus quantum mechanically, the particle can no longer access arbitrary values for the angular momentum. The angular momentum can only have values $\left|\vec{L}\right| = \sqrt{l\,(l+1)}\hbar$ with $l = 0, 1, 2, ...$ For large radii, the radial equation simplifies to

$$\frac{d^2R}{dr^2} + \frac{2m_0E}{\hbar^2}\,R = 0, \tag{4.108}$$

which indicates that the radial wave function must decay exponentially for large radii. Therefore, we rescale the radius accoring to

$$\rho = Ar \tag{4.109}$$

with

$$A^2 = -\frac{8m_0E}{\hbar^2},\ \text{because } E < 0, \tag{4.110}$$

and form the trial solution

$$R\,(\rho) = \rho^s\,w\,(\rho)\,e^{-\rho/2}. \tag{4.111}$$

Substitution into Eq.(4.109) leads to the following differential equation for $w\,(\rho)$

$$\rho^2\frac{d^2w}{d\rho^2} + \rho\,[2\,(s+1) - \rho]\,\frac{dw}{d\rho} + [\rho\,(\lambda - s - 1) + s\,(s+1) - l\,(l+1)]\ w = 0, \tag{4.112}$$

with

$$\lambda = \frac{m_0 e^2}{2\pi\varepsilon_0 \hbar^2 A} = \frac{\sqrt{m_0} e^2}{4\sqrt{2}\pi\varepsilon_0 \hbar \sqrt{-E}}. \tag{4.113}$$

Evaluation of this differential equation at $\rho = 0$ leads to

$$l = s,$$

and we are left with the much simpler equation

$$\rho \frac{d^2 w}{d\rho^2} + \left[2\left(l+1\right) - \rho\right] \frac{dw}{d\rho} + \left(\lambda - l - 1\right)\ w = 0\,. \tag{4.114}$$

One way to solve this equation is by using a polynomial trial solution.

$$w\left(\rho\right) = b_0 + b_1\rho + b_2\rho^2 + \ldots b_p\rho^p \tag{4.115}$$

Substitution into Eq.(4.114) leads to the following recursion relation for the coefficients

$$b_{k+1} = \frac{k + l + 1 - \lambda}{\left(k+1\right)\ \left(k + 2l + 2\right)}\ b_k \tag{4.116}$$

For

$$\lambda = p + l + 1 \tag{4.117}$$

the recursion breaks off and we obtain a polynomial of finite order. If λ is not an integer the polynomial does not stop and the corresponding series converges against a $w\left(\rho\right)$ that has an asymptotic behavior $w\left(\rho\right) \tilde{} e^{\rho}$, which leads to a radial function not normalizable. Thus we have the condition

$$\lambda \equiv n, \text{ with } n \geqslant l + 1 \tag{4.118}$$

and in total

$$w\left(\rho\right) = L^{21+1}_{n-l+1}\left(\rho\right) \tag{4.119}$$

with the Laguerre Polynomials

$$L^r_s\left(x\right) = \sum_{q=0}^{s} \left(-1\right)^q \frac{\left(s+r\right)!^2}{\left(s-q\right)!\ \left(r+q\right)!}\ \frac{x^q}{q!}. \tag{4.120}$$

The lowest order Laguerre Polynomials are summarized in Table 4.2 The radial wave function is then a Laguerre function

$$F_{n1}(\rho) = \ \rho^1\, L^{21+1}_{n-l=1}\left(\rho\right)\ e^{-\rho/2}, \tag{4.121}$$

$$L_0^1(x) = 1 \quad , \quad L_1^1(x) = 4 - 2x \quad , \quad L_2^1(x) = 18 - 18x + 3x^2 \quad ,$$

$$L_3^1(x) = 96 - 144x + 48x^2 - 4x^3 \quad , \quad L_0^2(x) = 2 \quad ,$$

$$L_1^2(x) = 18 - 6x \quad , \quad L_2^2(x) = 144 - 96 + 12x^2 \quad ,$$

$$L_3^3(x) = 6 \quad , \quad L_1^3(x) = 96 - 24x \quad ,$$

$$L_0^4(x) = 24 \quad .$$

Table 4.2: Lowest order Laguerre Polynomials

and they again form an orthogonal system of functions

$$\int_0^\infty F_{nl}(\rho) F_{n'l}(\rho) \rho^2 d\rho = \frac{2n\left[(n+l)!\right]^3}{(n-l-1)!} \delta_{nn'} \quad . \tag{4.122}$$

We now reverse the normalization of the radial coordinate and from Eqs.(4.109,4.110) and (4.113) we find

$$\rho = \frac{2r}{na_0} \tag{4.123}$$

with the Bohr radius

$$a_0 = \frac{4\pi\varepsilon_0\hbar^2}{e_0^2 m_0}, \tag{4.124}$$

which we found already in the Bohr-Sommerfeld model, see section 3.4. The radial wave function is then

$$R_{n1}(r) = N_{nl} F_{nl}(\rho) \quad . \tag{4.125}$$

And the normalization factor is determined by

$$\int_0^\infty R_{nl}(r) \; R_{n'l}(r) \; r^2 \, dr = \delta_{n,n'}, \tag{4.126}$$

which gives

$$N_{nl} = \frac{2}{n^2}\sqrt{\frac{(n-l-1)!}{\left[(n+l)!\right]^3}} \; a_0^{-3/2} . \tag{4.127}$$

The radial wave functions of the hydrogen atom are listed in Table 4.3 and plots of the lowest order radial wave functions are presented in Figure 4.10

$$R_{10}(r) = \frac{2}{\sqrt{a_0^3}} e^{-r/a_0}, \qquad R_{20}(r). = \frac{1}{2\sqrt{2}\sqrt{a_0^3}} \left(2 - \frac{r}{a_0}\right) e^{-r/2a_0}$$

$$R_{21}(r) = \frac{1}{2\sqrt{6}\sqrt{a_0^3}} \frac{r}{a_0} e^{-r/2a_0}$$

$$R_{30}(r) = \frac{1}{81\sqrt{3}\sqrt{a_0^3}} \left(27 - 18\frac{r}{a_0} + 2\frac{r^2}{a_0^2}\right) e^{-r/3a_0}$$

$$R_{31}(r) = \frac{4}{81\sqrt{6}\sqrt{a_0^3}} \left(6 - \frac{r}{a_0}\right) \frac{r}{a_0} e^{-r/3a_0}, \quad R_{32}(r) = \frac{4}{81\sqrt{30}\sqrt{a_0^3}} \frac{r^2}{a_0^2} e^{-r/3a_0}$$

Table 4.3: Lowest order radial wavefunctions $R_{n,l}(r)$.

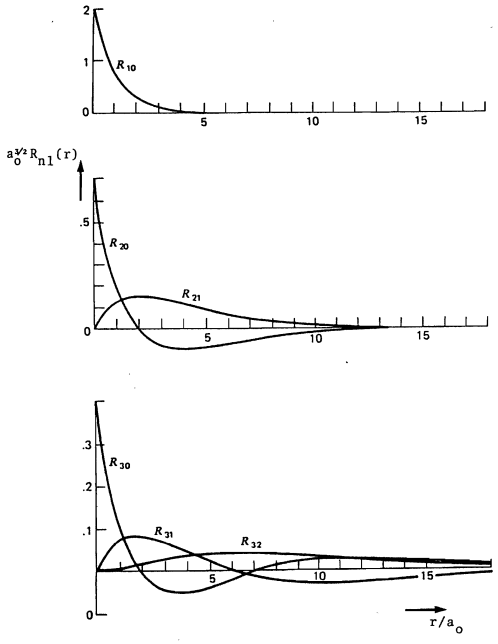

Figure 4.10: Radial wavefunctions $R_{nl}(r)$ of the hydrogen atom.

4.5.3 Stationary States of Hydrogen

In total we found the stationary states, or the energy eigenfunctions, of the hydrogen atom. Those are

$$\psi_{nlm}(r,\vartheta,\varphi) = R_{nl}(r)\ Y_l^m(\vartheta,\varphi). \tag{4.128}$$

The lower order wave functions are listed in Table 4.4 and plots of the resulting probability densities of the lowest order energy eigenstates of the hydrogen atom are shown in Figure 4.11

$$\psi_{100}(r,\vartheta,\varphi) = \frac{1}{\sqrt{\pi}\sqrt{a_0^3}} e^{-r/a_0}$$

$$\psi_{200}(r,\vartheta,\varphi) = \frac{1}{4\sqrt{2\pi}\sqrt{a_0^3}} \left(2 - \frac{r}{a_0}\right) e^{-r/2a_0}$$

$$\psi_{210}(r,\vartheta,\varphi) = \frac{1}{4\sqrt{2\pi}\sqrt{a_0^3}} \frac{r}{a_0} e^{-r/2a_0} \cos\vartheta$$

$$\psi_{21\pm1}(r,\vartheta,\varphi) = \frac{1}{8\sqrt{\pi}\sqrt{a_0^3}} \frac{r}{a_0} e^{-r/2a_0} \sin\vartheta e^{\pm \mathrm{j}\varphi},$$

$$\psi_{300}(r,\vartheta,\varphi) = \frac{1}{81\sqrt{3\pi}\sqrt{a_0^3}} \left(27 - 18\frac{r}{a_0} + 2\frac{r^2}{a_0^2}\right) e^{-r/3a_0}$$

Table 4.4: Lowest order hydrogen wavefunctions $\psi_{n,l,m}(r,\vartheta,\varphi)$.

4.5.4 Energy Spectrum of Hydrogen

We haven't yet discussed the energy eigenspectrum of hydrogen. From Eqs.(4.113) and (4.118) we find this to be

$$E = -\frac{m_0 e^4}{8\ \varepsilon_0^2 h^2} \frac{1}{n^2}, \tag{4.129}$$

which also agrees with the energy spectrum of the Bohr-Sommerfeld model, see section 3.4. The lowest energy eigenstate is

$$E_1 = -\frac{m_0 e^4}{8\ \varepsilon_0^2 h^2} = -13.7\mathrm{eV}. \tag{4.130}$$

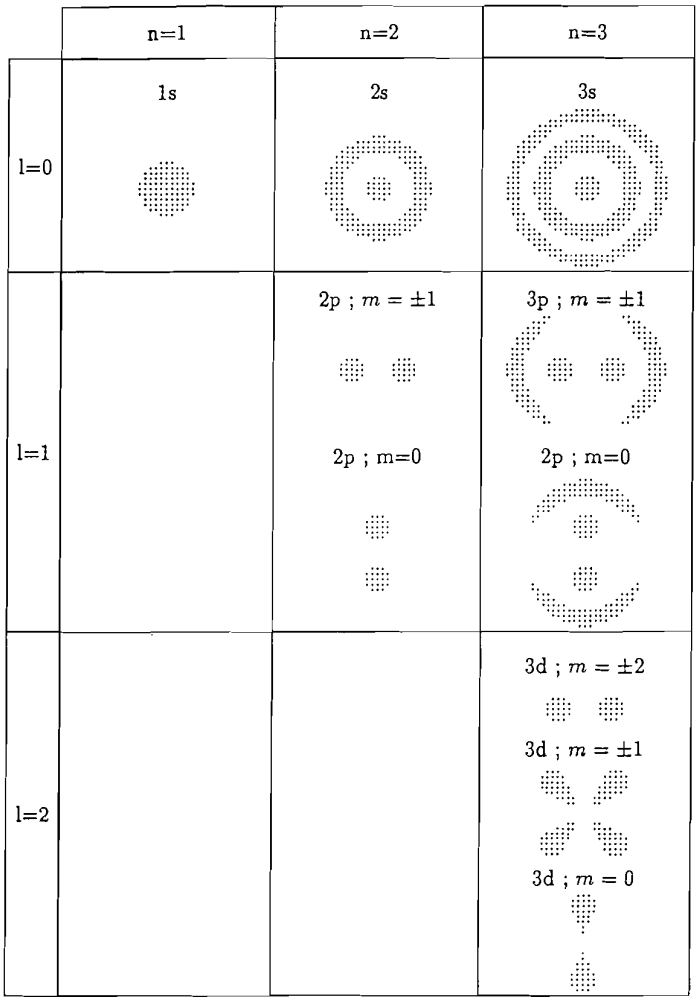

Figure 4.11: Probability densities of the lowest order hydrogen wavefunctions. (The density is presented along the meridial plane).

$$\psi_{310}(r,\vartheta,\varphi) = \frac{1}{81\sqrt{\pi}\sqrt{a_0^3}}\left(6-\frac{r}{a_0}\right)e^{-r/3a_0}\cos\vartheta$$

$$\psi_{31\pm1}(r,\vartheta,\varphi) = \frac{1}{81\sqrt{\pi}\sqrt{a_0^3}}\left(6-\frac{r}{a_0}\right)\frac{r}{a_0}e^{-r/3a_0}\sin\vartheta e^{\pm\mathrm{j}\varphi}$$

$$\psi_{320}(r,\vartheta,\varphi) = \frac{1}{81\sqrt{6\pi}\sqrt{a_0^3}}\frac{r^2}{a_0^2}e^{-r/3a_0}\left(3\cos^2\vartheta-1\right)$$

$$\psi_{32\pm1}(r,\vartheta,\varphi) = \frac{1}{81\sqrt{\pi}\sqrt{a_0^3}}\frac{r^2}{a_0^2}e^{-r/3a_0}\sin\vartheta\cos\vartheta e^{\pm\mathrm{j}\varphi},$$

$$\psi_{32\pm2}(r,\vartheta,\varphi) = \frac{1}{162\sqrt{3\pi}\sqrt{a_0^3}}\frac{r^2}{a_0^2}e^{-r/3a_0}\sin^2 e^{\pm2\mathrm{j}\varphi}$$

Table 4.5: Lowest order hydrogen wavefunctions $\psi_{n,l,m}(r,\vartheta,\varphi)$.continued.

The energy eigenvalues constitute a sequence that converges for large $n \to \infty$ towards 0, which corresponds to removing the electron from the atom. The energy to do so is $E_{\infty 1} - E_1 = 13.7\mathrm{eV}$.

Figure 4.12 shows the energy levels and the term diagram of the hydrogen atom and how the Lyman, Balmer, Paschen, Brackett and Pfund series arise from it. Each wavefunction is uniquely described by the set of quantum numbers (n,l,m). The first quantum number n specifies the energy eigen value E_n. As we will show in problem sets, the second quantum number l determines the eigenvalue of the squared angular momentum operator $\vec{L}^2$ with eigenvalues

$$\vec{L}^2\ \psi_{nlm}(r,\vartheta,\varphi) = l(l+1)\hbar^2\ \psi_{nlm}(r,\vartheta,\varphi)\,, \tag{4.131}$$

and the third quantum number m detemines the eigenvalue of the operator describing the z-component of the angular momentum operator

$$L_z\ \psi_{nlm}(r,\vartheta,\varphi) = m\hbar\ \psi_{nlm}(r,\vartheta,\varphi)\,. \tag{4.132}$$

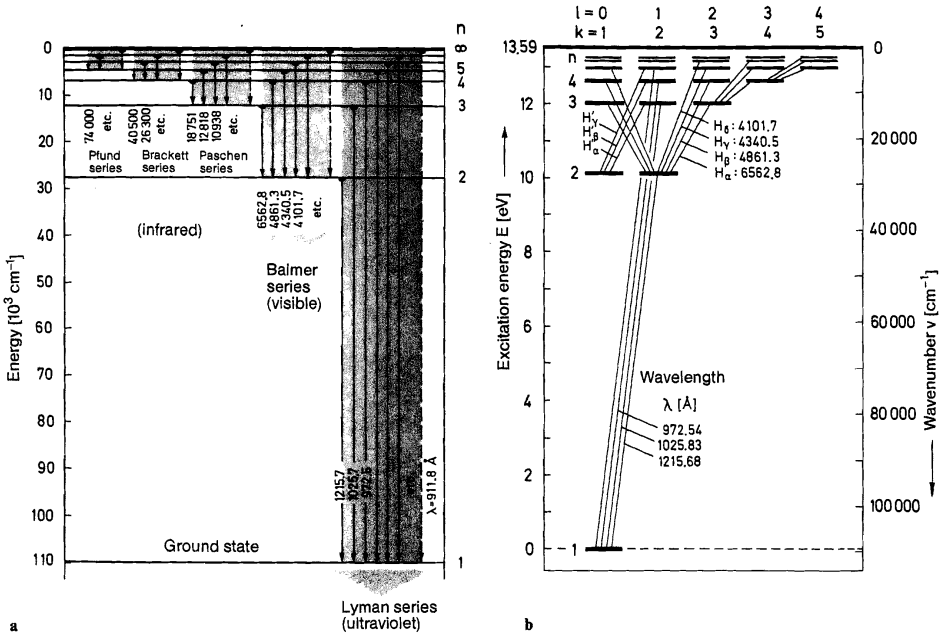

Figure 4.12: Energy levels and term diagram for the hydrogen atom [3]

In fact, the description of the electron wave functions is not yet complete, because the electron has an internal degree of freedom, that is its spin. The spin is an internal angular momentum of the electron that carries a magnetic moment with it. The Stern-Gerlach experiment shows that this degree of freedom has two eigenstates, i.e. the spin can be oriented parallel or anti-parallel to the direction of an applied magnetic field. The values of the internal angluar mometum with respect to the quantization axis defined by an external field, that shall be chosen along the z-axis, are $s = \pm\hbar/2$. Thus the energy eigenstates of an electron in hydrogen are uniquely characterized by four quantum numbers, n, l, m, and s. As Figure 4.12 shows, the energy

spectrum is degenerate, i.e. for $n > 1$, there exist to each energy eigenvalue several eigenfunctions, that are only uniquely characterized by the additional quantum numbers for angular momentum and spin. This is called degeneracy because there exist to a given energy eigenvalue several states.

4.6 Wave Mechanics

In this section, we generalize the concepts we have learned in the previous sections. The goal here is to give a broader description of quantum mechanics in terms of wave functions that are solutions to the Schroedinger Equation.

In classical mechanics the particle state is determined by its position and momentum and the state evolution is determined by Newton's law. In quantum mechanics the particle state is completely described by its wave function and the state evolution is determined by the Schroedinger equation.

The wave function as a complete description of the particle enables us to compute expected values of physical quantities of the particle when a corresponding measurement is performed. The measurement results are real numbers, like the energy, o4 position or momentum the particle has in this state. The physically measureable quantities are called observables. In classical mechanics these observables or real variables like x for position, p for momentum or functions thereof, like the energy, which is called the Hamiltonian $H(p,x) = \frac{p^2}{2m} + V(x)$ in classical mechanics. For simplicity, we state the results only for one-dimensional systems but it is straight forward to extend these results to multi-dimensional sytems. In quantum mechanics these observables become operators:

$$x \quad : \quad \text{position operator} \tag{4.133}$$

$$p \quad = \quad \frac{\hbar}{\mathrm{j}}\frac{\partial}{\partial x} : \text{ momentum operator} \tag{4.134}$$

$$H(p,x) \quad = \quad -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) : \text{ Hamiltonian operator} \tag{4.135}$$

If we carry out measurements of these observables, the result is a real number in each measurement and after many measurements on identical systems we can make a statistics of these measurements and the statistics is completely described by the moments of the observable.

4.6.1 Position Statistics

The statistical interpretation of quantum mechanics enables us to compute the expected value of the position operator or any of its moments according

to

$$\langle x \rangle = \int_{-\infty}^{\infty} \Psi^* (x,t)\ x\ \Psi (x,t)\ dx \tag{4.136}$$

$$\langle x^m \rangle = \int_{-\infty}^{\infty} \Psi^* (x,t)\ x^m\ \Psi (x,t)\ dx \tag{4.137}$$

The expectation value of functions of operators can always be evaluated by defining the operator by its Taylor expansion

$$\begin{aligned} \langle f(x) \rangle &= \int_{-\infty}^{\infty} \Psi^* (x,t)\ f(x)\ \Psi (x,t)\ dx \\ &= \left\langle \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0)\ x^n \right\rangle \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0) \left\langle \int_{-\infty}^{\infty} \Psi^* (x,t)\ x^n\ \Psi (x,t)\ dx \right\rangle \end{aligned} \tag{4.138}$$

4.6.2 Momentum Statistics

The momentum statistics is then

$$\langle p \rangle = \int_{-\infty}^{\infty} \Psi^* (x,t)\ \frac{\hbar}{\mathrm{j}} \frac{\partial}{\partial x}\ \Psi (x,t)\ dx \tag{4.139}$$

which can be written in terms of the wave function in the wave number space, which we define now for symmetry reasons as the Fourier transform of the wave function where the 2π is symmetrically distributed between Fourier and inverse Fourier transform

$$\phi (k,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Psi (x,t)\ e^{-\mathrm{j}kx}\ dx, \tag{4.140}$$

$$\Psi (x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi (k,t)\ e^{\mathrm{j}kx}\ dk. \tag{4.141}$$

Using the differentiation theorem of the Fourier transform and the generalized Parseval relation

$$\int_{-\infty}^{\infty} \phi_1^* (k)\ \phi_2 (k)\ dk = \int_{-\infty}^{\infty} \Psi_1^* (x)\, \Psi_2 (x)\ dx \tag{4.142}$$

we find

$$\langle p \rangle = \int_{-\infty}^{\infty} \phi^* (k,t) \ \hbar k \ \phi (k,t) \ dk \tag{4.143}$$

$$= \int_{-\infty}^{\infty} \hbar k \ |\phi (k,t)|^2 \ dk. \tag{4.144}$$

The introduction of the symmetrically defined expectation value of an operator according Eq.(4.136), where x can stand for any operator can be carried out using the wave function in the position space or the wave number space using the corresponding represenation of the wave function and of the operator.

4.6.3 Energy Statistics

The analysis for the measurement of position or moment carries over to every observable in an analogous way. Thus the expectation value of the energy is

$$\langle H(x,p) \rangle = \int_{-\infty}^{\infty} \Psi^* (x,t) \ H(x,p) \ \Psi (x,t) \ dx \tag{4.145}$$

$$= \int_{-\infty}^{\infty} \Psi^* (x,t) \ \left(-\frac{1}{2m\hbar^2} \frac{\partial^2}{\partial x^2} + V(x) \right) \Psi (x,t) \ dx \tag{4.146}$$

If the system is in an energy eigenstate, i.e.

$$\Psi (x,t) \ = \psi_n (x) \ e^{\mathrm{j}\omega_n t} \tag{4.147}$$

with

$$H(x,p) \ \psi_n (x) = E_n \ \psi_n (x) , \tag{4.148}$$

we obtain

$$\langle H(x,p) \rangle = \int_{-\infty}^{\infty} \Psi^* (x,t) \ E_n \ \Psi (x,t) \ dx = E_n. \tag{4.149}$$

If the system is in a superposition of energy eigenstates

$$\Psi (x,t) = \sum_{n=0}^{\infty} c_n \psi_n (x) e^{\mathrm{j}\omega_n t}. \tag{4.150}$$

we obtain

$$\langle H(x,p) \rangle = \sum_{n=0}^{\infty} E_n \ |c_n|^2 . \tag{4.151}$$

4.6.4 Arbitrary Observable

There may also occur observables that are not simple to translate from the classical to the quantum domain, such as the product

$$p_{cl} \cdot x_{cl} = x_{cl} \cdot p_{cl} \tag{4.152}$$

Classically it does not matter which variable comes first. However, if we tranfer this expression into quantum mechanics, the corresponding operator depends on the odering, for example

$$\begin{aligned} p_{qm} \ \cdot \ x_{qm}\Psi(x,t) &= \frac{\hbar}{\mathrm{j}}\frac{\partial}{\partial x} \ (x\Psi(x,t)) = & (4.153)\\ &= \frac{\hbar}{\mathrm{j}}\Psi(x,t) + \ \frac{\hbar}{\mathrm{j}}x\frac{\partial}{\partial x}\Psi(x,t)\,, & (4.154)\\ &= \left(\frac{\hbar}{\mathrm{j}} + \ x_{qm} \cdot p_{qm}\right)\Psi(x,t)\,. & (4.155) \end{aligned}$$

The decision of which expression represents the correct quantum mechanical operator or eventually even a linear combination of the possible expressions, has to be based on a close examination of the actual measurement apparatus that would measure the corresponding observable. Finally, the expression also has to deliver results that are in agreement with experimental findings.

If we have an operator that is a function of x and p and we have decided on a unique expression in terms of a power expansion in x and p

$$g(x,p) \rightarrow \ g_{op}(x, \ \frac{\hbar}{\mathrm{j}}\frac{\partial}{\partial x}) \tag{4.156}$$

then we can compute its expected value either in the space domain or the wave number domain

$$\begin{aligned} \langle g_{op}\rangle &= \int_{-\infty}^{\infty}\Psi^*(x,t) \ g_{op}(x, \ \frac{\hbar}{\mathrm{j}}\frac{\partial}{\partial x}) \ \Psi(x,t) \ dx & (4.157)\\ &= \int_{-\infty}^{\infty}\phi^*(k,t) \ g_{op}(j\frac{\partial}{\partial k}, \ \hbar k) \ \phi(k,t) \ dk & (4.158) \end{aligned}$$

That is this operator can be represented either in real space or in k-space as $g_{op}(x, \ \frac{\hbar}{\mathrm{j}}\frac{\partial}{\partial x})$ or $g_{op}(j\frac{\partial}{\partial k}, \ \hbar k)$.

4.6.5 Eigenfunctions and Eigenvalues of Operators

A differential operator has in general eigenfunctions and corresponding eigenvalues

$$g_{op}(x,\ \frac{\hbar}{\mathrm{j}}\frac{\partial}{\partial x})\ \psi_n(x) = g_n\psi_n(x)\,, \tag{4.159}$$

where g_n is the eigenvalue to the eigenfunction $\psi_n(x)$. An example for a differential operator is the Hamiltonian operator describing a partical moving in a potential

$$H_{op} = -\frac{1}{2m\hbar^2}\frac{\partial^2}{\partial x^2} + V(x) \tag{4.160}$$

the corresponding eigenvalue equation is the stationary Schroedinger Equation

$$H_{op}\psi_n(x) = E_n\psi_n(x)\,. \tag{4.161}$$

Thus the energy levels of a quantum system are the eigenvalues of the corresponding Hamiltonian operator.

The operator for which

$$\int \psi_n^*(x)\,(H_{op}\psi_m(x))\,dx = \int (H_{op}\psi_n(x))^*\,\psi_m(x)\,dx, \tag{4.162}$$

for arbitrary wave functions ψ_n and ψ_m is called a hermitian operation. From this equation we find immediately that the expected values of a hermitian operator are real, which also has the consequence that the eigenvalues of hermitian operators are real. This is important since operators that represent observables must have real expected values and real eigenvalues since these are results of physical measurements, which are real. Thus observables are represented by hermitian operators. This is easy to proove. Let's assume we have found two eigenfunctions and the corresponding eigen values

$$g_{op}\psi_m = g_m\psi_m, \tag{4.163}$$

$$g_{op}\psi_n = g_n\psi_n. \tag{4.164}$$

Then

$$\int \psi_n^* g_{op}\psi_m\ dx = g_m\int \psi_n^*\psi_m. \tag{4.165}$$

By taking advantage of the fact that the operator is hermitian we can also write

$$\int \psi_n^* g_{op} \psi_m \, dx = \int \left(g_{op}\psi_n\right)^* \psi_m \, dx = g_n^* \int \psi_n^* \psi_m \, dx \tag{4.166}$$

The right sides of Eqs.(4.165) and (4.166) must be equal

$$\left(g_m - g_n^*\right) \int \psi_n^* \psi_m \, dx = 0 \tag{4.167}$$

If $n = m$ the integral can not vanish and Eq.(4.167) enforces $g_n = g_n^*$, i.e. the corresponding eigenvalues are real. If $n \neq m$ and the corresponding eigenvalues are not degenerate, i.e. different eigenfunctions have different eigenvalues, then Eq.(4.167) enforces that the eigenfunctions are orthogonal to each other

$$\int \psi_n^* \psi_m \, dx = 0, \text{ for } n \neq m. \tag{4.168}$$

Thus, if there is no degneracy, the eigenfunctions of a hermitian operator are orthogonal to each other. If there is degeneracy, one can always choose an orthogonal set of eigenfunctions. If the eigenfunctions are properly normalized $\int \psi_n^* \psi_n \, dx = 1$, then the eigenfunctions build an orthonormal system

$$\int \psi_n^* \psi_m \, dx = \delta_{nm}, \tag{4.169}$$

and are complete, i.e. any arbitrary function $f(x)$ can be expressed as a superposition of the orthonormal basis functions $\psi_n(x)$

$$f(x) = \sum_{n=0}^{\infty} c_n \psi_n(x). \tag{4.170}$$

Thus we can freely change the basis in which we describe a certain physical problem. To account fully for this fact, we no longer wish to use wave mechanics, ie. express the wave function as a function in position space or in k-space. Instead we will utilize a vector in an abstract function space, i.e. a Hilbert space. In this way, we can formulate a physical problem, without using a fixed representation for the state of the system (wave function) and the corresponding operator representations. This description enables us to make full use of the mathematical structure of Hilbert spaces and the algebraic properties of operators.

Bibliography

[1] Introduction to Quantum Mechanics, Griffiths, David J., Prentice Hall, 1995.

[2] Quantum Mechanics I, C. Cohen-Tannoudji, B. Diu, F. Laloe, John Wiley and Sons, Inc., 1978.

[3] The Physics of Atoms and Quanta, Haken and Wolf, Springer Verlag 1994.

[4] Practical Quantum Mechanics, S. Flügge, Springer Verlag, Berlin, 1999.

[5] Handbook of Mathematical Functions, Abramowitz and Stegun, Dover Publications, NY 1970.

Chapter 5

The Dirac Formalism and Hilbert Spaces

In the last chapter we introduced quantum mechanics using wave functions defined in position space. We identified the Fourier transform of the wave function in position space as a wave function in the wave vector or momentum space. Expectation values of operators that represent observables of the system can be computed using either representation of the wavefunction. Obviously, the physics must be independent whether represented in position or wave number space. P.A.M. Dirac was the first to introduce a representation-free notation for the quantum mechanical state of the system and operators representing physical observables. He realized that quantum mechanical expectation values could be rewritten. For example the expected value of the Hamiltonian can be expressed as

$$\int \psi^*(x)\ H_{op}\ \psi(x)\, dx = \langle\psi| H_{op} |\psi\rangle\,, \tag{5.1}$$

$$= \langle\psi|\, \varphi\rangle\,, \tag{5.2}$$

with

$$|\varphi\rangle = H_{op}\,|\psi\rangle\,. \tag{5.3}$$

Here, $|\psi\rangle$ and $|\varphi\rangle$ are vectors in a Hilbert-Space, which is yet to be defined. For example, complex functions of one variable, $\psi(x)$, that are square inte-

grable, i.e.

$$\int \psi^* (x) \psi (x) dx < \infty, \tag{5.4}$$

form the Hilbert-Space of square integrable functions denoted as L^2. In Dirac notation this is

$$\int \psi^* (x) \psi (x) dx = \langle \psi | \psi \rangle . \tag{5.5}$$

Orthogonality relations can be rewritten as

$$\int \psi_m^* (x) \psi_n (x) dx = \langle \psi_m | \psi_n \rangle = \delta_{mn}. \tag{5.6}$$

As see above expressions look like a bracket he called the vector $|\psi_n\rangle$ a ket-vector and $\langle \psi_m|$ a bra-vector.

5.1 Hilbert Space

A Hilbert Space is a linear vector space, i.e. if there are two elements $|\varphi\rangle$ and $|\psi\rangle$ in this space the sum of the elements must also be an element of the vector space

$$|\varphi\rangle + |\psi\rangle = |\varphi + \psi\rangle \ . \tag{5.7}$$

The sum of two elements is commutative and associative

$$Commutative: \quad |\varphi\rangle + |\psi\rangle = |\psi\rangle + |\varphi\rangle \ , \tag{5.8}$$

$$Associative: \quad |\varphi\rangle + |\psi + \chi\rangle = |\varphi + \psi\rangle + |\chi\rangle \quad . \tag{5.9}$$

The product of the vector with a complex quantity c is again a vector of the Hilbert-Space

$$c |\varphi\rangle \equiv |c\varphi\rangle \quad . \tag{5.10}$$

The product between vectors and numbers is distributive

$$Distributive: \quad c |\varphi + \psi\rangle = c |\varphi\rangle + c |\psi\rangle \quad . \tag{5.11}$$

In short every linear combination of vectors in a Hilbert space is again a vector in the Hilbert space.

5.1.1 Scalar Product and Norm

There is a bilinear form defined by two elments of the Hilbert Space $|\varphi\rangle$ and $|\psi\rangle$, which is called a scalar product resulting in a complex number

$$\langle\varphi|\,\psi\rangle = a\,. \tag{5.12}$$

Ths scalar product obtained by exchanging the role of $|\varphi\rangle$ and $|\psi\rangle$ results in the complex conjugate number

$$\langle\psi\,|\varphi\rangle = \langle\varphi\,|\psi\rangle\,^{*} = a^{*}\,. \tag{5.13}$$

The scalar product is distributive

$$\textit{Distributive}: \quad \langle\varphi\,|\psi_1+\psi_2\rangle = \langle\varphi\,|\psi_1\rangle + \langle\varphi\,|\psi_2\rangle \quad . \tag{5.14}$$

$$\langle\varphi\,|c\psi\rangle = c\,\langle\varphi\,|\psi\rangle \quad . \tag{5.15}$$

And from Eq.(5.13) follows

$$\langle c\psi|\,\varphi\rangle = \langle\varphi|\,c\psi\rangle^{*} = c^{*}\,\langle\psi|\,\varphi\rangle \ . \tag{5.16}$$

Thus if the complex number is pulled out from a bra-vector it becomes its complex conjugate. The bra- and ket-vectors are hermitian, or adjoint, to each other. The adjoint vector is denoted by the symbol^{+}

$$(|\varphi\rangle)^{+} = \langle\varphi| \ \ , \tag{5.17}$$

$$(\langle\varphi|)^{+} = |\varphi\rangle \quad . \tag{5.18}$$

The vector spaces of bra- and ket-vectors are dual to each other. To transform an arbitrary expression into its adjoint, one has to replace all operators and vectors by the adjoint operator or vector and in addition the order of the elements must be reversed. For example

$$(c|\varphi\rangle)^{+} = c^{*}\,\langle\varphi| \ \ , \tag{5.19}$$

$$\langle\varphi|\,\psi\rangle^{+} = \langle\varphi|\,\psi\rangle^{*} = \langle\psi|\,\varphi\rangle \ . \tag{5.20}$$

This equation demands that a scalar product of a vector with itself is always real. Here, we even demand that it is positive

$$\langle\varphi\,|\varphi\rangle \geqslant 0, \quad \textit{real} \quad . \tag{5.21}$$

The equal sign in Eq.(5.21) is only fulfilled for the null vector, which is defined by

$$|\varphi\rangle + 0 = |\varphi\rangle \,. \tag{5.22}$$

Note, we denote the null vector not with the symbol $|0\rangle$ but rather with the scalar 0. Because the symbol $|0\rangle$ is reserved for the ground state of a system.

If the scalar product of a vector with itself is always positive, Eq.(5.21), then one can derive from the scalar product the norm of a vector according to

$$\|\varphi\| = \sqrt{\langle\varphi\,|\varphi\rangle}\,. \tag{5.23}$$

For vectors that are orthogonal to each other we have

$$\langle\varphi\,|\psi\rangle = 0 \tag{5.24}$$

without having one of them be the null vector.

5.1.2 Vector Bases

The dimensions of a Hilbert space are countable, i.e. each dimension can be assigned a whole number and thereby all dimensions are referenced in a unique way with $1, 2, 3, ...$ A vector space that is a Hilbert space has the following additional properties.

Completeness:

If there is a sequence of vectors in a Hilbert space $|\varphi_1\rangle\,, |\varphi_2\rangle\,, |\varphi_3\rangle\,, |\varphi_4\rangle\,, ...$ that fulfills Cauchy's convergence criterion then the limit vector $|\varphi\rangle$ is also an element of the Hilbert space. Cauchy's convergence criterion states that if $\|\varphi_n - \varphi_m\| < \varepsilon$, for some $n, m > N(\varepsilon)$ the sequence converges uniformly [2].

Separability:

The Hilbert space is separable. This indicates that for every element $|\varphi\rangle$ in the Hilbert space there is a sequence with $|\varphi\rangle$ as the limit vector.

Every vector in the Hilbert space can be decomposed into linear independent basis vectors $|\psi_n\rangle$. The number of basis vectors can be infinite

$$|\varphi\rangle = \sum_n c_n\,|\psi_n\rangle\,. \tag{5.25}$$

The components c_n of the vector $|\varphi\rangle$ with respect to the basis $|\psi_n\rangle$ are complex numbers denoted with index n. It is advantageous to orthonormalize the basis vectors

$$\langle\psi_n|\ \psi_m\rangle = \delta_{nm}\,. \tag{5.26}$$

The components of the vector $|\varphi\rangle$ can then be determined easily by

$$\langle\psi_m|\ \varphi\rangle = \sum_n c_n\,\langle\psi_m|\ \psi_n\rangle = \sum_n c_n\delta_{mn}, \tag{5.27}$$

or

$$c_m = \langle\psi_m|\ \varphi\rangle\,, \tag{5.28}$$

This leads to

$$|\varphi\rangle = \sum_n |\psi_n\rangle\,\langle\psi_n|\ \varphi\rangle\,. \tag{5.29}$$

5.2 Linear Operators in Hilbert Spaces

5.2.1 Properties of Linear Operators

An operator $\mathbf{L}$ is defined as a mapping of a vector $|\varphi\rangle$ onto another vector $|\psi\rangle$ of the Hilbert space

$$\mathbf{L}\,|\varphi\rangle = |\psi\rangle\,. \tag{5.30}$$

A linear operator $\mathbf{L}$ has the property that it maps a linear combination of input vectors to the linear combination of the correponding maps

$$\begin{aligned}\mathbf{L}\,(a\ |\varphi_1\rangle + b\ |\varphi_2\rangle) &= (a\ \mathbf{L}\,|\varphi_1\rangle + b\ \mathbf{L}\,|\varphi_2\rangle)\\ &= a\ |\psi_1\rangle + b\ |\psi_2\rangle\,,\ \text{for } a,b\in\mathbb{C}.\end{aligned} \tag{5.31}$$

The sum of two linear operators is defined as

$$(\mathbf{L}+\mathbf{M})\ |\varphi\rangle = \mathbf{L}\ |\varphi\rangle + \mathbf{M}\ |\varphi\rangle\,. \tag{5.32}$$

And the product of two operators is defined as

$$\mathbf{L}\,\mathbf{M}\ |\varphi\rangle = \mathbf{L}\ (\mathbf{M}\ |\varphi\rangle)\,. \tag{5.33}$$

The null element and 1-element of the operators is denoted as $\mathbf{0}$, and $\mathbf{1}$. Often we will not bold face these operators, especially in products, where a

scalar has the same meaning. The two operators are defined by their action on arbitrary vectors of the Hilbert space

$$\mathbf{0}\,|\varphi\rangle \;=\; 0,\ \forall\ |\varphi\rangle\,, \tag{5.34}$$

$$\mathbf{1}\,|\varphi\rangle \;=\; |\varphi\rangle\,,\ \forall\ |\varphi\rangle \tag{5.35}$$

In generally, the multiplication of two operators is not commutative

$$\mathbf{L}\ \mathbf{M}\ \ |\varphi\rangle \neq\ \mathbf{ML}\ \ |\varphi\rangle\,, \forall\ |\varphi\rangle\,, \tag{5.36}$$

or in short

$$\mathbf{L}\ \mathbf{M}\ \neq\ \mathbf{M}\ \mathbf{L}. \tag{5.37}$$

The expression

$$[\mathbf{L},\mathbf{M}] = \mathbf{L}\ \mathbf{M}\ -\ \mathbf{M}\ \mathbf{L} \tag{5.38}$$

is therefore called the commutator between $\mathbf{L}$ and $\mathbf{M}$. If $[\mathbf{L},\mathbf{M}] = \mathbf{0}$, the operators commute. The following rules for commutators apply:

$$[\mathbf{L},\mathbf{M}] = -\,[\mathbf{M},\mathbf{L}]\ \ , \tag{5.39}$$

$$[\mathbf{L},\mathbf{L}] = \mathbf{0}\ , \tag{5.40}$$

$$[\mathbf{L},\mathbf{1}] = \mathbf{0}\ , \tag{5.41}$$

$$\left[\mathbf{L},\mathbf{L}^{-1}\right] = \mathbf{0}\ , \tag{5.42}$$

$$[\mathbf{L},a\mathbf{M}] = a\ \ [\mathbf{L},\mathbf{M}]\ \ , \tag{5.43}$$

$$[\mathbf{L}_1+\mathbf{L}_2,\mathbf{M}] = [\mathbf{L}_1,\mathbf{M}] + [\mathbf{L}_2,\mathbf{M}]\ \ , \tag{5.44}$$

$$[\mathbf{L},\mathbf{M}] = -\,[\mathbf{M},\mathbf{L}] \tag{5.45}$$

$$[\mathbf{L}_1\mathbf{L}_2,\mathbf{M}] = [\mathbf{L}_1,\mathbf{M}]\ \ \mathbf{L}_2 + \mathbf{L}_1\ \ [\mathbf{L}_2,\mathbf{M}]\ \ , \tag{5.46}$$

$$[\mathbf{M},\mathbf{L}_1\mathbf{L}_2] = [\mathbf{M},\mathbf{L}_1]\ \ \mathbf{L}_2 + \mathbf{L}_1\ \ [\mathbf{M},\mathbf{L}_2]\,. \tag{5.47}$$

Often the anticommutator is also used. It is defined as

$$[\mathbf{L},\mathbf{M}]_+ = \mathbf{L}\ \mathbf{M}\ +\ \mathbf{M}\ \mathbf{L}\ . \tag{5.48}$$

If $[\mathbf{L},\mathbf{M}]_+ = \mathbf{0}$, the operators are called anti-commuting.

5.2.2 The Dyadic Product

Two vectors in the Hilbert space can not only be used to build a scalar product but rather what is called a dyadic product, which is an operator

$$|\alpha\rangle \langle\beta| . \tag{5.49}$$

The dyadic product is the formal product between a ket- and a bra-vector. If applied to a vector, it projects the vector onto the state $|\beta\rangle$ and generates a new vector in parallel to $|\alpha\rangle$ with a magnitude equal to the projection

$$|\alpha\rangle \langle\beta|\psi\rangle = \langle\beta|\psi\rangle \ |\alpha\rangle . \tag{5.50}$$

As we have seen from Eq.(5.29), if $|\psi_n\rangle$ built a complete orthonormal basis, then

$$|\varphi\rangle = \sum_n |\psi_n\rangle \langle\psi_n|\varphi\rangle , \ \forall \ |\varphi\rangle , . \tag{5.51}$$

Eq.(5.35) implies

$$\mathbf{1} = \sum_n |\psi_n\rangle \langle\psi_n| . \tag{5.52}$$

When applied to an operator from the left and right side

$$\begin{aligned} \mathbf{1 \ L \ 1} &= \sum_m |\psi_m\rangle \langle\psi_m| \left(\sum_n \mathbf{L} |\psi_n\rangle \langle\psi_n| \right) \\ &= \sum_m \sum_n L_{mn} |\psi_m\rangle \langle\psi_n| \end{aligned} \tag{5.53}$$

with the matrix elements

$$L_{mn} = \langle\psi_m| \mathbf{L} |\psi_n\rangle . \tag{5.54}$$

The matrix elments L_{mn} represent the operator in the chosen base $|\psi_n\rangle$. Once we choose a base and represent vectors and operators in term of this base, the components of the vector and the operator can be collected in a column vector and a matrix. The table below shows a comparison between a representation based on Hilbert space vectors and operators in term of vectors and matrices in an euclidian vector space. Initially matrix mechanics was

developed by Heisenberg independently from Schroedingers wave mechanics. The Dirac representation in terms of bra- and ket-vectors unifies them and shows that both forms are isomorph

Ket-vector	Column vector
$\|\varphi_a\rangle = \sum_n a_n \ \|\psi_n\rangle$	$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix}$
Bra-vector	Row vector
$\langle\varphi\| = \sum_n a_n^* \ \langle\psi_n\|$	$\begin{pmatrix} a_1^* & a_2^* & \cdots \end{pmatrix}$
Inner product	Scalar product
$\langle\varphi_a\|\varphi_b\rangle = \sum_m \sum_n a_m^* b_n \langle\psi_m\|\psi_n\rangle$	$\begin{pmatrix} a_1^* & a_2^* & \cdots \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \end{pmatrix}$
$= \sum_n a_n^* b_n$	$= a_1^* b_1 + a_2^* b_2 + \cdots$
Operator	Matrix
$L = \sum_{m,n} L_{mn} \|\psi_m\rangle \langle\psi_n\|$	$\begin{pmatrix} L_{11} & L_{12} & \cdots \\ L_{21} & L_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$
Dyadic Product	
$\|\varphi_a\rangle \langle\varphi_b\| = \sum_{m,n} a_m b_n^* \|\psi_m\rangle \langle\psi_n\|$	$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix} \cdot \begin{pmatrix} b_1^* & b_2^* & \cdots \end{pmatrix} =$
	$= \begin{pmatrix} a_1 b_1^* & a_1 b_2^* & \cdots \\ a_2 b_1^* & a_2 b_2^* & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$

5.2.3 Special Linear Operators

5.2.4 Inverse Operators

The operator inverse to a given operator $\mathbf{L}$ is denoted as $\mathbf{L}^{-1}$

$$\forall \, |\varphi\rangle \, , \ |\psi\rangle = \mathbf{L} \, |\varphi\rangle \ \implies |\varphi\rangle = \mathbf{L}^{-1} \, |\psi\rangle \, , \tag{5.55}$$

which leads to

$$\mathbf{L}\mathbf{L}^{-1} = \mathbf{1} \tag{5.56}$$

The inverse of a product is the product of the inverse in inverse order

$$(\mathbf{M}\mathbf{L})^{-1} = \mathbf{L}^{-1}\mathbf{M}^{-1} \tag{5.57}$$

5.2.5 Adjoint or Hermitian Conjugate Operators

The adjoint (hermitian conjugate) operator $\mathbf{L}^+$ is defined by

$$\langle\varphi|\,\mathbf{L}^+\,|\psi\rangle = (\langle\psi|\,\mathbf{L}\,|\varphi\rangle)^* = \langle\psi|\,\mathbf{L}\,|\varphi\rangle^*\,, \tag{5.58}$$

here $|\varphi\rangle$ and $|\psi\rangle$ are arbitrary vectors in a Hilbert space. Note, that if the adjoint of an expression is formed, each component gets conjugated and the order is reversed. For example

$$(\mathbf{L}\,|\varphi\rangle)^+ = \langle\varphi|\,\mathbf{L}^+, \tag{5.59}$$

$$\left(\mathbf{L}^+\,|\varphi\rangle\right)^+ = \langle\varphi|\,\mathbf{L} \tag{5.60}$$

If

$$|\mathbf{L}\varphi\rangle = \mathbf{L}\,|\varphi\rangle\,. \tag{5.61}$$

there is

$$\langle\mathbf{L}\varphi| = \langle\varphi|\,\mathbf{L}^+ \tag{5.62}$$

and

$$\langle\psi|\,\mathbf{L}\,|\varphi\rangle = \langle\psi|\,\mathbf{L}\varphi\rangle = \left\langle\mathbf{L}^+\psi\right|\varphi\rangle \tag{5.63}$$

The matrix elements of the adjoint operator are

$$L^+_{mn} = \langle\psi_m|\,\mathbf{L}^+\,|\psi_n\rangle = \langle\psi_n|\,\mathbf{L}\,|\psi_m\rangle = L^*_{nm} \tag{5.64}$$

The following rules apply to adjoint operators

$$\left(\mathbf{L}^+\right)^+ = \mathbf{L}, \tag{5.65}$$

$$(a\mathbf{L})^+ = a^*\mathbf{L}^+, \tag{5.66}$$

$$(\mathbf{L}+\mathbf{M})^+ = \mathbf{L}^+ + \mathbf{M}^+, \tag{5.67}$$

$$(\mathbf{L}\;\mathbf{M})^+ = \mathbf{M}^+\;\mathbf{L}^+. \tag{5.68}$$

5.2.6 Hermitian Operators

If the adjoint operator $\mathbf{L}^+$ is identical to the operator itself, then we call the operator hermitian

$$\mathbf{L} = \mathbf{L}^+. \tag{5.69}$$

Hermitian operators have real expected values. *Observables are represented by hermitian operators.*

5.2.7 Unitary Operators

If the inverse of an operator $\mathbf{U}$ is the adjoint operator

$$\mathbf{U}^{-1} = \mathbf{U}^+, \tag{5.70}$$

then this operator is called a unitary operator and

$$\mathbf{U}^+\mathbf{U} = \mathbf{U}\mathbf{U}^+ = \mathbf{1}. \tag{5.71}$$

If the operator $\mathbf{U}$ is unitary and $\mathbf{H}$ is a hermitian operator, then the product $\mathbf{U}\mathbf{H}\mathbf{U}^{-1}$ is also a hermitian operator.

$$\left(\mathbf{U}\mathbf{H}\mathbf{U}^{-1}\right)^+ = \left(\mathbf{U}^{-1}\right)^+ \mathbf{H}^+\mathbf{U}^+ = \mathbf{U}\mathbf{H}\mathbf{U}^{-1}. \tag{5.72}$$

5.2.8 Projection Operators

The dyadic product

$$\mathbf{P}_n = |\psi_n\rangle \langle\psi_n| , \tag{5.73}$$

is a projection operator $\mathbf{P}_n$ that projects a given state $|\varphi\rangle$ onto the unit vector $|\psi_n\rangle$

$$\mathbf{P}_n |\varphi\rangle = |\psi_n\rangle \quad \langle\psi_n|\varphi\rangle . \tag{5.74}$$

If $|\varphi\rangle$ is represented in the orthonormal base $|\psi_n\rangle$

$$|\varphi\rangle = \sum_n c_n |\psi_n\rangle , \tag{5.75}$$

we obtain

$$\mathbf{P}_n |\varphi\rangle = c_n |\psi_n\rangle \ . \tag{5.76}$$

By construction, projection operators are hermitian operators. Besides operators that project on vectors, there are also operators that project on subspaces of the Hilbert space

$$\mathbf{P}_U = \sum_U |\psi_n\rangle \ \langle\psi_n| \,. \tag{5.77}$$

Here, the orthonormal vectors $|\psi_n\rangle$ span the sub-space U. Projection operators are idempotent

$$\mathbf{P}_n^k = \mathbf{P}_n, \text{ for } k > 1. \tag{5.78}$$

5.3 Eigenvalues of Operators

In chapter 4, we studied the eigenvalue problem of differential operators. Here, we want to formulate the eigenvalue problem of operators in a Hilbert space. An operator $\mathbf{L}$ in a Hilbert space with eigenvectors $|\psi_n\rangle$ fulfills the equations

$$\mathbf{L}\,|\psi_n\rangle = L_n\,|\psi_n\rangle\,, \tag{5.79}$$

with eigenvalues L_n. If there exist several different eigenvectors to the same eigenvalue L_n, this eigenvalue is called degenerate. For example, the energy eigenfunctions of the hydrogen atom are degenerate with respect to the indices l and m. The set of all eigenvalues is called the eigenvalue spectrum of the operator $\mathbf{L}$. As shown earlier the eigenvalues of hermitian operators are real and the eigenvectors to different eigenvalues are orthogonal to each other, because

$$\langle\boldsymbol{\psi}_m|\,\mathbf{L}\,|\psi_n\rangle = L_n\,\langle\boldsymbol{\psi}_m\,|\psi_n\rangle = L_m\,\langle\boldsymbol{\psi}_m\,|\psi_n\rangle\,, \tag{5.80}$$

or

$$(L_n - L_m)\,\langle\boldsymbol{\psi}_m\,|\psi_n\rangle = 0. \tag{5.81}$$

If the eigenvectors of the operator $\mathbf{L}$ form a complete base of the Hilbert space, the operator $\mathbf{L}$ is represented in this base by a diagonal matrix

$$L_{mn} = \langle\psi_m|\,\mathbf{L}\,|\psi_n\rangle = L_n\,\langle\psi_m\,|\psi_n\rangle = L_n\,\delta_{mn} \tag{5.82}$$

The operator can then be written in its spectral representation

$$\mathbf{L} = \sum_n L_n\,|\psi_n\rangle \ \langle\psi_n| = \sum_n L_n P_n\,. \tag{5.83}$$

5.4 Eigenvectors of Commuting Operators

Two operators, $\mathbf{A}$ and $\mathbf{B}$, that commute with each other have a common set of eigenvectors. To proove this theorem, we assume that the eigenvalue spectrum of the operator $\mathbf{A}$ is non degenerate. The eigenvectors and eigenvalues of operator $\mathbf{A}$ are $|\psi_n\rangle$ and A_n, respectively

$$\mathbf{A}\,|\psi_n\rangle \;\; = A_n\,|\psi_n\rangle \;\; . \tag{5.84}$$

Using

$$[\mathbf{A},\mathbf{B}] = 0, \tag{5.85}$$

we find

$$\begin{aligned} \langle\psi_m|\,\mathbf{AB}-\mathbf{BA}\,|\psi_n\rangle &= 0 \qquad (5.86)\\ \langle\psi_m|\,\mathbf{A}\left(\sum_n |\psi_n\rangle\;\langle\psi_n|\right)\mathbf{B}-\mathbf{BA}\,|\psi_n\rangle &= 0 \\ (A_m-A_n)\,\langle\psi_m|\,\mathbf{B}\,|\psi_n\rangle &= (A_m-A_n)\,B_{mn}=0\;. \end{aligned}$$

Since the eigenvalues are assumed to be not degenerate, i.e. $A_m \neq A_n$, the matrix B_{mn} must be diagonal, which means that the vector $|\psi_n\rangle$ has also to be an eigenvector of operator $\mathbf{B}$. If the eigenvalues are degenerate, one can always choose, in the sub-space that belongs to the degenerate eigenvalue, a base that are also eigenvectors of $\mathbf{B}$. The operator $\mathbf{B}$ thus eventually has no degeneracies in this sub-space and therefore, the eigenvalues of $\mathbf{B}$ may help to uniquely characterize the set of joint eigenvectors.

Also the reverse is true. If two operators have a joint system of eigenvectors, they commute. This is easy to see from the spectral representation of both operators.

Example: We define the parity operator P_x which, when applied to a wave function of a particle in one dimension $\psi(x)$, changes the sign of the position x

$$P_x\psi(x) = \psi(-x). \tag{5.87}$$

The Hamiltonian of a particle in an inversion symmetric potential $V(x)$, i.e.

$$V(x) = V(-x), \tag{5.88}$$

commutes with the parity operator. Then the eigenfunctions of the Hamiltonian are also eigenfunctions of the parity operator. The eigenfunctions of the

parity operator are the symmetric or antisymmetric functions with eigenvalues 1 and -1, respectively. Therefore, the eigenfunctions of a Hamiltonian with symmetric potential has symmetric and antisymmetric eigenfunctions, see box potential and harmonic oscillator potential.

5.5 Complete System of Commuting Operators

In the case of the hydrogen atom, we had to use three qunatum numbers n, l, and m to characterize the energy eigenfunctions completely. Without proof, the indices l and m characterize the eigenvalues of the square of the angular momentum operator $\vec{L}^2$, and of the z-component of the angular moment L_z with eigenvalues $l\,(l+1)\,\hbar^2$ and $m\hbar$, respectively. One can show, that the Hamilton operator of the hydrogen atom, the square of the angular momentum operator and the z-component of the angular moment operator commute with each other and build a complete system of commuting operators (CSCO), whose eigenvalues enable a unique characterization of the energy eigenstates of the hydrogen atom.

5.6 Product Space

Very often in quantum mechanics one deals with interacting systems, for example system A and system B. The state space of each isolated system is Hilbert space H_A and Hilbert space H_B spanned by a complete base $|\psi_n\rangle_A$ and $|\psi_n\rangle_B$, respectively. $\mathbf{L}_A$ and $\mathbf{M}_B$ are operators on each of the Hilber spaces of the individual systems. The Hilbert space of the total system is the product space

$$H = H_A \otimes H_B. \tag{5.89}$$

The vectors in this Hilbert space are given by the direct product of the individual vectors and one could choose as a base in the product space

$$|\chi_{nm}\rangle = |\psi_n\rangle_A \otimes |\psi_m\rangle_B = |\psi_n\rangle_A\,|\psi_m\rangle_B\,. \tag{5.90}$$

Operators that only act on system A can be extended to operate on the product space by

$$\mathbf{L} = \mathbf{L}_A \otimes \mathbf{1}_B. \tag{5.91}$$

or similar for operators acting on system B

$$\mathbf{M} = \mathbf{1}_A \otimes \mathbf{M}_B. \tag{5.92}$$

The product of both operators is then

$$\mathbf{LM} = \mathbf{L}_A \otimes \mathbf{M}_B. \tag{5.93}$$

An operator in this product space acts on a vector in the following way

$$\mathbf{LM} \left|\chi_{nm}\right\rangle = \mathbf{L} \left|\psi_n\right\rangle_A \otimes \mathbf{M} \left|\psi_m\right\rangle_B . \tag{5.94}$$

Since, the vecotrs $\left|\psi_n\right\rangle_A$ and $\left|\psi_m\right\rangle_B$ build a complete base for system A and B, respectively, the product vectors in Eq.(5.90) build a complete base for the interacting system and each state can be written in terms of this base

$$\left|\chi\right\rangle = \sum_{m,n} a_{mn} \left|\chi_{nm}\right\rangle = \sum_{m,n} a_{mn} \left|\psi_n\right\rangle_A \otimes \left|\psi_m\right\rangle_B . \tag{5.95}$$

5.7 Quantum Dynamics

In chapter 4, we derived the Schroedinger Equation in the x-represenation. The stationary Schroedinger Equation was written as an eigenvalue problem to the Hamiltonian operator, which was then a differential operator. With the Dirac formulation we can rewrite these equations without refering to a special representation.

5.7.1 Schroedinger Equation

In the Dirac notation the Schroedinger Equation is

$$\mathrm{j}\hbar \frac{\partial \left|\Psi\left(t\right)\right\rangle}{\partial t} = \mathbf{H} \left|\Psi\left(t\right)\right\rangle . \tag{5.96}$$

$\mathbf{H}$ is the Hamiltonian operator; it determines the dynamics of the quantum system.

$$\mathbf{H} = \frac{\tilde{\mathbf{p}}^2}{2m} + \mathbf{V}(\tilde{\mathbf{x}}). \tag{5.97}$$

The Hamiltonian operator is the generator of motion in a quantum system. Here $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{x}}$ and functions of them are operators in the Hilbert space. $|\Psi(t)\rangle$ is the Hilbert space vector describing fully the system's quantum state at time t. When looking for states that have a harmonic temporal behaviour

$$|\Psi(t)\rangle = e^{\mathrm{j}E_n t/\hbar}\,|\psi_n\rangle\,, \tag{5.98}$$

we obtain the stationary Schroedinger Equation

$$\mathbf{H}\,|\psi_n\rangle = E_n\,|\psi_n\rangle\,, \tag{5.99}$$

that determines the energy eigenstates of the system. If the $|\psi_n\rangle$ build a complete basis of the Hilbert space, H, the system is dynamically evolving, the most general time dependent solution of the Schroedinger Equation is then a superposition of all energy eigenstates

$$|\Psi(t)\rangle = \sum_n a_n\; e^{\mathrm{j}E_n t/\hbar}\;\;|\psi_n\rangle\,. \tag{5.100}$$

5.7.2 Schroedinger Equation in x-representation

We can return to wave mechanics by rewriting the abstract Schroedinger Equation in the eigenbase $|x\rangle$ of the position operator. For simplicity in notation, we only consider the one dimensional case and define that there exists the following eigenvectors

$$\mathbf{x}\,|x'\rangle = x'\,|x'\rangle\,, \tag{5.101}$$

with the orthogonality relation

$$\langle x\,|x'\rangle = \delta(x-x'). \tag{5.102}$$

Note, since the position operator has a continuous spectrum of eigenvalues the orthogonality relation is a dirac delta function rather than a delta-symbol. The completness relation using this base is expressed in the unity operator

$$\mathbf{1} = \int |x'\rangle\,\langle x'|\,dx'\;, \tag{5.103}$$

rather then a sum as in Eq.(5.52). Inserting this unity operator in the Schroedinger Equation (5.96) and projecting from the left with $\langle x|$, we obtain

$$\mathrm{j}\,\hbar\frac{\partial}{\partial t}\,\langle x\,|\Psi(t)\rangle = \langle x|\;\mathbf{H}\int |x'\rangle\,\langle x'|\,\Psi(t)\;\rangle\,dx'. \tag{5.104}$$

The expression $\langle x \,|\Psi\,(t)\rangle$ is the probability amplitude that a position measurement on the system in state $|\Psi\,(t)\rangle$ yields the value x, which is precisely the meaning of the wave function

$$\Psi\,(x,t) = \langle x \,|\Psi\,(t)\rangle \tag{5.105}$$

in chapter 4. Using the eigenvalue property of the states and the orthogonality relations we obtain from Eq.(5.104)

$$\mathrm{j}\,\hbar\frac{\partial}{\partial t}\Psi\,(x,t) = H(x, p = \frac{\hbar}{\mathrm{j}}\frac{\partial}{\partial x})\Psi\,(x,t)\,. \tag{5.106}$$

5.7.3 Canonical Quantization

Thus the dynamics of a quantum system is fully determined by its Hamiltonian operator. The Hamiltonian operator is usually derived from the classical Hamilton function according to the Hamilton-Jacobi formulation of Classical Mechanics [3]. The classical Hamilton function $H(\{q_i\}\,,\{p_i\})$ is a function of the position coordinates of a particle x_i or generalized coordinates q_i and the corresponding momentum coordinates p_i. The classical equations of motion are found by

$$\dot{q}_i(t) = \frac{\partial}{\partial p_i}H(\{q_i\}\,,\{p_i\})\;, \tag{5.107}$$

$$\dot{p}_i(t) = -\frac{\partial}{\partial q_i}H(\{q_i\}\,,\{p_i\})\;. \tag{5.108}$$

In quantum mechanics the Hamiltonian function and the position and momentum coordinates become operators and quantization is achieved by imposing on position and momentum operators that are related to the same degree of freedom, for example the x-coordinate of a particle and the associate momentum p_x, canonical commutation relations

$$H(\{q_i\}\,,\{p_i\}) \;\Rightarrow\; \mathbf{H}(\{\mathbf{q}_i\}\,,\{\mathbf{p}_i\}), \tag{5.109}$$

$$[\mathbf{q}_i, \mathbf{p}_j] = \mathrm{j}\,\hbar\delta_{ij}. \tag{5.110}$$

Imposing this commutation relation implies that position and momentum related to one degree of freedom can not be measured simultaneously with arbitrary precision and Heisenberg's uncertainty relation applies to the possible states the system can take on.

5.7.4 Schroedinger Picture

In the Schroedinger picture the quantum mechanical state of the system is evolving with time. If there is no explicit time dependence in the operators then the operators stay time independent. The Schroedinger Equation (5.96)

$$\mathrm{j}\hbar\frac{\partial\left|\Psi\left(t\right)\right\rangle}{\partial t}=\mathbf{H}\left|\Psi\left(t\right)\right\rangle, \tag{5.111}$$

plus the initial state

$$\left|\Psi\left(t=0\right)\right\rangle=\left|\Psi\left(0\right)\right\rangle, \tag{5.112}$$

unquely determine the dynamics of the system. The evolution of the quantum state vector can be described as a mapping of the initial state by a time evolution operator.

$$\left|\Psi\left(t\right)\right\rangle=\ \mathbf{U}(t)\left|\Psi\left(0\right)\right\rangle. \tag{5.113}$$

If this solution is substituted into the Schroedinger Equation (5.111) it follows that the time evolution operator has to obey the equation

$$\mathrm{j}\,\hbar\frac{\partial}{\partial t}\mathbf{U}(t)=\mathbf{H}\ \mathbf{U}(t). \tag{5.114}$$

For a time independent Hamiltonian Operator the formal integration of this equation is

$$\mathbf{U}(t)=\exp\left[-\mathrm{j}\mathbf{H}t/\hbar\right]\ . \tag{5.115}$$

The time evolution operator is unitary

$$\mathbf{U}^{-1}(t)=\mathbf{U}^{+}(t)\ , \tag{5.116}$$

because the Harmiltonian operator is hermitian, and therefore the norm of an initial state is preserved. The initial value for the time evolution operator is

$$\mathbf{U}(t=0)=\mathbf{1}. \tag{5.117}$$

The expected value of an arbitrary operator $\mathbf{A}$ is given by

$$\left\langle\Psi\left(t\right)\right|\ \left.\Psi\left(t\right)\right\rangle=\left\langle\Psi\left(0\right)\right|\mathbf{U}^{+}(t)\mathbf{A}\mathbf{U}(t)\left|\Psi\left(0\right)\right\rangle. \tag{5.118}$$

5.7.5 Heisenberg Picture

Since the physically important quantities are the expected values, i.e. the outcome of experiments, Eq.(5.118) can be used to come up with an alternative formulation of quantum mechanics. In this formulation, called the Heisenberg picture, the operators are evolving in time according to

$$\mathbf{A}_H(t) = \mathbf{U}^+(t)\mathbf{A}\mathbf{U}(t), \tag{5.119}$$

and the state of the system is time independent and equal to its initial state

$$|\Psi_H(t)\rangle = |\Psi_S(0)\rangle . \tag{5.120}$$

Clearly an expected value for a time dependent operator using the Heisenberg state (5.120) is identical with Eq.(5.116).

This is identical to describing a unitary process in an eucledian vector space. Scalar products between vectors are preserved, if all vectors are undergoing a unitary transformation, i.e. a rotation for example. An alternative description is that the vectors are time independent but the coordinate system rotates in the opposite direction. When the coordinate system changes, the operators described in the time dependent coordinate system become time dependent themselves.

From the definition of the time evolution operator we find immediately an equation of motion for the time dependent operators of the Heisenberg picture

$$\begin{aligned} \mathrm{j}\hbar\frac{\partial \mathbf{A}_H(t)}{\partial t} &= \left(\mathrm{j}\,\hbar\frac{\partial \mathbf{U}^+(t)}{\partial t}\right)\mathbf{A}_S\mathbf{U}(t) + \mathbf{U}^+(t)\mathbf{A}_S\left(\mathrm{j}\,\hbar\frac{\partial \mathbf{U}(t)}{\partial t}\right) \\ &\quad +\mathbf{U}^+(t)\left(\mathrm{j}\hbar\frac{\partial \mathbf{A}_S}{\partial t}\right)\mathbf{U}(t) \end{aligned} \tag{5.121}$$

$$\begin{aligned} \mathrm{j}\hbar\frac{\partial \mathbf{A}_H(t)}{\partial t} &= -\mathbf{U}^+(t)\mathbf{H}^+\mathbf{A_S}\mathbf{U}(t) + \mathbf{U}^+(t)\mathbf{A}_S\mathbf{H}\mathbf{U}(t) \\ &\quad +\mathbf{U}^+(t)\left(\mathrm{j}\hbar\frac{\partial \mathbf{A}_S}{\partial t}\right)\mathbf{U}(t) \end{aligned} \tag{5.122}$$

Using the relation $\mathbf{U}^+(t)\mathbf{U}(t) = \mathbf{U}(t)\mathbf{U}^+(t) = \mathbf{1}$ and inserting it between the Hamiltonian operator and the operator $\mathbf{A}$, we finally end up with the

Heisenberg equations of motion for the Heisenberg operators

$$\begin{aligned} \mathrm{j}\hbar\frac{\partial}{\partial t}\mathbf{A}_H(t) &= -\mathbf{H}_H\mathbf{A}_H + \mathbf{A}_H\mathbf{H}_H + \mathrm{j}\hbar\left(\frac{\partial \mathbf{A}}{\partial \mathbf{t}}\right)_H & (5.123)\\ &= [\mathbf{A}_H, \mathbf{H}_H] + \mathrm{j}\,\hbar\left(\frac{\partial \mathbf{A}}{\partial \mathbf{t}}\right)_H & (5.124)\end{aligned}$$

with

$$\begin{aligned} \mathbf{H}_H &= \mathbf{U}^+(t)\mathbf{H}_S\mathbf{U}(t) & (5.125)\\ &= \mathbf{H}_S \ \text{ for conservative systems, i.e. } \mathbf{H}_S \neq \mathbf{H}_S(t) & (5.126)\end{aligned}$$

Note, that the last term in Eq.(5.124) is only present if the Schroedinger operators do have an explicit time dependence, a case which is beyond the scope of this class.

5.8 The Harmonic Oscillator

To illustrate the beauty and efficiency in describing the dynamics of a quantum system using the dirac notation and operator algebra, we reconsider the one-dimensional harmonic oscillator discussed in section 4.4.2 and described by the Hamiltonian operator

$$\mathbf{H} = \frac{\mathbf{p}^2}{2m} + \frac{1}{2}\,K\,\mathbf{x}^2, \tag{5.127}$$

with

$$[\mathbf{x}, \mathbf{p}] = \mathrm{j}\hbar. \tag{5.128}$$

5.8.1 Energy Eigenstates, Creation and Annihilation Operators

It is advantageous to introduce the following normalized position and momentum operators

$$\begin{aligned} \mathbf{X} &= \sqrt{\frac{K}{\hbar\omega_0}}\,\mathbf{x} & (5.129)\\ \mathbf{P} &= \sqrt{m\hbar\omega_0}\,\mathbf{p} & (5.130)\end{aligned}$$

with $\omega_0 = \sqrt{\frac{K}{m}}$. The Hamiltonian operator and the commutation relationship of the normalized position and momentum operator resume the simpler forms

$$\begin{aligned} \mathbf{H} &= \frac{\hbar\omega_0}{2}\left(\mathbf{P}^2 + \mathbf{X}^2\right), && (5.131)\\ [\mathbf{X}, \mathbf{P}] &= \mathrm{j}\,. && (5.132) \end{aligned}$$

Algebraically, it is very useful to introduce the nonhermitian operators

$$\begin{aligned} \mathbf{a} &= \frac{1}{\sqrt{2}}\left(\mathbf{X} + \mathrm{j}\mathbf{P}\right), && (5.133)\\ \mathbf{a}^+ &= \frac{1}{\sqrt{2}}\left(\mathbf{X} - \mathrm{j}\mathbf{P}\right), && (5.134) \end{aligned}$$

which satisfy the commutation relation

$$\left[\mathbf{a}, \mathbf{a}^+\right] = 1. \tag{5.135}$$

We find

$$\begin{aligned} \mathbf{a}\mathbf{a}^+ &= \frac{1}{2}\left(\mathbf{X}^2 + \mathbf{P}^2\right) - \frac{\mathrm{j}}{2}[\mathbf{X}, \mathbf{P}] = \frac{1}{2}\left(\mathbf{X}^2 + \mathbf{P}^2 + 1\right), && (5.136)\\ \mathbf{a}^+\mathbf{a} &= \frac{1}{2}\left(\mathbf{X}^2 + \mathbf{P}^2\right) + \frac{\mathrm{j}}{2}[\mathbf{X}, \mathbf{P}] = \frac{1}{2}\left(\mathbf{X}^2 + \mathbf{P}^2 - 1\right), && (5.137) \end{aligned}$$

and the Hamiltonian operator can be rewritten in terms of the new operators $\mathbf{a}$ and $\mathbf{a}^+$ as

$$\begin{aligned} \mathbf{H} &= \frac{\hbar\omega_0}{2}\left(\mathbf{a}^+\mathbf{a} + \mathbf{a}\mathbf{a}^+\right) && (5.138)\\ &= \hbar\omega_0\left(\mathbf{a}^+\mathbf{a} + \frac{1}{2}\right). && (5.139) \end{aligned}$$

We introduce the operator

$$\mathbf{N} = \mathbf{a}^+\mathbf{a}, \tag{5.140}$$

which is a hermitian operator. Up to an additive constant $1/2$ and a scaling factor $\hbar\omega_0$ equal to the energy of one quantum of the harmonic oscillator it is equal to the Hamiltonian operator of the harmonic oscillator. Obviously, $\mathbf{N}$ is the number operator counting the number of energy quanta excited in a

harmonic oscillator. We assume that the number operator $\mathbf{N}$ has eigenvectors denoted by $|n\rangle$ and corresponding eigenvalues N_n

$$\mathbf{N}\ |n\rangle = \mathbf{a}^{+}\mathbf{a}\ |n\rangle = N_n\,|n\rangle\,. \tag{5.141}$$

We also assume that these eigenvectors are normalized and since $\mathbf{N}$ is hermitian they are also orthogonal to each other

$$\langle m\,|n\rangle\ = \delta_{mn}. \tag{5.142}$$

Multiplication of this equation with the operator $\mathbf{a}$ and use of the commutation relation (5.135) leads to

$$\begin{aligned}
\mathbf{a}\ \mathbf{a}^{+}\mathbf{a}\,|n\rangle &= N_n\mathbf{a}\,|n\rangle && (5.143)\\
\left(\mathbf{a}^{+}\mathbf{a}+\mathbf{1}\right)\mathbf{a}\,|n\rangle &= N_n\mathbf{a}\,|n\rangle && (5.144)\\
\mathbf{N}\ \mathbf{a}\,|n\rangle &= (N_n-1)\,\mathbf{a}\,|n\rangle && (5.145)
\end{aligned}$$

Eq.(5.143) indicates that if $|n\rangle$ is an eigenstate to the number operator $\mathbf{N}$ then the state $\mathbf{a}\ |n\rangle$ is a new eigenstate to $\mathbf{N}$ with eigenvalue N_n-1. Because of this property, the operator $\mathbf{a}$ is called a lowering operator or annihilation operator, since application of the annihilation operator to an eigenstate with N_n quanta leads to a new eigenstate that contains one less quantum

$$\mathbf{a}\,|n\rangle = \ C\,|n-1\rangle\,, \tag{5.146}$$

where C is a yet undetermined constant. This constant follows from the normalization of this state and being an eigenvector to the number operator.

$$\begin{aligned}
\langle n|\,\mathbf{a}^{+}\mathbf{a}\,|n\rangle &= |C|^2\,, && (5.147)\\
C &= \sqrt{n}. && (5.148)
\end{aligned}$$

Thus

$$\mathbf{a}\,|n\rangle = \ \sqrt{n}\,|n-1\rangle\,, \tag{5.149}$$

Clearly, if there is a state with $n=0$ application of the annihilation operator leads to the null-vector in this Hilbert space, i.e.

$$\mathbf{a}\,|0\rangle = \ 0, \tag{5.150}$$

and there is no other state with a lower number of quanta, i.e. $N_0=0$ and $N_n=n$. This is the ground state of the harmonic oscillator, the state with the lowest energy.

If $\mathbf{a}$ is an annihilation operator for energy quanta, $\mathbf{a}^+$ must be a creation operator for energy quanta, otherwise the state $|n\rangle$ would not fulfill the eigenvalue equation Eq.(5.141)

$$\begin{aligned} \mathbf{a}^+\mathbf{a}\,|n\rangle &= n\,|n\rangle & (5.151)\\ \mathbf{a}^+\sqrt{n}\,|n-1\rangle &= n\,|n\rangle & (5.152)\\ \mathbf{a}^+\,|n-1\rangle &= \sqrt{n}\,|n\rangle & (5.153)\end{aligned}$$

or

$$\mathbf{a}^+\,|n\rangle = \sqrt{n+1}\,|n+1\rangle\ . \tag{5.154}$$

Starting from the energy ground state of the harmonic oscillator $|0\rangle$ with energy $\hbar\omega_0/2$ we can generate the n-th energy eigenstate by n-fold application of the creation operator $\mathbf{a}^+$ and proper normalization

$$|n\rangle = \frac{1}{\sqrt{(n+1)!}}\ \left(\mathbf{a}^+\right)^n|0\rangle\,, \tag{5.155}$$

with

$$\mathbf{H}\,|n\rangle = E_n\,|n\rangle\,, \tag{5.156}$$

and

$$E_n\ = \hbar\omega_0\left(n+\frac{1}{2}\right). \tag{5.157}$$

5.8.2 Matrix Representation

We can express the normalized position and momentum operators as functions of the creation and annihilation operators

$$\begin{aligned} \mathbf{X} &= \frac{1}{\sqrt{2}}\left(\mathbf{a}^+ + \mathbf{a}\right), & (5.158)\\ \mathbf{P} &= \frac{\mathrm{j}}{\sqrt{2}}\left(\mathbf{a}^+ - \mathbf{a}\right). & (5.159)\end{aligned}$$

These operators do have the following matrix representations

$$\begin{aligned} \langle m|\,\mathbf{a}\,|n\rangle &= \sqrt{n}\delta_{m,n-1}\,, \qquad \langle m|\ \mathbf{a}^+\ |n\rangle = \sqrt{n+1}\delta_{m,n+1}\,, & (5.160)\\ \langle m|\,\mathbf{a}^+\mathbf{a}\,|n\rangle &= n\delta_{m,n}\,, \qquad \langle m|\ \mathbf{a}\mathbf{a}^+\ |n\rangle = (n+1)\,\delta_{m,n}\,, & (5.161)\\ \langle m|\,\mathbf{X}\,|n\rangle &= \frac{1}{\sqrt{2}}\left(\sqrt{n+1}\delta_{m,n+1}\ +\sqrt{n}\delta_{m,n-1}\right), & (5.162)\\ \langle m|\,\mathbf{P}\,|n\rangle &= \frac{\mathrm{j}}{\sqrt{2}}\left(\sqrt{n+1}\delta_{m,n+1}\ -\sqrt{n}\delta_{m,n-1}\right), & (5.163)\end{aligned}$$

$$\langle m|\,\mathbf{a}^2\,|n\rangle = \sqrt{n(n-1)}\delta_{m,n-2}\,, \tag{5.164}$$

$$\langle m|\;\mathbf{a}^{+2}\,|n\rangle = \sqrt{(n+1)\,(n+2)}\delta_{m,n+2}\,, \tag{5.165}$$

$$\langle m|\,\mathbf{X}^2\,|n\rangle = \frac{1}{2}\left(\begin{array}{c}(2n+1)\delta_{m,n}+\sqrt{n(n-1)}\delta_{m,n-2}\\ +\sqrt{(n+1)\,(n+2)}\delta_{m,n+2}\end{array}\right), \tag{5.166}$$

$$\langle m|\,\mathbf{P}^2\,|n\rangle = \frac{1}{2}\left(\begin{array}{c}(2n+1)\delta_{m,n}-\sqrt{n(n-1)}\delta_{m,n-2}\\ -\sqrt{(n+1)\,(n+2)}\delta_{m,n+2}\end{array}\right). \tag{5.167}$$

5.8.3 Minimum Uncertainty States or Coherent States

From the matrix elements calculated in the last section, we find that the energy or quantum number eigenstates $|n\rangle$ have vanishing expected values for position and momentum. This also follows from the x-representation $\psi_n(x) = \langle x\,|n\rangle$ studied in section 4.4.2

$$\langle n|\,\mathbf{X}\,|n\rangle = 0\,, \qquad \langle n|\,\mathbf{P}\,|n\rangle = 0\,, \tag{5.168}$$

and the fluctuations in position and momentum are then simply

$$\langle n|\,\mathbf{X}^2\,|n\rangle = n+\frac{1}{2}\;, \qquad \langle n|\,\mathbf{P}^2\,|n\rangle = n+\frac{1}{2}\,. \tag{5.169}$$

The minimum uncertainty product for the fluctuations

$$\Delta X = \sqrt{\langle n|\,\mathbf{X}^2\,|n\rangle - \langle n|\,\mathbf{X}\,|n\rangle^2} = n+\frac{1}{2}\;, \tag{5.170}$$

$$\Delta P = \sqrt{\langle n|\,\mathbf{P}^2\,|n\rangle - \langle n|\,\mathbf{P}\,|n\rangle^2} = n+\frac{1}{2}. \tag{5.171}$$

is then

$$\Delta X\cdot\Delta P = n+\frac{1}{2}\;. \tag{5.172}$$

Only the ground state $n=0$ is a minimum uncertainty wave packet, since it satisfies the eigenvalue equation

$$\mathbf{a}\,|0\rangle = \;0, \tag{5.173}$$

where

$$\mathbf{a} = \frac{1}{\sqrt{2}}\,(\mathbf{X}+\mathrm{j}\mathbf{P})\,, \tag{5.174}$$

see problem set 8. In fact we can show that every eigenstate to the annihilation operator

$$\mathbf{a}\left|\alpha\right\rangle = \alpha\left|\alpha\right\rangle, \text{ for } \alpha\epsilon\mathbb{C} \tag{5.175}$$

is a minimum uncertainty state. We obtain for expected values of position or momentum in these states

$$\begin{aligned}
\left\langle\alpha\right|\mathbf{a}\left|\alpha\right\rangle &= \alpha, \qquad \left\langle\alpha\right|\mathbf{a}^{+}\left|\alpha\right\rangle = \alpha^{*}, && (5.176)\\
\left\langle\alpha\right|\mathbf{a}^{+}\mathbf{a}\left|\alpha\right\rangle &= \left|\alpha\right|^{2}, \qquad \left\langle\alpha\right|\mathbf{a}\mathbf{a}^{+}\left|\alpha\right\rangle = \left(\left|\alpha\right|^{2}+1\right), && (5.177)\\
\left\langle\alpha\right|\mathbf{X}\left|\alpha\right\rangle &= \frac{1}{\sqrt{2}}\left(\alpha+\alpha^{*}\right), \left\langle\alpha\right|\mathbf{P}\left|\alpha\right\rangle = \frac{\mathrm{j}}{\sqrt{2}}\left(\alpha-\alpha^{*}\right), && (5.178)
\end{aligned}$$

and for its squares

$$\begin{aligned}
\left\langle\alpha\right|\mathbf{a}^{+}\mathbf{a}\left|\alpha\right\rangle &= \left|\alpha\right|^{2}, \qquad \left\langle\alpha\right|\mathbf{a}\mathbf{a}^{+}\left|\alpha\right\rangle = \left(\left|\alpha\right|^{2}+1\right), && (5.179)\\
\left\langle\alpha\right|\mathbf{a}^{2}\left|\alpha\right\rangle &= \alpha^{2}, \qquad \left\langle\alpha\right|\mathbf{a}^{+2}\left|\alpha\right\rangle = \alpha^{*2}, && (5.180)\\
\left\langle\alpha\right|\mathbf{X}^{2}\left|\alpha\right\rangle &= \frac{1}{2}\left(\alpha^{2}+2\alpha^{*}\alpha+\alpha^{*2}+1\right) = \left\langle\alpha\right|\mathbf{X}\left|\alpha\right\rangle^{2}+\frac{1}{2}, && (5.181)\\
\left\langle\alpha\right|\mathbf{P}^{2}\left|\alpha\right\rangle &= \frac{1}{2}\left(-\alpha^{2}+2\alpha^{*}\alpha-\alpha^{*2}+1\right) = \left\langle\alpha\right|\mathbf{P}\left|\alpha\right\rangle^{2}+\frac{1}{2}. && (5.182)
\end{aligned}$$

Thus the uncertainty product is at its minimum

$$\Delta X\cdot\Delta P = \frac{1}{2}\ \forall\ \alpha\epsilon\mathbb{C}. \tag{5.183}$$

In fact one can show that the statistics of a position or momentum measurement for a harmonic oscillator in this state follows a Gaussian satistics with the average and variance given by Eqs.(5.178), (5.181) and (5.182). This can be represented pictorially in a phase space diagram as shown in Figure 5.1

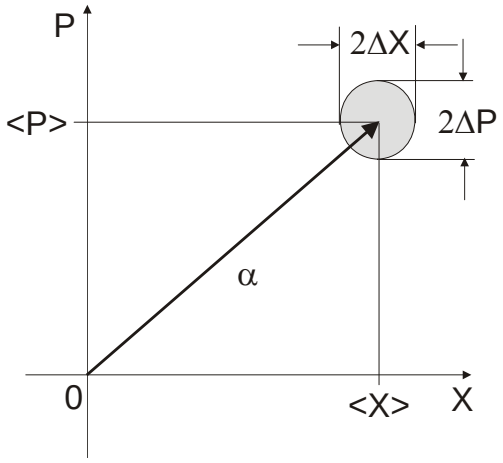

Figure 5.1: Representation of a minimum uncertainty state of the harmonic oscillator as a phase space distribution.

5.8.4 Heisenberg Picture

The Heisenberg equations of motion for a linear system like the harmonic oscillator are linear differential equations for the operators, which can be easily solved. From Eqs.(5.124) we find

$$\mathrm{j}\hbar\frac{\partial}{\partial t}\mathbf{a}_H(t) = [\mathbf{a}_H, \mathbf{H}] \tag{5.184}$$

$$= \hbar\omega_0\mathbf{a}_H \ , \tag{5.185}$$

with the solution

$$\mathbf{a}_H(t) = e^{-\mathrm{j}\omega_0 t}\,\mathbf{a}_S \ . \tag{5.186}$$

Therefore, the expectation values for the creation, annihilation, position and momentum operators are identical to those of Eqs.(5.176) to (5.182); we only need to subsitute $\alpha \rightarrow \alpha e^{-\mathrm{j}\omega_0 t}$. We may again pictorially represent the time evolution of these states as a probability distribution in phase space, see Figure 5.2.

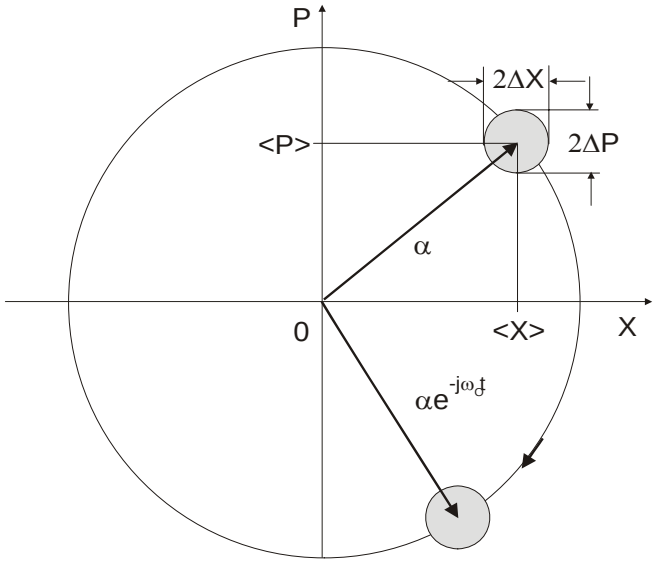

Figure 5.2: Time evolution of a coherent state in phase space.

5.9 The Kopenhagen Interpretation of Quantum Mechanics

5.9.1 Description of the State of a System

At a given time t the state of a system is described by a normalized vector $|\Psi(t)\rangle$ in the Hilbert space, H. The Hilbert space is a linear vector space. Therefore, any linear combination of vectors is again a possible state of the system. Thus superpositions of states are possible and with it come interferences.

5.9.2 Description of Physical Quantities

Measurable physical quantities, observables, are described by hermitian operators $\mathbf{A} = \mathbf{A}^+$.

5.9.3 The Measurement of Observables

An observable has a spectral representation in terms of eigenvectors and eigenvalues, which can be discrete or continuous, here we discuss the discrete case

$$\mathbf{A} = \sum_n A_n \left|A_n\right\rangle \ \left\langle A_n\right| \ , \tag{5.187}$$

The eigenvectors are orthogonal to each other and the eigenvalues are real

$$\left\langle A_n\right| \ \left. A_{n'}\right\rangle = \delta_{n,n'}. \tag{5.188}$$

Upon a measurement of the observable $\mathbf{A}$ of the system in state $\left|\Psi(t)\right\rangle$ the outcome can only be one of the eigenvalues A_n of the observable and the probability for that event to occur is

$$p_n = \left|\left\langle A_n\right| \ \left.\Psi(t)\right\rangle\right|^2 . \tag{5.189}$$

If the eigenvalue spectrum of the operator $\mathbf{A}$ is degenerate, the probabilities of the probabilities of the different states to the same eigenvector need to be added.

After the measurement the system is in the eigenstate $\left|A_n\right\rangle$ corresponding to the eigenvalue A_n found in the measurement, which is called the reduction of state[4]. This unphysical reduction of state is only necessary as a shortcut for the description of the measurement process and the fact that the system becomes entangled with the state of the macroscopic measurement equipment. This entanglement leads to a necessary decoherence of the superposition state of the measured system, which is equivalent to assuming a reduced state.

Bibliography

[1] Introduction to Quantum Mechanics, Griffiths, David J., Prentice Hall, 1995.

[2] Functional Analysis, G. Bachman and L. Naricci, Academic Press, 1966.

[3] Classical Mechanics, H. Goldstein, Addison and Wesley series in physics, 1959.

[4] Quantum Mechanics I, C. Cohen-Tannoudji, B. Diu, F. Laloe, John Wiley and Sons, Inc., 1978.

Chapter 6

Interaction of Light and Matter

Atomic or molecular gases in low concentration show sharp energy eigenspectra. This was shown for the hydrogen atom. Usually, there are infinitely many energy eigenstates in an atomic, molecular or solid-state medium and the spectral lines are associated with allowed transitions between two of these energy eigenstates. For many physical considerations it is already sufficient to take only two of these possible energy eigenstates into account, for example those which are related to the laser transition. The pumping of the laser can be later described by phenomenological relaxation processes into the upper laser level and out of the lower laser level. The resulting simple model is often called a two-level atom, which is mathematically also equivalent to a spin 1/2 particle in an external magnetic field, because the spin can only be parallel or anti-parallel to the field, i.e. it has two energy levels and energy eigenstates [4]. The interaction of the two-level atom with the electric field of an electromagnetic wave is described by the Bloch equations.

6.1 The Two-Level Model

An atom with only two energy eigenvalues is described by a two-dimensional state space spanned by the two energy eigenstates $|e\rangle$ and $|g\rangle$. The two states constitute a complete orthonormal system. The corresponding energy eigenvalues are E_e and E_g, see Fig. 6.1. In the position-, i.e. x-representation, these states correspond to the wave functions

$$\psi_g(x) = \langle x\,|g\rangle\ , \qquad \text{and} \quad \psi_e(x) = \langle x\,|e\rangle\ . \tag{6.1}$$

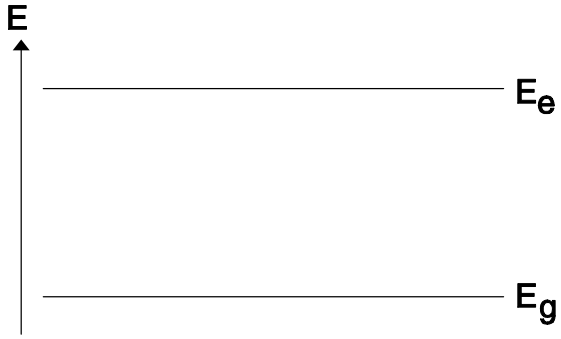

Figure 6.1: Two-level atom

The Hamiltonian operator of the two-level atom is in the energy representation

$$\mathbf{H}_A = E_e \left|e\right\rangle \left\langle e\right| + E_g \left|e\right\rangle \left\langle g\right| . \tag{6.2}$$

In this two-dimensional state space only $2 \times 2 = 4$ linearly independent linear operators are possible. A possible choice for an operator base in this space is

$$\begin{aligned} \mathbf{1} &= \left|e\right\rangle \left\langle e\right| + \left|g\right\rangle \left\langle g\right| , && (6.3)\\ \boldsymbol{\sigma}_z &= \left|e\right\rangle \left\langle e\right| - \left|g\right\rangle \left\langle g\right| , && (6.4)\\ \boldsymbol{\sigma}^+ &= \left|e\right\rangle \left\langle g\right| , && (6.5)\\ \boldsymbol{\sigma}^- &= \left|e\right\rangle \left\langle e\right| . && (6.6) \end{aligned}$$

The non-Hermitian operators $\boldsymbol{\sigma}^{\pm}$ could be replaced by the Hermitian operators $\boldsymbol{\sigma}_{x,y}$

$$\begin{aligned} \boldsymbol{\sigma}_x &= \boldsymbol{\sigma}^+ + \boldsymbol{\sigma}^-, && (6.7)\\ \boldsymbol{\sigma}_y &= -\mathrm{j}\boldsymbol{\sigma}^+ + \mathrm{j}\boldsymbol{\sigma}^- . && (6.8) \end{aligned}$$

The physical meaning of these operators becomes obvious, if we look at the action when applied to an arbitrary state

$$\left|\psi\right\rangle = c_g \left|g\right\rangle + c_e \left|e\right\rangle . \tag{6.9}$$

We obtain

$$\begin{aligned} \boldsymbol{\sigma}^+ \left|\psi\right\rangle &= c_g \left|e\right\rangle , && (6.10)\\ \boldsymbol{\sigma}^- \left|\psi\right\rangle &= c_e \left|g\right\rangle , && (6.11)\\ \boldsymbol{\sigma}_z \left|\psi\right\rangle &= c_e \left|e\right\rangle - c_g \left|g\right\rangle . && (6.12) \end{aligned}$$

The operator $\boldsymbol{\sigma}^+$ generates a transition from the ground to the excited state, and $\boldsymbol{\sigma}^-$ does the opposite. In contrast to $\boldsymbol{\sigma}^+$ and $\boldsymbol{\sigma}^-$, $\boldsymbol{\sigma}_z$ is a Hermitian operator, and its expectation value is an observable physical quantity with expectation value

$$\langle\psi|\,\boldsymbol{\sigma}_z\,|\psi\rangle = |c_e|^2 - |c_g|^2 = w, \tag{6.13}$$

the inversion w of the atom, since $|c_e|^2$ and $|c_g|^2$ are the probabilities for finding the atom in state $|e\rangle$ or $|g\rangle$ upon a corresponding measurement. If we consider an ensemble of N atoms the total inversion would be $W = N\,\langle\psi|\,\boldsymbol{\sigma}_z\,|\psi\rangle$. If we separate from the Hamiltonian (6.1) the term $(E_e + E_g)/2\cdot\mathbf{1}$, where $\mathbf{1}$ denotes the unity matrix, we rescale the energy values correspondingly and obtain for the Hamiltonian of the two-level system

$$\mathbf{H}_A = \frac{1}{2}\hbar\omega_{eg}\boldsymbol{\sigma}_z, \tag{6.14}$$

with the transition frequency

$$\omega_{eg} = \frac{1}{\hbar}(E_e - E_g). \tag{6.15}$$

This form of the Hamiltonian is favorable. There are the following commutator relations between operators (6.4) to (6.6)

$$\begin{aligned}
[\boldsymbol{\sigma}^+, \boldsymbol{\sigma}^-] &= \boldsymbol{\sigma}_z, && (6.16)\\
[\boldsymbol{\sigma}^+, \boldsymbol{\sigma}_z] &= -2\boldsymbol{\sigma}^+, && (6.17)\\
[\boldsymbol{\sigma}^-, \boldsymbol{\sigma}_z] &= 2\boldsymbol{\sigma}^-, && (6.18)
\end{aligned}$$

and anti-commutator relations, respectively

$$\begin{aligned}
[\boldsymbol{\sigma}^+, \boldsymbol{\sigma}^-]_+ &= \mathbf{1}, && (6.19)\\
[\boldsymbol{\sigma}^+, \boldsymbol{\sigma}_z]_+ &= \mathbf{0}, && (6.20)\\
[\boldsymbol{\sigma}^-, \boldsymbol{\sigma}_z]_+ &= \mathbf{0}, && (6.21)\\
[\boldsymbol{\sigma}^-, \boldsymbol{\sigma}^-]_+ &= [\boldsymbol{\sigma}^+, \boldsymbol{\sigma}^+]_+ = \mathbf{0}. && (6.22)
\end{aligned}$$

The operators $\boldsymbol{\sigma}_x$, $\boldsymbol{\sigma}_y$, $\boldsymbol{\sigma}_z$ fulfill the angular momentum commutator relations

$$\begin{aligned}
[\boldsymbol{\sigma}_x, \boldsymbol{\sigma}_y] &= 2\mathrm{j}\boldsymbol{\sigma}_z, && (6.23)\\
[\boldsymbol{\sigma}_y, \boldsymbol{\sigma}_z] &= 2\mathrm{j}\boldsymbol{\sigma}_x, && (6.24)\\
[\boldsymbol{\sigma}_z, \boldsymbol{\sigma}_x] &= 2\mathrm{j}\boldsymbol{\sigma}_y. && (6.25)
\end{aligned}$$

The two-dimensional state space can be represented as vectors in $\mathbb{C}^2$ according to the rule:

$$|\psi\rangle = c_g\,|g\rangle + c_e\,|e\rangle \quad \rightarrow \quad \begin{pmatrix} c_e \\ c_g \end{pmatrix}. \tag{6.26}$$

The operators are then represented by matrices

$$\boldsymbol{\sigma}^+ \quad \rightarrow \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{6.27}$$

$$\boldsymbol{\sigma}^- \quad \rightarrow \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{6.28}$$

$$\boldsymbol{\sigma}_z \quad \rightarrow \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{6.29}$$

$$\mathbf{1} \quad \rightarrow \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.30}$$

6.2 The Atom-Field Interaction In Dipole Approximation

The dipole moment of an atom $\vec{\mathbf{d}}$ is determined by the position operator $\vec{\mathbf{x}}$ via

$$\vec{\mathbf{d}} = -e_0\vec{\mathbf{x}}. \tag{6.31}$$

Then the expectation value for the dipole moment of an atom in state (6.9) is

$$\begin{aligned}\langle\psi|\,\vec{\mathbf{d}}\,|\psi\rangle = -e_0(&|c_e|^2\,\langle e|\,\vec{\mathbf{x}}\,|e\rangle + c_e c_g^*\,\langle g|\,\vec{\mathbf{x}}\,|e\rangle \\ &+c_g c_e^*\,\langle e|\,\vec{\mathbf{x}}\,|g\rangle + |c_g|^2\,\langle g|\,\vec{\mathbf{x}}\,|g\rangle).\end{aligned} \tag{6.32}$$

For simplicity, we may assume that the medium is an atomic gas. The atoms posses inversion symmetry, therefore, energy eigenstates must be symmetric or anti-symmetric, i.e. $\langle e|\,\vec{\mathbf{x}}\,|e\rangle = \langle g|\,\vec{\mathbf{x}}\,|g\rangle = 0$, see problem set 8. We obtain

$$\langle\psi|\,\vec{\mathbf{d}}\,|\psi\rangle = -e_0\,(c_e c_g^*\,\langle g|\,\vec{\mathbf{x}}\,|e\rangle + c_g c_e^*\,\langle g|\,\vec{\mathbf{x}}\,|e\rangle^*). \tag{6.33}$$

Note, this means there is no permanent dipole moment in an atom, which is in an energy eigenstate. This might not be the case in a solid. The

atoms consituting the solid are oriented in a lattice, which may break the symmetry. If so, there are permanent dipole moments and consequently the matrix elements $\langle e|\,\vec{\mathbf{x}}\,|e\rangle$ and $\langle g|\,\vec{\mathbf{x}}\,|g\rangle$ would not vanish.

An atom does only exhibit a dipole moment, if the product $c_e c_g^* \neq 0$, i.e. the state of the atom is in a superposition of states $|e\rangle$ and $|g\rangle$. With the dipole matrix elements

$$\vec{M} = e_0 \langle g|\,\vec{\mathbf{x}}\,|e\rangle \tag{6.34}$$

the expectation value for the dipole moment can be written as

$$\langle\psi|\,\vec{\mathbf{d}}\,|\psi\rangle = -(c_e c_g^* \vec{M} + c_g c_e^* \vec{M}^*) = -\langle\psi|\,(\boldsymbol{\sigma}^- \vec{M}^* + \boldsymbol{\sigma}^+ \vec{M})\,|\psi\rangle\,. \tag{6.35}$$

Since this is true for an arbitrary state, the dipole operator (6.31) is represented by

$$\vec{\mathbf{d}} = -\left(\boldsymbol{\sigma}^- \vec{M}^* + \boldsymbol{\sigma}^+ \vec{M}\right). \tag{6.36}$$

The energy of an electric dipole in an electric field is

$$\mathbf{H}_{A-F} = -\vec{\mathbf{d}} \cdot \vec{E}(\vec{x}_A, t). \tag{6.37}$$

We assume that the electric field is due to a monochromatic electromagntic wave. Then the electric field at the position of the atom, $\vec{x}_A$, can be written as

$$\vec{E}(\vec{x}_A, t) = \frac{1}{2}\left(\underline{E}_0 e^{\mathrm{j}\omega t}\ \vec{e}_p + \underline{E}_0^* e^{-\mathrm{j}\omega t}\ \vec{e}_p^*\right), \tag{6.38}$$

where $\underline{E}_0$ denotes the complex electric field amplitude at the position of the atom and $\vec{e}_p$ is the polarization vector of the wave. As we will see shortly, when there is a strong interaction of the wave with the atomic levels, the frequency of the electromagnetic wave is close to the atomic transistion frequency $\omega \approx \omega_{eg}$. The atom-field interaction Hamiltonian operator is then

$$\mathbf{H}_{A-F} = -\vec{\mathbf{d}} \cdot \vec{E}(\vec{x}_A, t) = \left(\boldsymbol{\sigma}^- \vec{M}^* + \boldsymbol{\sigma}^+ \vec{M}\right) \frac{1}{2}\left(\underline{E}_0 e^{\mathrm{j}\omega t}\ \vec{e}_p + \underline{E}_0^* e^{-\mathrm{j}\omega t}\ \vec{e}_p^*\right) \tag{6.39}$$

In the Rotating-Wave Approximation (RWA)[3], we only keep the slowly varying components in the interaction Hamiltonian. If there is no field, the operator $\boldsymbol{\sigma}^+$ evolves in the Heisenberg picture of the atom according to $\boldsymbol{\sigma}^+(t) = \boldsymbol{\sigma}^+(0) e^{\mathrm{j}\omega_{eg} t}$, thus terms proportional to the products $\boldsymbol{\sigma}^+ e^{\mathrm{j}\omega t}$ rotate at twice the optical frequency and will be neglected in the following

$$\mathbf{H}_{A-F} \approx \mathbf{H}_{A-F}^{RWA} = \frac{1}{2}\left(\vec{M} \cdot \vec{e}_p^*\right) \underline{E}_0^* e^{-\mathrm{j}\omega t} \boldsymbol{\sigma}^+ \ + h.c.. \tag{6.40}$$

The Schrödinger Equation for a two-level atom in a classical field is then

$$\begin{aligned} \mathrm{j}\hbar\frac{d}{dt}\left|\psi\right\rangle &= \left(\mathbf{H}_A + \mathbf{H}_{A-F}\right)\left|\psi\right\rangle \\ &\approx \left(\mathbf{H}_A + \mathbf{H}_{A-F}^{RWA}\right)\left|\psi\right\rangle . \end{aligned} \tag{6.41}$$

Written in the energy representation, we obtain

$$\frac{d}{dt}c_e = -\mathrm{j}\frac{\omega_{eg}}{2}c_e - \mathrm{j}\Omega_r e^{-\mathrm{j}\omega t}c_g, \tag{6.42}$$

$$\frac{d}{dt}c_g = +\mathrm{j}\frac{\omega_{eg}}{2}c_g - \mathrm{j}\Omega_r^* e^{+\mathrm{j}\omega t}c_e, \tag{6.43}$$

with the Rabi-frequency defined as

$$\Omega_r = \frac{\vec{M}\cdot\vec{e}_p^*}{2\hbar}\underline{E}_0^*. \tag{6.44}$$

For the time being, we assume that the the Rabi-frequency is real. If this is not the case, a transformation including a phase shift in the amplitudes $c_{\in,g}$ would be necessary to eliminate this phase. As expected the field couples the energy eigenstates.

6.3 Rabi-Oscillations

If the incident light has a constant field amplitude, $\underline{E}_0$, Eqs. (6.42) and (6.43) can be solved and we observe an oscillation in the population difference, the Rabi-oscillation [1]. To show this we introduce the detuning between field and atomic resonance

$$\Delta = \frac{\omega_{eg} - \omega}{2} \tag{6.45}$$

and the new probability amplitudes

$$C_e = c_e e^{\mathrm{j}\frac{\omega}{2}t}, \tag{6.46}$$

$$C_g = c_g e^{-\mathrm{j}\frac{\omega}{2}t}. \tag{6.47}$$

This leads to the new system of equations with constant coefficients

$$\frac{d}{dt}C_e = -\mathrm{j}\Delta C_e - \mathrm{j}\Omega_r C_g, \tag{6.48}$$

$$\frac{d}{dt}C_g = +\mathrm{j}\Delta C_g - \mathrm{j}\Omega_r C_e. \tag{6.49}$$

Note, these equations are identical to the coupled mode equations between two waveguide modes as studied in section 2.7.4. But now the coupling is between modes in time, i.e. resonances. The modes are electronic ones instead of photonic modes. But otherwise what has been said in section 2.7.4 applies in the same way. For the case of vanishing detuning it is especially easy to eliminate one of the variables and we arrive at

$$\frac{d^2}{dt^2}C_e = -\Omega_r^2 C_e \tag{6.50}$$

$$\frac{d^2}{dt^2}C_g = -\Omega_r^2 C_g. \tag{6.51}$$

The solution to this set of equations are the oscillations we are looking for. If the atom is at time $t = 0$ in the ground-state, i.e. $C_g(0) = 1$ and $C_e(0) = 0$, respectively, we arrive at

$$C_g(t) = \cos(\Omega_r t) \tag{6.52}$$

$$C_e(t) = -\mathrm{j}\sin(\Omega_r t). \tag{6.53}$$

Then, the probabilities for finding the atom in the ground or excited state are

$$|c_b(t)|^2 = \cos^2(\Omega_r t) \tag{6.54}$$

$$|c_a(t)|^2 = \sin^2(\Omega_r t), \tag{6.55}$$

as shown in Fig. 6.2. For the expectation value of the dipole operator under the assumption of a real dipole matrix element $\vec{M} = \vec{M}^*$ we obtain

$$\langle\psi|\,\vec{\mathbf{d}}\,|\psi\rangle = -\vec{M}c_e c_g^* + c.c. \tag{6.56}$$

$$= -\vec{M}\sin(2\Omega_r t)\sin(\omega_{eg} t). \tag{6.57}$$

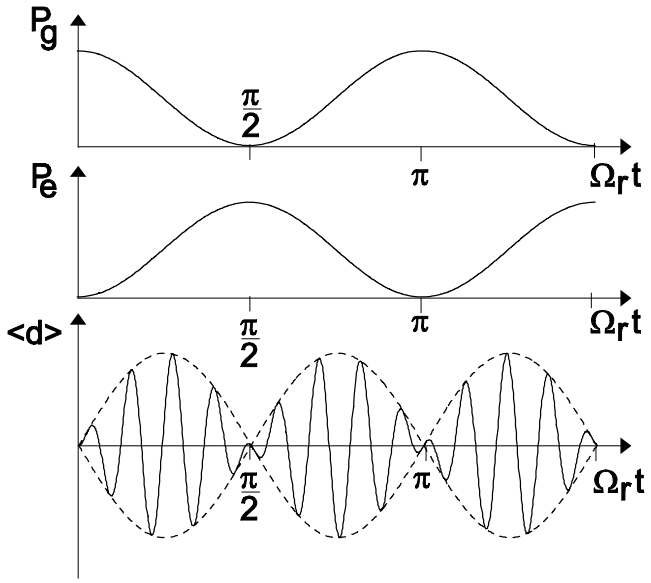

Figure 6.2: Evolution of occupation probabilities of ground and excited state and the average dipole moment of a two-level atom in resonant interaction with a coherent classical field.

The coherent external field drives the population of the atomic system between the two available states with a period $T_r = \pi/\Omega_r$. Applying the field only over half of this period leads to a complete inversion of the population. These Rabi-oscillations have been observed in various systems ranging from gases to semiconductors. Interestingly, the light emitted from the coherently driven two-level atom is not identical in frequency to the driving field. If we look at the Fourier spectrum of the polarization according to Eq.(6.57), we obtain lines at frequencies $\omega_\pm = \omega_{eg} \pm 2\Omega_r$. This is clearly a nonlinear output and the sidebands are called Mollow-sidebands [2] . Most important for the existence of these oscillations is the coherence of the atomic system over at least one Rabi-oscillation. If this coherence is destroyed fast enough, the Rabi-oscillations cannot happen and it is then impossible to generate inversion in a two-level system by interaction with light. This is the case for a large class of situations in light-matter interaction and especially for typical laser materials. So we are interested in finding out what happens in the case of loss of coherence in the atomic system due to additional interaction of the atoms with its environment.

6.4 The Density Operator

To study incoherent or dissipative processes it is necessary to switch to a statistical description. That is, we investigate not only the interaction of the atoms with the light field, via the Schroedinger Equation, leading to Rabi-oscillations but rather the interaction of an atomic ensemble with the light field. This is achieved by using the density operator instead of deterministic wave functions, similar to classical statistical mechanics, where the deterministic trajectories of particles are replaced by probability distributions.

The density operator of a pure state is defined by the dyadic product of the state with itself

$$\boldsymbol{\rho} = |\psi\rangle \langle\psi| \tag{6.58}$$

or in the energy representation by a $2 \times 2-$matrix

$$\boldsymbol{\rho} = \begin{pmatrix} \rho_{ee} & \rho_{eg} \\ \rho_{ge} & \rho_{gg} \end{pmatrix}. \tag{6.59}$$

. In the case of a pure state (6.9) this is

$$\boldsymbol{\rho} = \begin{pmatrix} c_e c_e^* & c_e c_g^* \\ c_g c_e^* & c_g c_g^* \end{pmatrix}. \tag{6.60}$$

For the rather simple case of a two-level system, each element of the density matrix corresponds to a physical quantity. The main diagonal contains the population probabilities for the levels; the off-diagonal element is the expectation value of the positive or negative frequency component of the dipole moment of the atom, i.e. its contribution to the polarization in the medium.

However, the concept of a density operator can be applied to any quantum mechanical system, not just the two-level atom. If an ensemble is described by a density operator, the expectation value of an arbitrary operator $\mathbf{A}$ can be computed using the trace formula

$$\langle \mathbf{A} \rangle = Tr\{\boldsymbol{\rho}\mathbf{A}\}. \tag{6.61}$$

The trace of an operator is defined as

$$Tr\{\mathbf{O}\} = \sum_n \langle n| \mathbf{O} |n\rangle . \tag{6.62}$$

where $|n\rangle$ can be any complete orthonormal base (ONB) in the Hilbert space. For example for the density matrix of the pure state (6.58) we find

$$\langle \mathbf{A} \rangle = Tr\{\boldsymbol{\rho}\mathbf{A}\} = \sum_n \langle n| \boldsymbol{\rho}\mathbf{A} |n\rangle \tag{6.63}$$

$$= \sum_n \langle n|\psi\rangle \langle\psi| \mathbf{A} |n\rangle = \langle\psi| \mathbf{A} \sum_n |n\rangle \langle n|\psi\rangle \tag{6.64}$$

$$= \langle\psi| \mathbf{A} |\psi\rangle . \tag{6.65}$$

The advantage of the density operator is that it can also be applied to a statistical mixture of pure states. For example, if the atom is in state $|e\rangle$ with probability p_e and in state $||g\rangle$ with probability p_g a density operator

$$\boldsymbol{\rho} = p_e |e\rangle \langle e| + p_g |g\rangle \langle g| \tag{6.66}$$

is defined, which can be used to compute the average values of observables in the proper statistical sense

$$\langle \mathbf{A} \rangle = Tr\{\boldsymbol{\rho}\mathbf{A}\} = p_e \langle e| \mathbf{A} |e\rangle + p_g \langle g| \mathbf{A} |g\rangle . \tag{6.67}$$

Since the matrices (6.27) to (6.30) build a complete base in the space of $2 \times 2-$matrices, we can express the density matrix as

$$\boldsymbol{\rho} = \rho_{ee}\frac{1}{2}(\mathbf{1} + \boldsymbol{\sigma}_z) + \rho_{gg}\frac{1}{2}(\mathbf{1} - \boldsymbol{\sigma}_z) + \rho_{eg}\boldsymbol{\sigma}^+ + \rho_{ge}\boldsymbol{\sigma}^- \tag{6.68}$$

$$= \frac{1}{2}\mathbf{1} + \frac{1}{2}(\rho_{ee} - \rho_{gg})\boldsymbol{\sigma}_z + \rho_{eg}\boldsymbol{\sigma}^+ + \rho_{ge}\boldsymbol{\sigma}^-, \tag{6.69}$$

since the trace of the density matrix is always one (normalization). Choosing the new base $\mathbf{1}, \boldsymbol{\sigma}_x, \boldsymbol{\sigma}_y, \boldsymbol{\sigma}_z$, we obtain

$$\boldsymbol{\rho} = \frac{1}{2}\mathbf{1} + \frac{1}{2}(\rho_{ee} - \rho_{gg})\boldsymbol{\sigma}_z + d_x\boldsymbol{\sigma}_x + d_y\boldsymbol{\sigma}_y, \tag{6.70}$$

with

$$d_x = \frac{1}{2}\left(\rho_{eg} + \rho_{ge}\right) = \Re\{\langle\boldsymbol{\sigma}^{(+)}\rangle\}, \tag{6.71}$$

$$d_y = \frac{\mathrm{j}}{2}\left(\rho_{eg} - \rho_{ge}\right) = \Im\{\langle\boldsymbol{\sigma}^{(+)}\rangle\}. \tag{6.72}$$

The expectation value of the dipole operator is given by (6.36)

$$\left\langle \vec{\mathbf{d}} \right\rangle = Tr\{\boldsymbol{\rho}\vec{\mathbf{d}}\} = -\vec{M}^* Tr\{\boldsymbol{\rho}\,\boldsymbol{\sigma}^+\} + c.c. = -\vec{M}^*\rho_{ge} + c.c. \tag{6.73}$$

From the Schrödinger equation for the wave function $|\psi>$ we can easily derive the equation of motion for the density operator called the von Neumann equation

$$\begin{aligned}\dot{\boldsymbol{\rho}} &= \frac{d}{dt}|\psi\rangle\langle\psi| + h.c. = \frac{1}{\mathrm{j}\hbar}\mathbf{H}|\psi\rangle\langle\psi| - \frac{1}{\mathrm{j}\hbar}|\psi\rangle\langle\psi|\mathbf{H} \qquad (6.74)\\ &= \frac{1}{\mathrm{j}\hbar}[\mathbf{H},\boldsymbol{\rho}].\end{aligned}$$

Due to the linear nature of this equation, this is also the correct equation for a density operator describing an arbitrary mixture of states. In case of a two-level atom, the von Neumann equation is

$$\dot{\boldsymbol{\rho}} = \frac{1}{\mathrm{j}\hbar}[\mathbf{H}_A,\boldsymbol{\rho}] = -\mathrm{j}\frac{\omega_{\in g}}{2}[\boldsymbol{\sigma}_z,\boldsymbol{\rho}]. \qquad (6.75)$$

Using the commutator relations (6.16) - (6.18), the result is

$$\begin{aligned}\dot{\rho}_{\in e} &= 0, &(6.76)\\ \dot{\rho}_{gg} &= 0, &(6.77)\\ \dot{\rho}_{eg} &= -\mathrm{j}\omega_{eg}\rho_{eg} \quad\rightarrow\quad \rho_{eg}(t) = e^{-\mathrm{j}\omega_{eg}t}\rho_{eg}(0), &(6.78)\\ \dot{\rho}_{ge} &= \mathrm{j}\omega_{eg}\rho_{ge} \quad\rightarrow\quad \rho_{ge}(t) = e^{\mathrm{j}\omega_{eg}t}\rho_{ge}(0). &(6.79)\end{aligned}$$

Again the isolated two-level atom has rather simple dynamics. The populations are constant. If there is a dipole moment induced at $t = 0$, i.e. the system is in a superposition state, then this dipole moment oscillates with the transition frequency $\omega_{\in g}$.

6.5 Energy- and Phase-Relaxation

In reality one has to work very hard to isolated an atom from its environment. Indeed in the case of laser active media, we are interested at radiating atoms, i.e. atoms that have a dipole interaction with the field. The coupling with the infinitely many modes of the free field leads already to spontaneous emission, an irreversible process. We could treat this process by using the Hamiltonian

$$\mathbf{H} = \mathbf{H}_A + \mathbf{H}_F + \mathbf{H}_{A-F}. \qquad (6.80)$$

Here, $\mathbf{H}_A$ is the Hamiltonian of the atom, $\mathbf{H}_F$ of the free field in thermal equilibrium at temperature T, and $\mathbf{H}_{A-F}$ describes the interaction between

them. A complete treatment along these lines would be straight forward using the techniques we learned so far, however it is beyond the scope of this class. The result of this calculation leads to the von Neumann equation of the reduced density matrix, i.e. the density matrix of the atom alone. In fact the result of such a calculation gives for the diagonal elements of the density operator, i.e. the state population probabilities, equations identical to those in section 3.3 involving Einstein's A and B coefficients. With the spontaneous emission rate $A = 1/\tau_{sp}$,i.e. the inverse spontaneous life time τ_{sp}, the populations change due to the induced and spontaneious emission processes and the absorption processes

$$\frac{d}{dt}|c_e(t)|^2 = \frac{d}{dt}\rho_{ee} = -\Gamma_e\rho_{ee} + \Gamma_a\rho_{gg} \tag{6.81}$$

with the abbreviations

$$\Gamma_e = \frac{1}{\tau_{sp}}(n_{th}+1), \tag{6.82}$$

$$\Gamma_a = \frac{1}{\tau_{sp}}n_{th}. \tag{6.83}$$

see Figure 6.3.

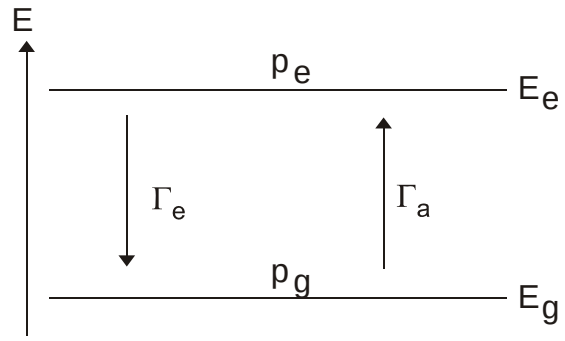

Figure 6.3: Two-level atom with transistion rates due to induced and spontaneous emission and absorption.

Here n_{th} is the number of thermally excited photons in the modes of the free field with frequency ω_{eg}, $n_{th} = 1/(\exp(\hbar\omega_{eg}/kT) - 1)$, at temperature T.

The total probability of being in the excited or the ground state has to be maintained, that is

$$\frac{d}{dt}\rho_{gg} = -\frac{d}{dt}\rho_{ee} = \Gamma_e\rho_{ee} - \Gamma_a\rho_{gg}. \tag{6.84}$$

If the populations decay, the polarization does as well, since $\rho_{ge} = c_e^* c_g$. It turns out that the polarization dynamics according to Eq.(6.78), besides the coherent oscillation, also aquires a decay process due to the finite lifetime of the excited state

$$\frac{d}{dt}\rho_{ge} = \mathrm{j}\omega_{eg}\rho_{eg} - \frac{\Gamma_e + \Gamma_a}{2}\rho_{ge}. \tag{6.85}$$

Thus the absorption as well as the emission processes are destructive to the phase. Therefore, the corresponding rates add up in the phase decay rate.

Taking the coherent (6.76)-(6.79) and incoherent processes (6.84-6.85) into account results in the following equations for the normalized average dipole moment $d = d_x + \mathrm{j}d_y$ and the inversion w

$$\dot{d} = \dot{\rho}_{ge} = (\mathrm{j}\omega_{eg} - \frac{1}{T_2})d, \tag{6.86}$$

$$\dot{w} = \dot{\rho}_{ee} - \dot{\rho}_{gg} = -\frac{w - w_0}{T_1}, \tag{6.87}$$

with the time constants

$$\frac{1}{T_1} = \frac{2}{T_2} = \Gamma_e + \Gamma_a = \frac{2n_{th}+1}{\tau_{sp}} \tag{6.88}$$

and the equilibrium inversion w_0, due to the thermal excitation of the atom by the thermal field

$$w_0 = \frac{\Gamma_a - \Gamma_e}{\Gamma_a + \Gamma_e} = \frac{-1}{1+2n_{th}} = -\tanh\left(\frac{\hbar\omega_{eg}}{2kT}\right). \tag{6.89}$$

The time constant T_1 denotes the energy relaxation in the two-level system and T_2 the phase relaxation. T_2 is the correlation time between amplitudes c_e and c_g. The coherence between the excited and the ground state described by ρ_{ge} is destroyed by the interaction of the two -level system with the environment.

In this basic model, the energy relaxation is half the phase relaxation rate or

$$T_2 = 2T_1. \tag{6.90}$$

The atoms in a laser medium do not only interact with the electromagnetic field, but also with phonons, i.e. acoustic vibrations of the host lattice in solid state laser material. Atoms might collide with each other in a gas laser and so on. All these processes must be considered when determining the energy and phase relaxation rates. Thus it might be not only radiative transistions that lead to a finite energy relaxation time T_1. Some of the processes are elastic, i.e. there is no energy relaxation but only the phase is influenced during the collision. Therefore, these processes reduce T_2 but have no influence on T_1. In real systems the phase relaxation time is most often much shorter than twice the energy relaxation time.

$$T_2 \leq 2T_1. \tag{6.91}$$

If the inversion deviates from its equilibrium value, w_0, it relaxes back into equilibrium with a time constant T_1. Eq. (6.89) shows that for all temperatures $T > 0$ the inversion is negative, i.e. the lower level is stronger populated than the upper level. Thus with incoherent thermal light, inversion in a two-level system cannot be achieved. Inversion can only be achieved by pumping with incoherent light, if there are more levels and subsequent relaxation processes into the upper laser level. Due to these relaxation processes the rate Γ_a deviates from the equilibrium expression (6.83), and it has to be replaced by the pump rate Λ. If the pump rate Λ exceeds Γ_e, the inversion corresponding to Eq. (6.89) becomes positive,

$$w_0 = \frac{\Lambda - \Gamma_e}{\Lambda + \Gamma_e}. \tag{6.92}$$

If we allow for artificial negative temperatures, we obtain with $T < 0$ for the ratio of relaxation rates

$$\frac{\Gamma_e}{\Gamma_a} = \frac{1 + \bar{n}}{\bar{n}} = e^{\frac{\hbar\omega_{eg}}{kT}} < 1. \tag{6.93}$$

Thus the pumping of the two-level system drives the system far away from thermal equilibrium. Now, we have a correct description of an ensemble of atoms in thermal equilibrium with its environment, which is a much more realistic description of media especially of typical laser media.

6.6 The Bloch Equations

If there is a coherent additional field in addition to the coupling to the environment, the Hamiltonian has to be extended by the dipole interaction with

that field,

$$\mathbf{H}_E = -\vec{\mathbf{d}}\cdot\vec{E}(\vec{x}_A, t). \tag{6.94}$$

Again we use the interaction Hamiltonian in RWA according to Eq.(6.40) for a time harmonic field Eq.(6.38) with polarization vector $\vec{e}_p$

$$\mathbf{H}_E = \frac{1}{2}\left(\vec{M}\cdot\vec{e}_p^*\right)\underline{E}_0^* e^{-\mathrm{j}\omega t}\boldsymbol{\sigma}^+ \; + h.c.. \tag{6.95}$$

In the von Neumann equation this leads to the additional term

$$\begin{aligned}
\dot{\boldsymbol{\rho}}|_E &= \frac{1}{\mathrm{j}\hbar}[\mathbf{H}_E, \boldsymbol{\rho}] && (6.96)\\
&= -\mathrm{j}\Omega_r e^{-\mathrm{j}\omega t}[\boldsymbol{\sigma}^+, \boldsymbol{\rho}] + \mathrm{j}\Omega_r^* e^{\mathrm{j}\omega t}[\boldsymbol{\sigma}^-, \boldsymbol{\rho}]. && (6.97)
\end{aligned}$$

With the density operator expressed as

$$\boldsymbol{\rho} = \frac{1}{2}\mathbf{1} + \frac{1}{2}(\rho_{ee} - \rho_{gg})\boldsymbol{\sigma}_z + \rho_{eg}\boldsymbol{\sigma}^+ + \rho_{ge}\boldsymbol{\sigma}^-, \tag{6.98}$$

and the commutation relations (6.16) - (6.18) we find

$$\begin{aligned}
\dot{\boldsymbol{\rho}}|_E &= \frac{1}{2}(\dot{\boldsymbol{\rho}}_{ee} - \dot{\boldsymbol{\rho}}_{gg})\boldsymbol{\sigma}_z + \dot{\boldsymbol{\rho}}_{eg}\boldsymbol{\sigma}^+ + \dot{\boldsymbol{\rho}}_{ge}\boldsymbol{\sigma}^- && (6.99)\\
&= -\mathrm{j}\Omega_r e^{-\mathrm{j}\omega t}\left\{\frac{1}{2}(\boldsymbol{\rho}_{ee} - \boldsymbol{\rho}_{gg})\left[\boldsymbol{\sigma}^+, \boldsymbol{\sigma}_z\right] + \boldsymbol{\rho}_{ge}\left[\boldsymbol{\sigma}^+, \boldsymbol{\sigma}^-\right]\right\} + \\
&\quad +\mathrm{j}\Omega_r^* e^{\mathrm{j}\omega t}[\left\{\frac{1}{2}(\boldsymbol{\rho}_{ee} - \boldsymbol{\rho}_{gg})\left[\boldsymbol{\sigma}^-, \boldsymbol{\sigma}_z\right] + \boldsymbol{\rho}_{eg}\left[\boldsymbol{\sigma}^-, \boldsymbol{\sigma}^+\right]\right\} \\
&= \mathrm{j}\Omega_r e^{-\mathrm{j}\omega t}\left\{(\boldsymbol{\rho}_{ee} - \boldsymbol{\rho}_{gg})\boldsymbol{\sigma}^+ + \boldsymbol{\rho}_{ge}\boldsymbol{\sigma}_z\right\} \\
&\quad +\mathrm{j}\Omega_r^* e^{\mathrm{j}\omega t}[\left\{(\boldsymbol{\rho}_{ee} - \boldsymbol{\rho}_{gg})\boldsymbol{\sigma}^- - \boldsymbol{\rho}_{eg}\boldsymbol{\sigma}_z\right\}
\end{aligned}$$

or expressed by the components of the density operator

$$\begin{aligned}
(\dot{\boldsymbol{\rho}}_{ee} - \dot{\boldsymbol{\rho}}_{gg})|_E &= 2\mathrm{j}\Omega_r e^{-\mathrm{j}\omega t}\rho_{ge} + c.c., && (6.100)\\
\dot{\rho}_{ge}|_E &= \mathrm{j}\Omega_r^* e^{\mathrm{j}\omega t}(\rho_{ee} - \rho_{gg}),. && (6.101)
\end{aligned}$$

The interaction with the external field leads to the following contributions in the dynamics of the dipole moment and the inversion

$$\begin{aligned}
\dot{d}|_E &= \dot{\rho}_{ge}|_E = \mathrm{j}\Omega_r^* e^{\mathrm{j}\omega t}\, w, && (6.102)\\
\dot{w}|_E &= \dot{\rho}_{ee}|_E - \dot{\rho}_{gg}|_E = 2\mathrm{j}\Omega_r e^{-\mathrm{j}\omega t} d + c.c. && (6.103)
\end{aligned}$$

Thus, the total dynamics of the two-level system including the pumping and dephasing processes from Eqs.(6.86) and (6.87) is given by

$$\dot{d} = -(\frac{1}{T_2} - \mathrm{j}\omega_{eg})d + \mathrm{j}\Omega_r^* e^{\mathrm{j}\omega t}\, w, \tag{6.104}$$

$$\dot{w} = -\frac{w - w_0}{T_1} + 2\mathrm{j}\Omega_r e^{-\mathrm{j}\omega t} d - 2\mathrm{j}\Omega_r^* e^{\mathrm{j}\omega t} d^*. \tag{6.105}$$

These equations are called the Bloch Equations. They describe the dynamics of a statistical ensemble of two-level atoms interacting with a classical electric field. Together with the Maxwell-Equations, where the polarization of the medium is related to the expectation value of the dipole moment of the atomic ensemble these result in the Maxwell-Bloch Equations.

6.7 Dielectric Susceptibility and Saturation

We have assumed that the external field is time harmonic

$$\vec{E}(\vec{x}_A, t) = \frac{1}{2}\left(\underline{E}_0 e^{\mathrm{j}\omega t}\ \vec{e}_p + \underline{E}_0^* e^{-\mathrm{j}\omega t}\ \vec{e}_p^*\right). \tag{6.106}$$

The Bloch Equations are nonlinear. However, for moderate field intensities, i.e. the magnitude of the Rabi-frequency is much smaller than the optical frequency, $|\Omega_r| << \omega$, the inversion does not change much within an optical cycle of the field. We assume that the inversion w of the atom will only be slowly changing and it adjusts itself to a steady state value w_s. If the inversion can be assumed time independent, $w = w_s$ the equation for the dipole moment is linear and the dipole moment will oscillate with the same frequency as the driving field

$$\underline{d} = \underline{d}_0 e^{\mathrm{j}\omega t}. \tag{6.107}$$

With the time harmonic solution (6.107) we find from Eqs. (6.104) and (6.105) for the dipole amplitude and the steady state inversion

$$\underline{d}_0 = \frac{\mathrm{j}}{2\hbar}\frac{\left(\vec{M}^* \cdot \vec{e}_p\right) w_s}{1/T_2 + \mathrm{j}(\omega - \omega_{eg})}\underline{E}_0 \tag{6.108}$$

$$w_s = \frac{w_0}{1 + \frac{T_1}{\hbar^2}\frac{1/T_2\,|\vec{M}^* \cdot \vec{e}_p|^2}{(1/T_2)^2 + (\omega_{eg} - \omega)^2}|\underline{E}_0|^2}. \tag{6.109}$$

We introduce the normalized lineshape function, which is in this case a Lorentzian**,**

$$L(\omega) = \frac{(1/T_2)^2}{(1/T_2)^2 + (\omega_{eg} - \omega)^2}, \tag{6.110}$$

and connect the square of the field $|\underline{E}_0|^2$ to the intensity I of a propagating plane wave, according to Eq. (2.38), $I = \frac{1}{2Z_F}|\underline{E}_0|^2$**,**

$$w_s = \frac{w_0}{1 + \frac{I}{I_s}L(\omega)}. \tag{6.111}$$

Thus the stationary inversion depends on the intensity of the incident light. Therefore, w_0 is called the unsaturated inversion, w_s the saturated inversion and I_s,with

$$I_s = \left[\frac{2T_1T_2Z_F}{\hbar^2}|\vec{M}^* \cdot \vec{e}_p|^2\right]^{-1}, \tag{6.112}$$

is the saturation intensity. The expectation value of the dipole operator (6.31) is then given by

$$\left\langle \tilde{\mathbf{d}} \right\rangle = -(\vec{M}^* \underline{d} + \vec{M}\underline{d}^*). \tag{6.113}$$

Multiplication with the number of atoms per unit volume, N, relates the dipole moment of the atom to the complex polarization $\underline{\vec{P}}_0$ of the medium, and therefore to the susceptibility according to

$$\underline{\vec{P}}_0 = -2N\vec{M}^*\underline{d}_0, \tag{6.114}$$

$$\underline{\vec{P}}_0 = \epsilon_0\chi(\omega)\vec{e}_p\underline{E}_0. \tag{6.115}$$

From the definitions (6.114), (6.115) and Eq. (6.108) we obtain for the linear susceptibility of the medium

$$\chi(\omega) = \vec{M}^*\vec{M}^T\frac{\mathrm{j}N}{\hbar\epsilon_0}\frac{w_s}{1/T_2 + \mathrm{j}(\omega - \omega_{eg})}, \tag{6.116}$$

which is a tensor. In the following we assume that the direction of the atom is random, i.e. the alignment of the atomic dipole moment, $\vec{M}$, and the electric field is random. Therefore, we have to average over the angle enclosed

between the electric field of the wave and the atomic dipole moment, which results in

$$\overline{\begin{pmatrix} M_xM_x & M_xM_y & M_xM_z \\ M_yM_x & M_yM_y & M_yM_z \\ M_zM_x & M_zM_y & M_zM_z \end{pmatrix}} = \begin{pmatrix} \overline{M_x^2} & 0 & 0 \\ 0 & \overline{M_y^2} & 0 \\ 0 & 0 & \overline{M_z^2} \end{pmatrix} = \frac{1}{3}|\vec{M}|^2\,\mathbf{1}. \quad (6.117)$$

Thus, for homogeneous and isotropic media the susceptibility tensor shrinks to a scalar

$$\chi(\omega) = \frac{1}{3}|\vec{M}|^2 \frac{\mathrm{j}N}{\hbar\epsilon_0} \frac{w_s}{1/T_2 + \mathrm{j}(\omega - \omega_{eg})}. \quad (6.118)$$

Real and imaginary part of the susceptibility

$$\chi(\omega) = \chi'(\omega) + \mathrm{j}\chi''(\omega) \quad (6.119)$$

are then given by

$$\chi'(\omega) = -\frac{|\vec{M}|^2 N w_s T_2^2(\omega_{eg} - \omega)}{3\hbar\epsilon_0} L(\omega), \quad (6.120)$$

$$\chi''(\omega) = \frac{|\vec{M}|^2 N w_s T_2}{3\hbar\epsilon_0} L(\omega). \quad (6.121)$$

If the incident radiation is weak, i.e.

$$\frac{I}{I_s} L(\omega)) \ll 1 \quad (6.122)$$

we obtain $w_s \approx w_0$. For optical transitions there is no thermal excitation of the excited state and $w_0 = -1$. For an inverted system, $w_0 > 0$, the real and imaginary parts of the susceptibility are shown in Fig. 6.4.

The shape of the susceptibility computed quantum mechanically compares well with the classical susceptibility (2.43) derived from the harmonic oscillator model close to the transistion frequency ω_{eg} for a transition with reasonably high $Q = T_2\omega_{eg}$. Note, the quantum mechanical susceptibility is identical to the complex Lorentzian introduced in Eq.(2.90). There is an appreciable deviation, however, far away from resonance. Far off resonance the rotating wave approximation should not be used.

The physical meaning of the real and imaginary part of the susceptibility is of course identical to section 2.1.8. The propagation constant k of a TEM-wave in such a medium is related to the susceptibility by

$$k = \omega\sqrt{\mu_0\epsilon_0(1 + \chi(\omega))} \approx k_0\left(1 + \frac{1}{2}\chi(\omega)\right), \quad \text{with} \quad k_0 = \omega\sqrt{\mu_0\epsilon_0} \quad (6.123)$$

for $|\chi| \ll 1$. Under this assumption we obtain

$$k = k_0(1 + \frac{\chi'}{2}) + \mathrm{j}k_0\frac{\chi''}{2}. \tag{6.124}$$

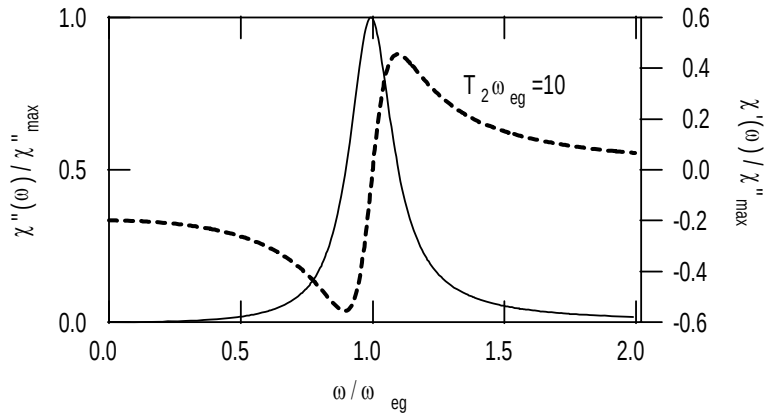

Figure 6.4: Real and imaginary part of the complex susceptibility for an inverted medium $w_s > 0$. The positive imaginary susceptibility indicates exponential growth of an electromagnetic wave propagating in the medium.

The real part of the susceptibility contributes to the refractive index $n = 1 + \chi'/2$. In the case of $\chi'' < 0$, the imaginary part leads to an exponential damping of the wave. For $\chi'' > 0$ amplification takes place. Amplification of the wave is possible for $w_0 > 0$, i.e. an inverted medium.

The phase relaxation rate $1/T_2$ of the dipole moment determines the width of the absorption line or the bandwidth of the amplifier. The amplification can not occur forever, because the amplifier saturates when the intensity reaches the saturation intensity. This is a strong deviation from the linear susceptibility we derived from the classical oscillator model. The reason for this saturation is two fold. First, the light can not extract more energy from the atoms then there is energy stored in them, i.e. energy conservation holds. Second the induced dipole moment in a two-level atom is limited by the maximum value of the matrix element. In contrast the induced dipole moment in a classical oscillator grows proportionally to the applied field without limits.

6.8 Rate Equations and Cross Sections

In many cases the fastest process in the atom-field interaction dynamics is the dephasing of the dipole moment, i. e. $T_2 \to 0$. For example, in semiconductors $T_2 < 50fs$. In those cases the magnitude of the dipole moment relaxes instantaneously into the steady state and follows the magnitude of the intensity envelope $I(t)$ of the electromagntic field, which evolves on a much slower time scale. From Eq.(6.104) we obtain with the steady state solutions for the dipole moment (6.107) and (6.108) for the time dependent inversion in the atomic system after adiabatic elimination of dipole moment

$$\dot{w} = -\frac{w(t) - w_0}{T_1} - \frac{w(t)}{T_1 I_s} L(\omega) I(t), \tag{6.125}$$

where I(t) is the intensity of the electromagntic wave interaction with the two-level atom. In this limit the Bloch Equations became simple rate equations. We only take care of the counting of population differences due to spontaneous and stimulated emissions.

The interaction of an atom with light at a given transition with the stream of photons on resonance, i.e. $\omega = \omega_{eg}$ is often discribed by the mass action law, that is that the number of induced transistions, for example from the excited to the ground state, is proportional to the product of the number of atoms in the excited state and the photon flux density $I_{ph} = I/\hbar\omega_{eg}$

$$\dot{w}|_{induced} = -\sigma w I_{ph} = -\frac{w}{T_1 I_s} I. \tag{6.126}$$

Thus the interaction cross section can be expressed in terms of the saturation intensity as

$$\sigma = \frac{\hbar\omega_{eg}}{T_1 I_s} \tag{6.127}$$

$$= \frac{2\omega_{eg} T_2 Z_F}{\hbar} |\vec{M}^* \cdot \vec{e}_p|^2. \tag{6.128}$$

In this chapter, we have introduced the most important spectroscopic quantities that characterize an atomic transisition, which are the lifetime of the excited state or often called upper-state lifetime or longitudinal lifetime T_1, the phase relaxation time or transverse relaxation time T_2 which is the inverse half-width at half maximum of the line and the interaction cross-section that only depends on the dipole matrix element and the linewidth of the transition.

Bibliography

[1] I. I. Rabi: "Space Quantization in a Gyrating Magnetic Field," Phys. Rev. **51**, 652-654 (1937).

[2] B. R. Mollow, "Power Spectrum of Light Scattered by Two-Level Systems," Phys. Rev **188**, 1969-1975 (1969).

[3] P. Meystre, M. Sargent III: Elements of Quantum Optics, Springer Verlag (1990).

[4] L. Allen and J. H. Eberly: Optical Resonance and Two-Level Atoms, Dover Verlag (1987).

Chapter 7

Lasers

After having derived the quantum mechanically correct suszeptibility for an inverted atomic system that can provide gain, we can use the two-level model to study the laser and its dynamics. After discussing the laser concept briefly we will investigate various types of gain media, gas, liquid and solid-state, that can be used to construct lasers and amplifiers. Then the dynamics of lasers, threshold behavior, steady state behavior and relaxation oscillations are discussed. A short introduction in the generation of high energy and ultrashort laser pulses using Q-switching and mode locking will be given at the end.

7.1 The Laser (Oscillator) Concept

Since the invention of the vacuum amplifier tube by Robert von Lieben and Lee de Forest in 1905/06 it was known how to amplify electromagnetic waves over a broad wavelength range and how to build oscillator with which such waves could be generated. This was extended into the millimeter wave region with advances in amplifier tubes and later solid-state devices such as transistors. Until the 1950's thermal radiation sources were mostly used to generate electromagnetic waves in the optical frequency range. The generation of coherent optical waves was only made possible by the Laser. The first amplifier based on discrete energy levels (quantum amplifier) was the MASER (Microwave Amplification by Stimulated Emission of Radiation), which was invented by Gordon, Townes and Zeiger 1954. In 1958 Schawlow and Townes proposed to extend the MASER principle to the optical regime.

The amplification should arise from stimulated emission between discrete energy levels that must be inverted, as discussed in the last section. Amplifiers and oscillators based on this principle are called LASER (Light Amplification by Stimulated Emission of Radiation). Maiman was the first to demonstrate a laser based on the solid-state laser material Ruby.

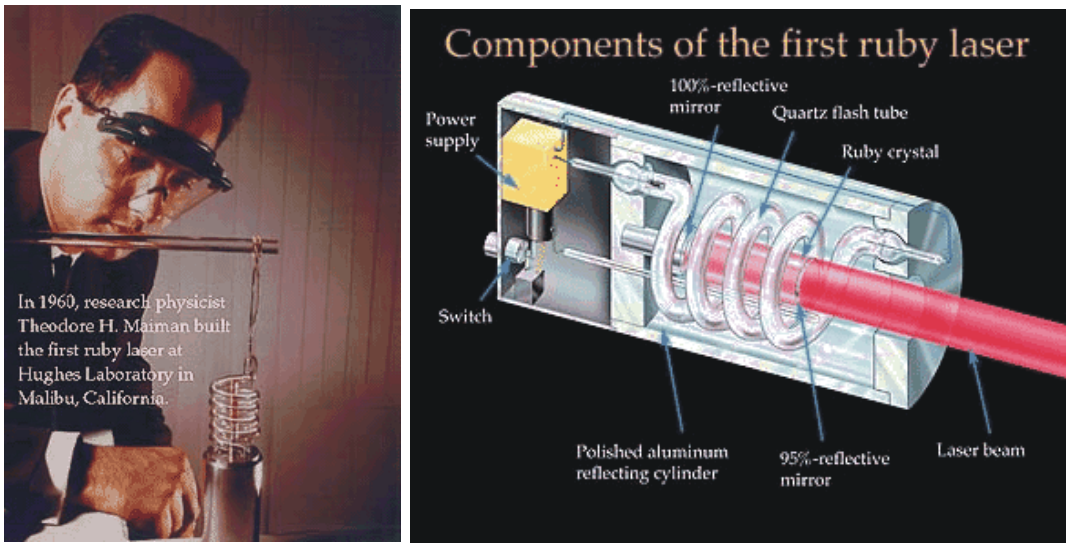

Figure 7.1: Theodore Maiman with the first Ruby Laser in 1960 and a cross sectional view of the first device [4].

The first HeNe-Laser, a gas laser followed in 1961. It is a gas laser built by Ali Javan at MIT, with a wavelength of 632.8 nm and a linewidth of only 10kHz.

The basic principle of an oscillator is a feedback circuit that is unstable, i.e. there is positive feedback at certain frequencies or certain frequency ranges, see Figure 7.2. It is the feedback circuit that determines the frequency of oscillation. Once the oscillation starts, the optical field will build up to an intensity approaching, or even surpassing, the saturation intensity of the amplifier medium by many times, until the amplifier gain is reduced to a value equal to the losses that the signal experiences after one roundtrip in the feedback loop, see Figure 7.3

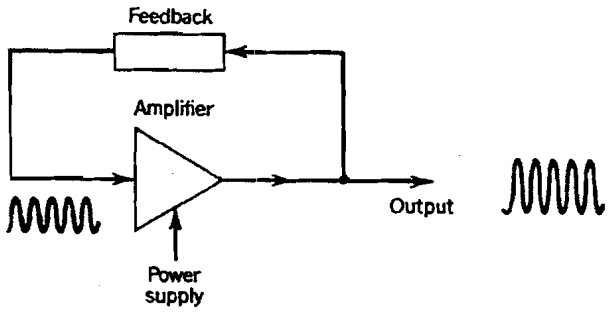

Figure 7.2: Principle of an oscillator circuit: an amplifier with positive feedback [6] p. 495.

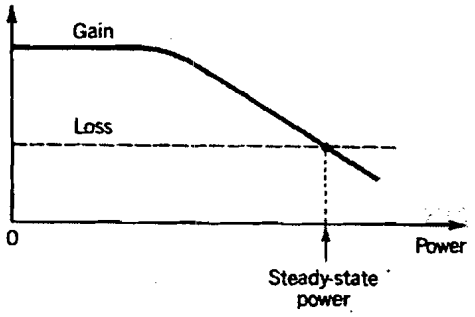

Figure 7.3: Saturation of amplification with increasing signal power leads to a stable oscillation [6], p. 496.

In the radio frequency range the feedback circuit can be an electronic feedback circuit. At optical frequencies we use an optical resonator, which is in most cases well modeled as a one-dimensional Fabry-Perot resonator, which we analysed in depth in section 7.4. We already found back then that the transfer characterisitcs of a Fabry-Perot resonator can be understood as a feedback structure. All we need to do to construct an oscillator is provide amplification in the feedback loop, i.e. to compensate in the resonator for eventual internal losses or the losses due to the output coupling via the mirrors of the Fabry-Perot, see Figure 7.4.We have already discussed in section

2.6.2 various optical resonators, which have Gaussian beams as the fundamental resonator modes. One can also use waveguides or fibers that have semitransparent mirrors at its ends or form rings as laser resonators. In the latter ones output coupling of radiation is achieved with waveguide or fiber couplers in the rings.

Today lasers generating light continuosly or in the form of long, nanosecond, or very short, femtosecond pulses can be built. Typically these lasers are Q-switched or mode-locked, respectively. The average power level can vary from microwatt to kilowatts.

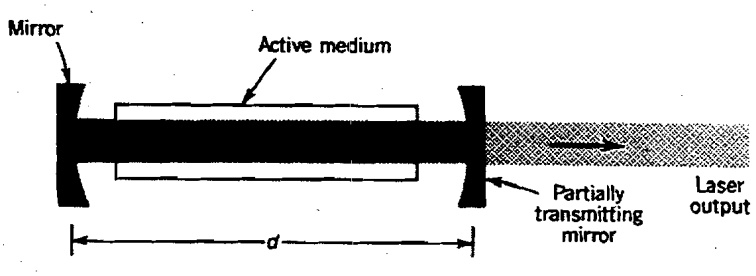

Figure 7.4: A laser consists of an optical resonator where the internal losses and/or the losses due to partially reflecting mirrors are compensated by a gain medium inside the resonator [6], p. 496.

7.2 Laser Gain Media

Important characteristics of laser gain media are whether it is a solid, a gase or liquid, how inversion can be achieved and what the spectroscopic paratmeters are, i.e. upperstate lifetime, $\tau_L = T_1$, linewdith $\Delta f_{FWHM} = \frac{2}{T_2}$ and the crosssection for stimulated emission.

7.2.1 Three and Four Level Laser Media

As we discussed before inversion can not be achieved in a two level system by optical pumping. The coherent regime is typically inaccesible by typcial optical pump sources. Inversion by optical pumping can only be achieved when using a three or four-level system, see Figures 7.5 and 7.6

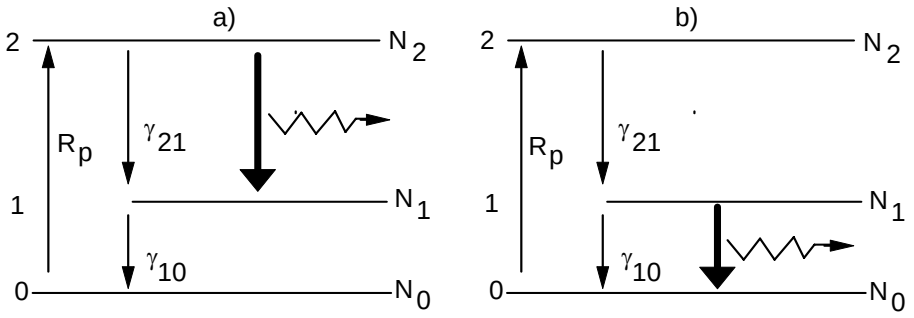

Figure 7.5: Three-level laser medium.

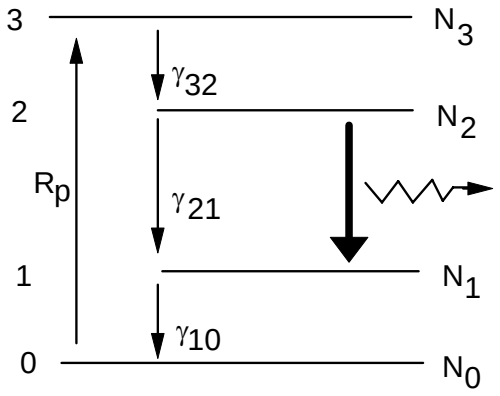

Figure 7.6: Four-level laser medium.

If the medium is in thermal equilibrium, typically only the ground state is occupied. By optical pumping with an intense lamp (flash lamp) or another laser one can pump a significant fraction of the atoms from the ground state with population N_0 into the excited state N_3 both for the three level laser operating according to scheme shown in figure 297 (a) or N_4 in the case

of the four level laser, see Figure 7.6. If the relaxation rate γ_{10} is very fast compared to γ_{21}, where the laser action should occur inversion can be achieved, i.e. $N_2 > N_1$. For the four level laser the relaxation rate γ_{32} should also be fast in comparison to γ_{21}. These systems are easy to analyze in the rate equation approximation, where the dipole moments are already adiabatically eliminated. For example, for the three level system in Figure 7.5 a). we obtain the rate equations of the three level system in analogy to the two-level system

$$\frac{d}{dt}N_2 = -\gamma_{21}N_2 - \sigma_{21}\left(N_2 - N_1\right)I_{ph} + R_p \tag{7.1}$$

$$\frac{d}{dt}N_1 = -\gamma_{10}N_1 + \gamma_{21}N_2 + \sigma_{21}\left(N_2 - N_1\right)I_{ph} \tag{7.2}$$

$$\frac{d}{dt}N_0 = \gamma_{10}N_1 - R_p \tag{7.3}$$

Here, σ_{21} is the cross section for stimulated emission between the levels 2 and 1 and I_{ph} is the photon flux at the transition frequency f_{21}.In most cases, there are any atoms available in the ground state such that optical pumping can never deplete the number of atoms in the ground state N_0. That is why we can assume a constant pump rate R_p. If the relaxation rate γ_{10} is much faster than γ_{21} and the number of possible stimulated emission events that can occur $\sigma_{21}\left(N_2 - N_1\right)I_{ph}$, then we can set $N_1 = 0$ and obtain only a rate equation for the upper laser level

$$\frac{d}{dt}N_2 = -\gamma_{21}\left(N_2 - \frac{R_p}{\gamma_{21}}\right) - \sigma_{21}N_2 \cdot I_{ph}. \tag{7.4}$$

This equation is identical to the equation for the inversion of the two-level system, see Eq.(6.125). Here, $\frac{R_p}{\gamma_{21}}$ is the equilibrium upper state population in the absence of photons, $\gamma_{21} = \frac{1}{\tau_L}$ is the inverse upper state lifetime due to radiative and non radiative processes.

Note, a similar analysis can be done for the three level laser operating according to the scheme shown in Figure 7.5 (b). Then the relaxation rate from level 3 to level 2, which is now the upper laser level has to be fast. But in addition the optical pumping must be so strong that essentially all the ground state levels are depleted. Undepleted groundstate populations would always lead to absorption of laser radiation.

In the following we want to discuss the electronic structure of a few often encountered laser media. A detail description of laser media can be found in [7].

7.3 Types of Lasers

7.3.1 Gas Lasers

Helium-Neon Laser

The HeNe-Laser is the most widely used noble gas laser. Lasing can be achieved at many wavelength 632.8nm (543.5nm, 593.9nm, 611.8nm, 1.1523μm, 1.52μm, 3.3913μm). Pumping is achieved by electrical discharge, see Figure 7.7.

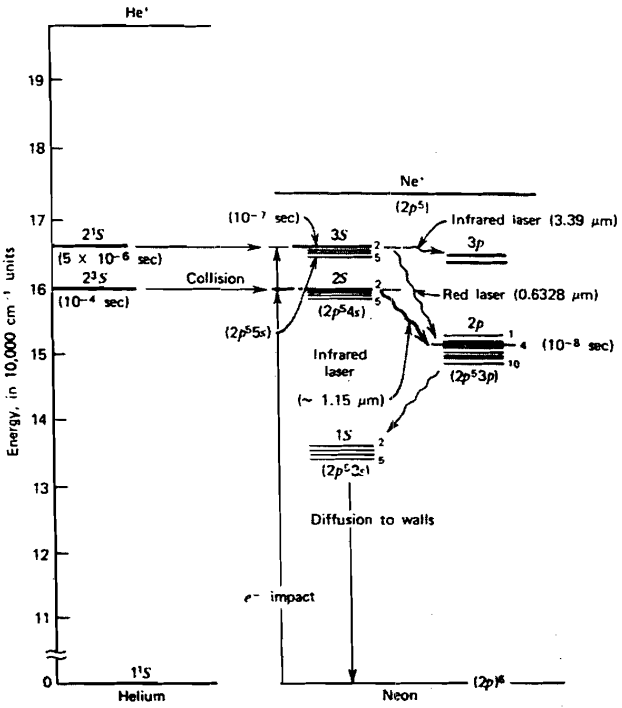

Figure 7.7: Energy level diagram of the transistions involved in the HeNe laser [9].

The helium is excited by electron impact. The energy is then transfered to Neon by collisions. The first HeNe laser operated at the 1.1523μm line [8]. HeNe lasers are used in many applications such as interferometry, holography, spectroscopy, barcode scanning, alignment and optical demonstrations.

Argon and Krypton Ion Lasers

Similar to the HeNe-laser the Argon ion gas laser is pumped by electric discharge and emitts light at wavelength: 488.0nm, 514.5nm, 351nm, 465.8nm, 472.7nm, 528.7nm.It is used in applications ranging from retinal phototherapy for diabetes, lithography, and pumping of other lasers.

The Krypton ion gas laser is analogous to the Argon gas laser with wavelength: 416nm, 530.9nm, 568.2nm, 647.1nm, 676.4nm, 752.5nm, 799.3nm. Pumped by electrical discharge. Applications range from scientific research. When mixed with argon it can be used as "white-light" lasers for light shows.

Carbon Lasers

In the carbon dioxide (CO_2) gas laser the laser transistions are related to vibrational-rotational excitations. CO_2 lasers are highly efficient approaching 30%. The main emission wavelengths are 10.6μm and 9.4μm. They are pumped by transverse (high power) or longitudinal (low power) electrical discharge. It is heavily used in the material processing industry for cutting, and welding of steel and in the medical area for surgery.

Carbon monoxide (CO) gas laser: Wavelength 2.6 - 4μm, 4.8 - 8.3μm pumped by electrical discharge. Also used in material processing such as engraving and welding and in photoacoustic spectroscopy. Output powers as high as 100kW have been demonstrated.

Excimer Lasers:

Chemical lasers emitting in the UV: 193nm (ArF), 248nm (KrF), 308nm (XeCl), 353nm (XeF) excimer (excited dimer). These are molecules that exist only if one of the atoms is electronically excited. Without excitation the two atoms repell each other. Thus the electronic groundstate is not stable and is therefore not populated, which is ideal for laser operation. These lasers are used for ultraviolet lithography in the semiconductor industry and laser surgery.

7.3.2 Dye Lasers:

The laser gain medium are organic dyes in solution of ethyl, methyl alcohol, glycerol or water. These dyes can be excited by optically with Argon lasers for example and emit at 390-435nm (stilbene), 460-515nm (coumarin 102),

570-640 nm (rhodamine 6G) and many others. These lasers have been widely used in research and spectroscopy because of there wide tuning ranges. Unfortunately, dyes are carcinogenic and as soon as tunable solid state laser media became available dye laser became extinct.

7.3.3 Solid-State Lasers

Ruby Laser

The first laser was indeed a solid-state laser: Ruby emitting at 694.3nm [5]. Ruby consists of the naturally formed crystal of aluminum oxide (Al_2O_3) called corundum. In that crystal some of Al^{3+} ions are replaced by Cr^{3+} ions. Its the chromium ions that give Ruby the pinkish color, i.e. its flourescence, which is related to the laser transisitons, see the level structure in Figure 7.8. Ruby is a three level laser.

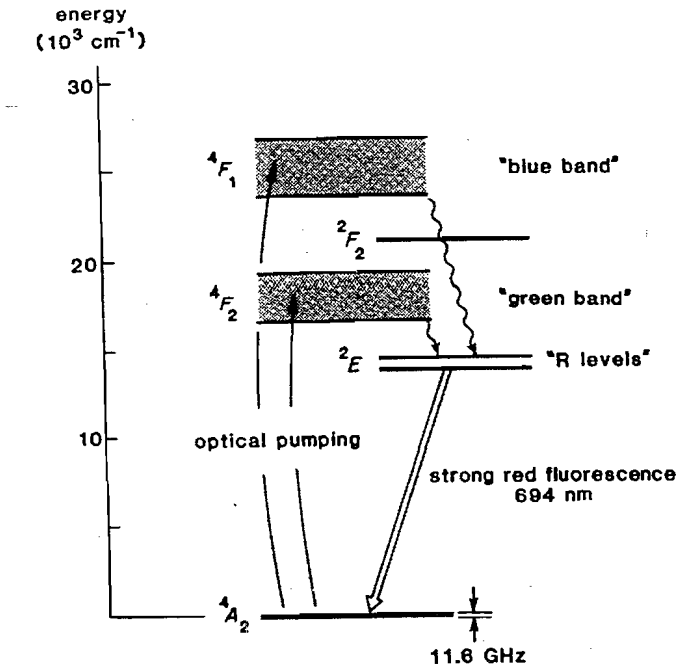

Figure 7.8: Energy level diagram for Ruby, [2], p. 13.

Today, for the manufacturing of ruby as a laser material, artificially grown

crystals from molten material which crystalizes in the form of sapphire is used. The liftetime of the upper laser level is 3ms. Pumping is usually achieved with flashlamps, see Figure 7.1.

Neodymium YAG (Nd:YAG)

Neodymium YAG consists of Yttrium-Aluminium-Garnet (YAG) $Y_3Al_5O_{12}$ in which some of the Y^{3+} ions are replaced by Nd^{3+} ions. Neodymium is a rare earth element, where the active electronic states are shielded inner $4f$ states. Nd:YAG is a four level laser, see Figure **??**.

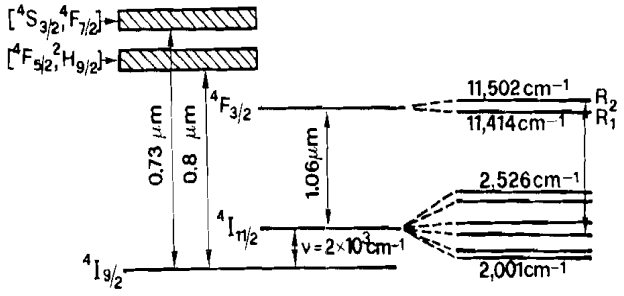

Figure 7.9: Energy level diagram for Nd:YAG, [3], p. 370.

The main emission of Nd:YAG is at 1.064μm. Another line with considerable less gain is at 1.32μm. Initially Nd:YAG was flashlamp pumped. Today, much more efficient pumping is possible with laser diodes and diode arrays. Diode pumped versions which can be very compact and efficient become a competition for the CO_2 laser in material processing, range finding, surgery, pumping of other lasers in combination with frequency doubling to produce a green 532nm beam).

Neodymium can also be doped in a host of other crystals such as YLF (Nd:YLF) emitting at 1047μm, YVO4 (Nd:YVO) emitting at 1.064μm, glass (Nd:Glass) at 1.062μm (Silicate glasses), 1.054μm (Phosphate glasses). Glass lasers have been used to build extremely high power (Terawatt), high energy (Megajoules) multiple beam systems for inertial confinement fusion. The big advantage of glass is that it can be fabricated on meter scale which is hard or even impossible to do with crystalline materials.

Other rare earth elements are Er^{3+}, Tm^{3+}, Ho^{3+}, Er^{3+}, which have emmission lines at 1.53μm and in the 2-3μm range.

Ytterbium YAG (Yb:YAG)

Ytterbium YAG is a quasi three level laser, see Figure 303 emitting at 1.030μm. The lower laser level is only 500-600cm^{-1} (60meV) above the ground state and is therefore at room temperature heavily thermally populated. The laser is pumped at 941 or 968nm with laser diodes to provide the high brighness pumping needed to achieve gain.

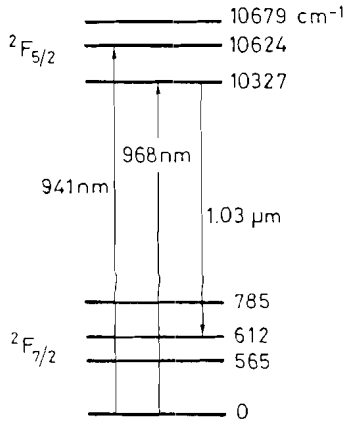

Figure 7.10: Energy level diagram of Yb:YAG, [3], p. 374.

However, Yb:YAG has many advantages over other laser materials:

- Very low quantum defect, i.e. difference between the photon energy necessary for pumping and photon energy of the emitted radiation, $(hf_P - hf_L)/hf_P \tilde{} 9\%$.
- long radiative lifetime of the upper laser level, i.e. much energy can be stored in the crystal.
- high doping levels can be used without upper state lifetime quenching
- broad emission bandwidth of $\Delta f_{FWHM} = 2.5$THz enabling the generation of sub-picosecond pulses

- with cryogenic cooling Yb:YAG becomes a four level laser.

Due to the low quantum defect and the good thermal properties of YAG, Yb:YAG lasers approaching an optical to optical efficiency of 80% and a wall plug efficiency of 40% have been demonstrated.

Titanium Sapphire (Ti:sapphire)

In contrast to Neodymium, which is a rare earth element, Titanium is a transition metal. The Ti^{3+} ions replace a certain fraction of the Al^{3+} ions in sapphire (Al_2O_3). In transistion metal lasers, the laser active electronic states are outer 3s electrons which couple strongly to lattice vibrations. These lattice vibrations lead to strong line broadening. Therefore, Ti:sapphire has an extremely broad amplification linewidth $\Delta f_{FWHM} \approx 100$THz. Ti:sapphire can provide gain from 650-1080nm. Therefore, this material is used in todays highly-tunable or very short pulse laser systems and amplifiers. Once Ti:sapphire was developed it rapidly replaced the dye laser systems. Figure 7.11 shows the absorption and emission bands of Ti:sapphire for polarization along its optical axis ($\pi-$polarization).

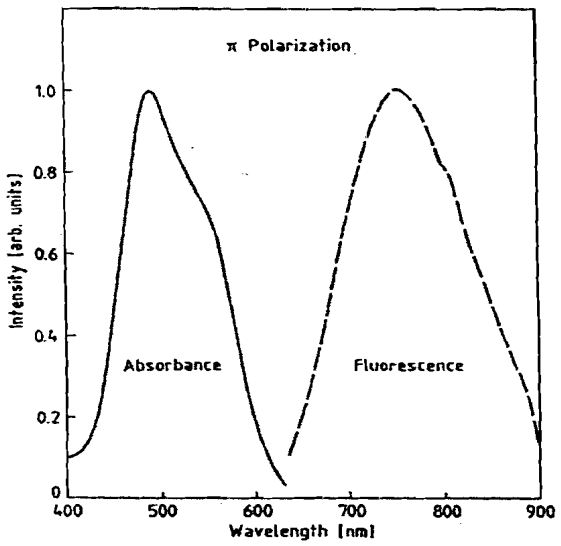

Figure 7.11: Absorption and flourescence spectra of Ti:sapphire, [10]

7.3.4 Semiconductor Lasers

An important class of solid-state lasers are semiconductor lasers. Depending on the semiconductor material used the emission wavelength can be further refined by using bandstructure engineering, 0.4 μm (GaN) or 0.63-1.55 μm (AlGaAs, InGaAs, InGaAsP) or 3-20 μm (lead salt). The AlGaAs based lasers in the wavelength range 670nm-780 nm are used in compact disc players and therefore are the most common and cheapest lasers in the world. In the semiconductor laser the electronic bandstructure is exploited, which arises from the periodic crystal potential, see problem set. The energy eigenstates can be characterized by the periodic crystal quasi momentum vector $\vec{k}$, see Figure

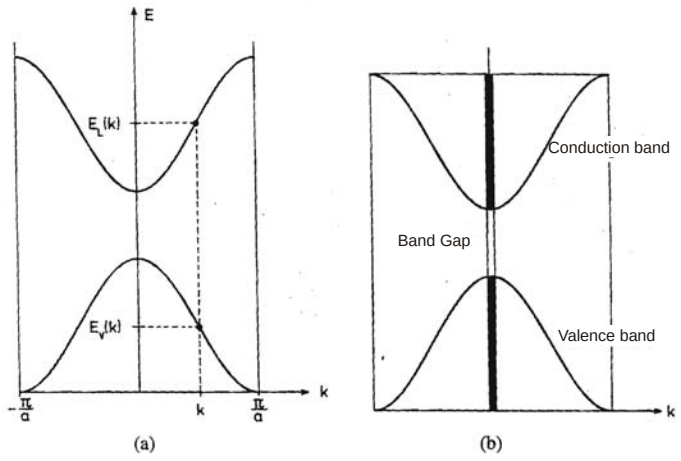

Figure 7.12: (a) Energy level diagram of the electronic states in a crystaline solid-state material. There is usually a highest occupied band, the valence band and a lowest unoccupied band the conduction band. Electronics states in a crystal can usually be characterized by their quasi momentum $\vec{k}$. b) The valence and conduction band are separated by a band gap.

Since the momentum carried along by an optical photon is very small compared to the momentum of the electrons in the crystal lattice, transistions of an electron from the valence band to the conduction band occur essentially vertically, see Figure 7.13 (a).

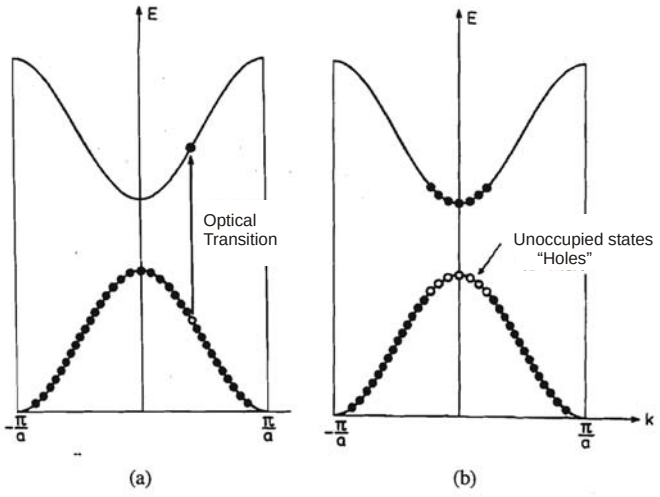

Figure 7.13: (a) At thermal equilibrium the valence band is occupied and the conduction band is unoccupied. Optical transistions occur vertically under momentum conservation, since the photon momentum is negligible compared to the momentum of the electrons. (b) To obtain amplification, the medium must be inverted, i.e. electrones must be accumulated in the conduction band and empty states in the valence band. The missing electron behave as a positively charged particles called holes.

Inversion, i.e. electrons in the conduction band and empty states in the valence band, holes, see Figure 7.13 (b) can be achieved by creating a pn-junction diode and forward biasing, see Figure 7.14.

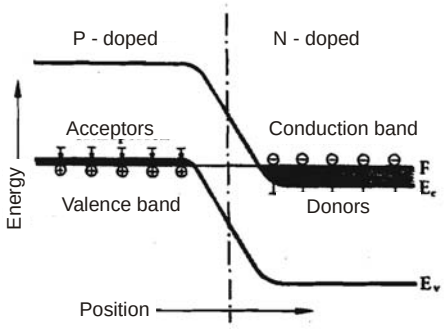

Figure 7.14: Forward biased pn-junction laser diode. Electrons and holes are injected into the space charge region of a pn-junction and emit light by recombination.

When forward-biased electrons and holes are injected into the space charge region. The carriers recombine and emit the released energy in the form of photons with an energy roughly equal to the band gap energy. A sketch of a typical pn-junction diode laser is shown in Figure 7.15.

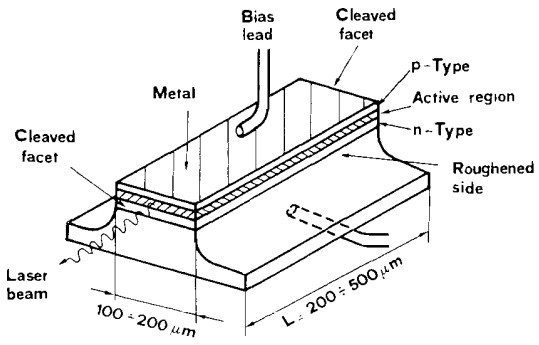

Figure 7.15: Typical broad area pn-homojunction laser, [3], p. 397.

The devices can be further refined by using heterojunctions so that the carriers are precisely confined to the region of the waveguide mode, see Figure

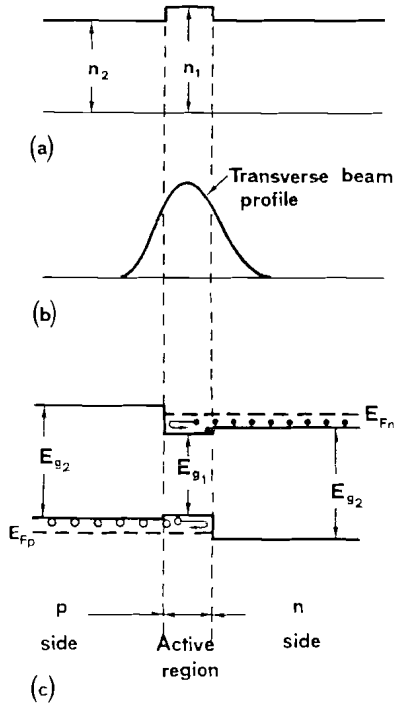

Figure 7.16: a) Refractive index profile. b) transverse beam profile, and c) band structure (shematic) of a double-heterostructure diode laser, [3], p. 399.

.

7.3.5 Quantum Cascade Lasers

A new form of semiconductor lasers was predicted in the 70's by the two russian physicists Kazarinov and Suris that is based only on one kind of electrical carriers. These are most often chosen to be electrons because of there higher mobility. This laser is therefore a unipolar device in contrast to the conventional semiconductor laser that uses both electrons and holes. the transitions are intraband transistions. A layout of a quantum cascade laser is shown in Figure 7.17.

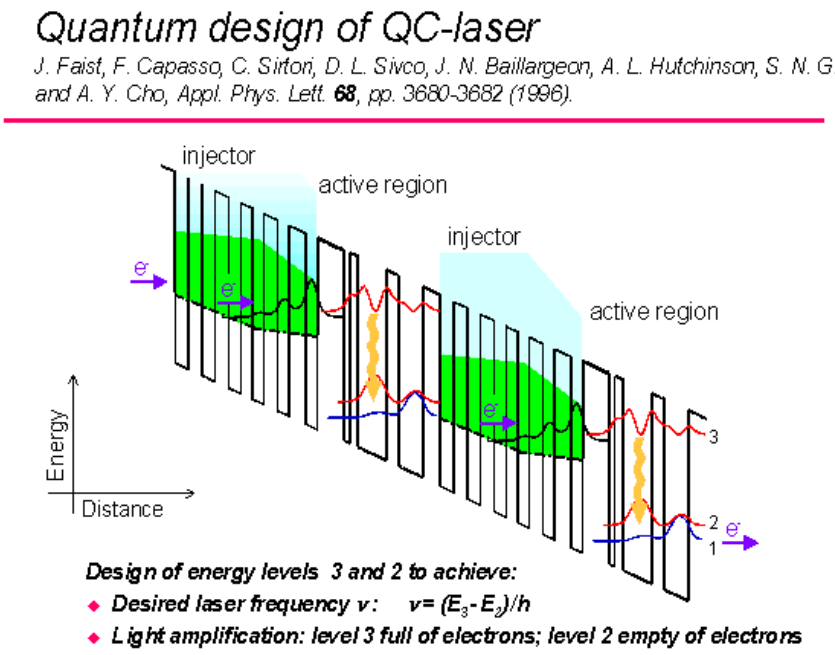

Figure 7.17: Quantum Cascade laser layout.

Like semiconductor lasers these lasers are electrically pumped. The first laser of this type was realized in 1994 by Federico Capasso's group at Bell Laboratories [9], 23 years after the theoretical prediction. The reason for this is the difficult layer growth, that are only possible using advanced semiconductor growth capabilities such as molecular beam epitaxy (MBE) and more recently metal oxide chemical vapor depostion (MOCVD). Lasers have been demostrated in the few THz range [13] up to the 3.5μm region.

Some of the most important spectroscopic parameters of often used laser media are summarized in table 7.1.

7.3.6 Homogeneous and Inhomogeneous Broadening

Laser media are also distinguished by the line broadening mechanisms involved. Very often it is the case that the linewidth observed in the absorption or emission spectrum is not only due to dephasing process that are acting on

Laser Medium	Wave-length $\lambda_0(n\text{m})$	Cross Section σ (cm^2)	Upper-St. Lifetime τ_L (μs)	Linewidth $\Delta f_{FWHM} = \frac{2}{T_2}$(THz)	Typ	Refr. index n
Nd^{3+}:YAG	1,064	$4.1 \cdot 10^{-19}$	1,200	0.210	H	1.82
Nd^{3+}:LSB	1,062	$1.3 \cdot 10^{-19}$	87	1.2	H	1.47 (ne)
Nd^{3+}:YLF	1,047	$1.8 \cdot 10^{-19}$	450	0.390	H	1.82 (ne)
Nd^{3+}:YVO_4	1,064	$2.5 \cdot 10^{-19}$	50	0.300	H	2.19 (ne)
Nd^{3+}:glass	1,054	$4 \cdot 10^{-20}$	350	3	H/I	1.5
Er^{3+}:glass	1,55	$6 \cdot 10^{-21}$	10,000	4	H/I	1.46
Ruby	694.3	$2 \cdot 10^{-20}$	1,000	0.06	H	1.76
Ti^{3+}:Al_2O_3	660-1180	$3 \cdot 10^{-19}$	3	100	H	1.76
Cr^{3+}:LiSAF	760-960	$4.8 \cdot 10^{-20}$	67	80	H	1.4
Cr^{3+}:LiCAF	710-840	$1.3 \cdot 10^{-20}$	170	65	H	1.4
Cr^{3+}:LiSGAF	740-930	$3.3 \cdot 10^{-20}$	88	80	H	1.4
He-Ne	632.8	$1 \cdot 10^{-13}$	0.7	0.0015	I	$\sim$1
Ar^+	515	$3 \cdot 10^{-12}$	0.07	0.0035	I	$\sim$1
CO_2	10,600	$3 \cdot 10^{-18}$	2,900,000	0.000060	H	$\sim$1
Rhodamin-6G	560-640	$3 \cdot 10^{-16}$	0.0033	5	H	1.33
semiconductors	450-30,000	$\sim 10^{-14}$	~ 0.002	25	H/I	3 - 4

Table 7.1: Wavelength range, cross-section for stimulated emission, upper-state lifetime, linewidth, typ of lineshape (H=homogeneously broadened, I=inhomogeneously broadened) and index for some often used solid-state laser materials, and in comparison with semiconductor and dye lasers.

all atoms in the same, i.e. homogenous way. Lattice vibrations that lead to a line broadening of electronic transisitions of laser ions in the crystal act in the same way on all atoms in the crystal. Such mechanisms are called homogeneous broadening. However, It can be that in an atomic ensemble there are groups of atoms with a different center frequency of the atomic transistion. The overall ensemble therefore may eventually show a very broad linewidth but it is not related to actual dephasing mechanism that acts upon each atom in the ensemble. This is partially the case in Nd:silicate glass lasers, see table 7.1 and the linewidth is said to be inhomogeneously broadened. Wether a transistion is homogenously or inhomogeneously broadened can be tested by using a laser to saturate the medium. In a homogenously

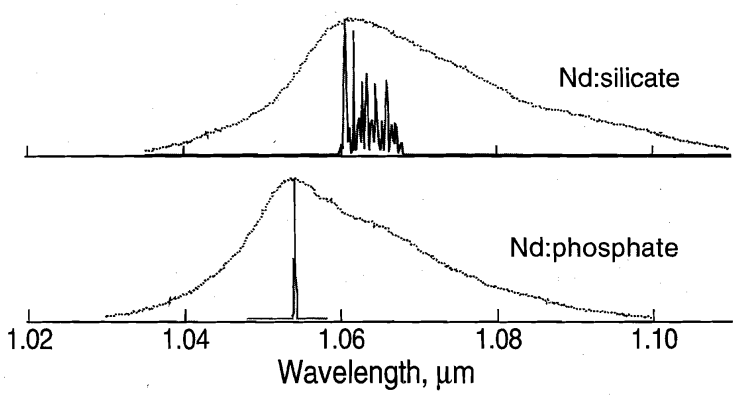

Figure 7.18: Laser with inhomogenously broaden laser medium (Nd:silicate glass) and homogenously broadened laser medium (Nd:phosphate glass), [14]

broadened medium the loss or gain saturates homogenously, i.e. the whole line is reduced. In an inhomogenously broadened medium a spectral hole burning occurs, i.e. only that sub-group of atoms that are sufficiently in resonance with the driving field saturate and the others not, which leads to a hole in the spectral distribution of the atoms. Figure 7.18 shows the impact of an inhomogeneously broadened gain medium on the continous wave output spectrum of a laser. In homogenous broadening leads to lasing of many longitudinal laser modes because inhomogenous saturation of the gain. In the homogenously broadened medium the gain saturates homogenously and only one or a few modes can lase. An important inhomogenous broadening mechanism in gases is doppler broadening. Due to the motion of the atoms in a gas relative to an incident electromagnetic beam, the center frequency of each atomic transistion is doppler shifted according to its velocity by

$$f = \left(1 \pm \frac{v}{c}\right) f_0, \tag{7.5}$$

where the plus sign is correct for an atom moving towards the beam and the minus sign for a atom moving with the beam. The velocity distribution of an ideal gas with atoms or molecules of mass m in thermal equilibrium is given by the Maxwell-Boltzman distribution

$$p(v) = \sqrt{\frac{m}{2\pi kT}} \exp\left(-\frac{mv^2}{2kT}\right). \tag{7.6}$$

This means that $p(v)dv$ is equal to the brobability that the atom or molecule has a velocity in the interval $[v, v + dv]$. Here, v is the component of the velocity that is in the direction of the beam. If the homogenous linewidth of the atoms is small compared to the doppler broading, we obtain the lineshape of the inhomogenously broadened gas simply by substituting the velocity by the induced frequency shift due to the motion

$$v = c\,\frac{f - f_0}{f_0} \tag{7.7}$$

Then the lineshape is a Gaussian

$$g(f) = \sqrt{\frac{mc}{2\pi kT f_0}}\exp\left[-\frac{mc^2}{2kT}\left(\frac{f - f_0}{f_0}\right)^2\right]. \tag{7.8}$$

The full width at half maximum of the line is

$$\Delta f = 8\ln(2)\sqrt{\frac{kT}{mc^2}}f_0. \tag{7.9}$$

7.4 Laser Dynamics (Single Mode)

In this section we want study the single mode laser dynamics. The laser typically starts to lase in a few closely spaced longitudinal modes, which are incoherent with each other and the dynamics is to a large extent similar to the dynamics of a single mode that carries the power of all lasing modes. To do so, we complement the rate equations for the populations in the atomic medium, that can be reduced to the population of the upper laser level Eq.(7.4) as discussed before with a rate equation for the photon population in the laser mode.

There are two different kinds of laser cavities, linear and ring cavities, see Figure 7.19

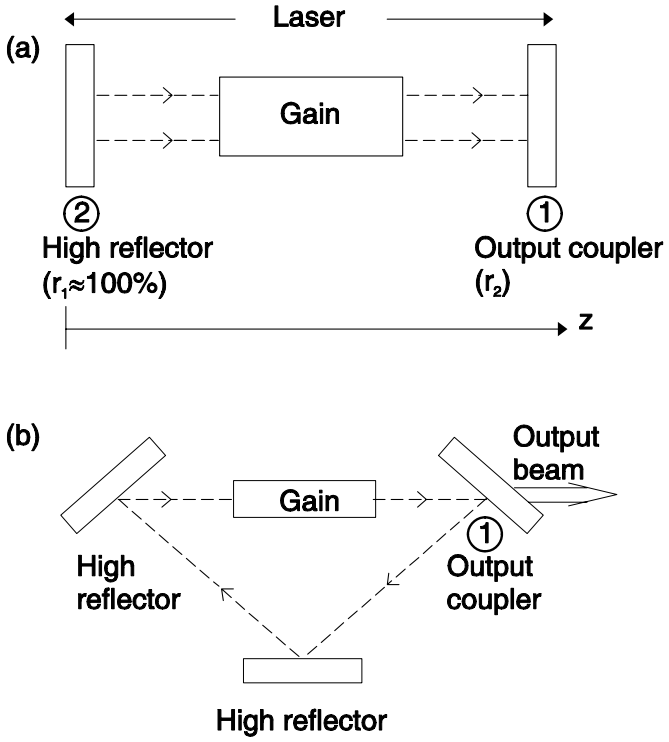

Figure 7.19: Possible cavity configurations. (a) Schematic of a linear cavity laser. (b) Schematic of a ring laser.

The laser resonators can be modelled as Fabry Perots as discussed in section . Typically the techniques are used to avoid lasing of transverse modes and only the longitudinal modes are of interest. The resonance frequencies of the longitudinal modes are determined by the round trip phase to be a multiple of 2π

$$\phi(\omega_m) = 2m\pi. \tag{7.10}$$

neighboring modes are space in frequency by the inverse roundtrip time

$$\phi(\omega_0 + \Delta\omega) = \phi(\omega_0) + T_R\Delta\omega = 2m\pi. \tag{7.11}$$

T_R is the round trip time in the resonator, which is

$$T_R = \frac{2^* L}{\nu_g}, \tag{7.12}$$

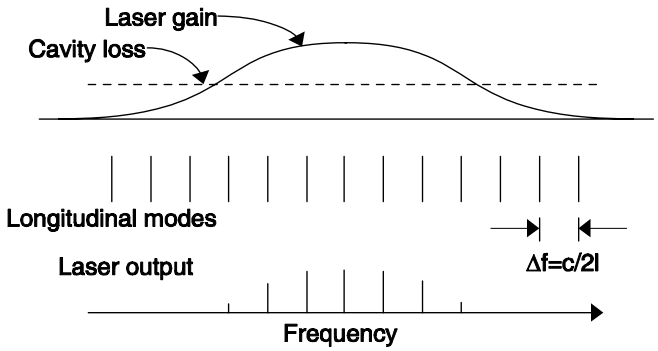

Figure 7.20: Laser gain and cavity loss spectra, longitudinal mode location, and laser output for multimode laser operation.

where ν_g is the group velocity in the cavity in the frequency range considered, and L is the cavity length of the linear or ring cavity and $2^* = 1$ for the ring cavity and $2^* = 2$ for the linear cavity. In the case of no dispersion, the longitudinal modes of the resonator are multiples of the inverse roundtrip timeand

$$f_m = m\frac{1}{T_R}. \tag{7.13}$$

The mode spacing of the longitudinal modes is

$$\Delta f = f_m - f_{m-1} = \frac{1}{T_R} \tag{7.14}$$

If we assume frequency independent cavity loss and Lorentzian shaped gain (see Fig. 7.20). Initially when the laser gain is larger then the cavity loss, many modes will start to lase. To assure single frequency operation a filter (etalon) can be inserted into the laser resonator, see Figure 7.21. If the laser is homogenously broadened the laser gain will satured to the loss level and only the mode at the maximum of the gain will lase. If the gain is not homogenously broadend and in the absence of a filter many modes will lase.

For the following we assume a homogenously broadend laser medium and only one cavity mode is able to lase. We want to derive the equations of motion for th population inversion, or population in the upper laser level and the photon number in that mode, see Figure 7.22.

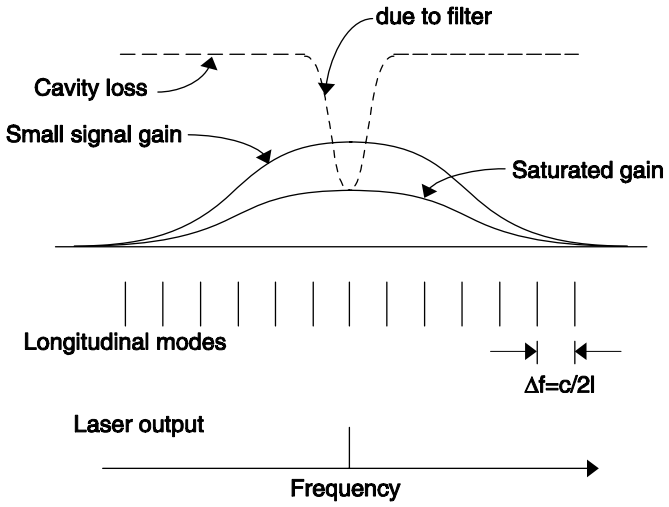

Figure 7.21: Gain and loss spectra, longitudinal mode locations, and laser output for single mode laser operation.

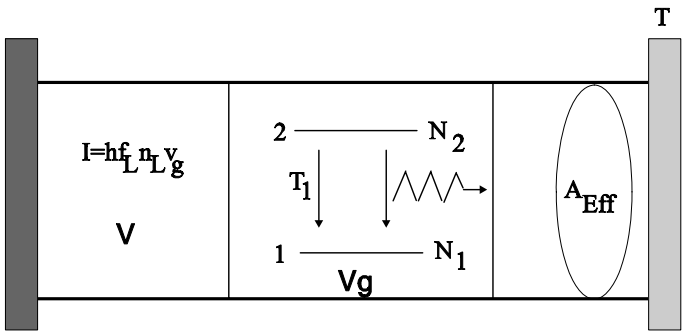

Figure 7.22: Rate equations for a laser with two-level atoms and a resonator.

The intensity I in a mode propagating at group velocity v_g with a mode volume V is related to the number of photons N_L or the number density $n_L = N_L/V$ stored in the mode with volume V by

$$I = hf_L \frac{N_L}{2^* V} v_g = \frac{1}{2^*} hf_L n_L v_g, \tag{7.15}$$

where hf_L is the photon energy. $2^* = 2$ for a linear laser resonator (then only half of the photons are going in one direction), and $2^* = 1$ for a ring

laser. In this first treatment we consider the case of space-independent rate equations, i.e. we assume that the laser is oscillating on a single mode and pumping and mode energy densities are uniform within the laser material. With the interaction cross section σ for stimulated emission defined as

$$\sigma = \frac{hf_L}{I_s T_1}, \tag{7.16}$$

and Eq. (7.4) with the number of atoms in the mode, we obtain

$$\frac{d}{dt}N_2 = -\frac{N_2}{\tau_L} - 2^*\sigma N_2 v_g n_L + R_p. \tag{7.17}$$

Here, $v_g n_L$ is the photon flux, σ is the stimulated emission cross section, $\tau_L = \gamma_{21}$ the upper state lifetime and R_p is the pumping rate into the upper laser level. A similar rate equation can be derived for the photon density

$$\frac{d}{dt}n_L = -\frac{n_L}{\tau_p} + 2^*\frac{\sigma v_g}{V}N_2\left(n_L + \frac{1}{V}\right). \tag{7.18}$$

Here, τ_p is the photon lifetime in the cavity or cavity decay time. The $1/V$ -term in Eq.(7.18) accounts for spontaneous emission which is equivalent to stimulated emission by one photon occupying the mode with mode volume V. For a laser cavity with a semi-transparent mirror with amplitude transmission T, see section 2.3.8, producing a power loss $2l = 2T$ per round-trip in the cavity, the cavity decay time is $\tau_p = 2l/T_R$, if $T_R = 2^*L/c_0$ is the roundtrip-time in linear cavity with optical length $2L$ or a ring cavity with optical length L. Eventual internal losses can be treated in a similar way and contribute to the cavity decay time. Note, the decay rate for the inversion in the absence of a field, $1/\tau_L$, is not only due to spontaneous emission, but is also a result of non radiative decay processes. See for example the four level system shown in Fig. 7.6.

So the two rate equations are

$$\frac{d}{dt}N_2 = -\frac{N_2}{\tau_L} - 2^*\sigma v_g N_2 n_L + R_p \tag{7.19}$$

$$\frac{d}{dt}n_L = -\frac{n_L}{\tau_p} + 2^*\frac{\sigma v_g}{V}N_2\left(n_L + \frac{1}{V}\right). \tag{7.20}$$

Experimentally, the photon number and the inversion in a laser resonator are not very convenient quantities, therefore, we normalize both equations to

the round-trip amplitude gain $g = 2^* \frac{\sigma v_g}{2V} N_2 T_R$ experienced by the light and the circulating intracavity power $P = I \cdot A_{eff}$

$$\frac{d}{dt}g = -\frac{g - g_0}{\tau_L} - \frac{gP}{E_{sat}} \tag{7.21}$$

$$\frac{d}{dt}P = -\frac{1}{\tau_p}P + \frac{2g}{T_R}\left(P + P_{vac}\right), \tag{7.22}$$

with

$$E_{sat} = \frac{hf_L}{2^*\sigma}A_{eff} = \frac{1}{2^*}I_s A_{eff}\tau_L \tag{7.23}$$

$$P_{sat} = E_s/\tau_L \tag{7.24}$$

$$P_{vac} = hf_L v_g/2^* L = hf_L/T_R \tag{7.25}$$

$$g_0 = 2^* \frac{R_p}{2A_{eff}}\sigma\tau_L, \tag{7.26}$$

the small signal round-trip gain of the laser. Note, the factor of two in front of gain and loss is due to the fact, that we defined g and l as gain and loss with respect to amplitude. Eq.(7.26) elucidates that the figure of merit that characterizes the small signal gain achievable with a certain laser material is the $\sigma\tau_L$-product.

7.5 Continuous Wave Operation

If $P_{vac} \ll P \ll P_{sat} = E_{sat}/\tau_L$, than $g = g_0$ and we obtain from Eq.(7.22), neglecting P_{vac}

$$\frac{dP}{P} = 2\left(g_0 - l\right)\frac{dt}{T_R} \tag{7.27}$$

or

$$P(t) = P(0)e^{2(g_0 - l)\frac{t}{T_R}}. \tag{7.28}$$

The laser power builts up from vaccum fluctuations, see Figure 7.23 until it reaches the saturation power, when saturation of the gain sets in within the built-up time

$$T_B = \frac{T_R}{2\left(g_0 - l\right)}\ln\frac{P_{sat}}{P_{vac}} = \frac{T_R}{2\left(g_0 - l\right)}\ln\frac{A_{eff}T_R}{\sigma\tau_L}. \tag{7.29}$$

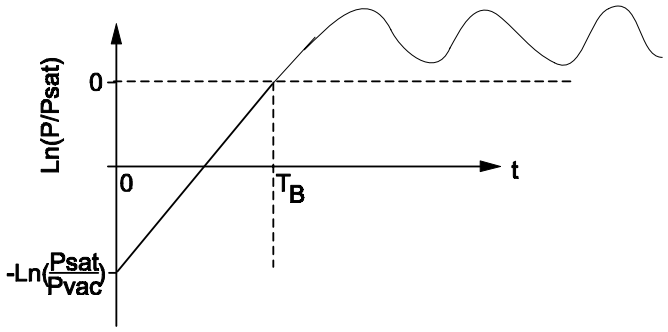

Figure 7.23: Built-up of laser power from spontaneous emission noise.

Some time after the built-up phase the laser reaches steady state, with the saturated gain and steady state power resulting from Eqs.(7.21-7.22), neglecting in the following the spontaneous emission, $P_{vac} = 0$, and for $\frac{d}{dt} = 0$:

$$g_s = \frac{g_0}{1 + \frac{P_s}{P_{sat}}} = l \tag{7.30}$$

$$P_s = P_{sat}\left(\frac{g_0}{l} - 1\right), \tag{7.31}$$

Figure 7.24 shows output power and gain as a function of small signal gain g_0, which is proportional to the pump rate. Below threshold, the output power is zero and the gain increases linearly with in crease pumping. After reaching threshold the gain stays clamped at the threshold value determined by gain equal loss and the output power increases linearly.

7.6 Stability and Relaxation Oscillations

How does the laser reach steady state, once a perturbation has occured?

$$g = g_s + \Delta g \tag{7.32}$$

$$P = P_s + \Delta P \tag{7.33}$$

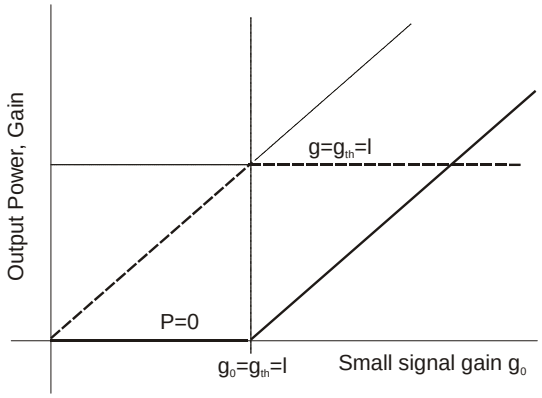

Figure 7.24: Output power and gain of a laser as a function of pump power.

Substitution into Eqs.(7.21-7.22) and linearization leads to

$$\frac{d\Delta P}{dt} = +2\frac{P_s}{T_R}\Delta g \tag{7.34}$$

$$\frac{d\Delta g}{dt} = -\frac{g_s}{E_{sat}}\Delta P - \frac{1}{\tau_{stim}}\Delta g \tag{7.35}$$

where $\frac{1}{\tau_{stim}} = \frac{1}{\tau_L}\left(1 + \frac{P_s}{Psat}\right)$ is the inverse stimulated lifetime. The stimulated lifetime is the lifetime of the upper laser state in the presence of the optical field. The perturbations decay or grow like

$$\begin{pmatrix} \Delta P \\ \Delta g \end{pmatrix} = \begin{pmatrix} \Delta P_0 \\ \Delta g_0 \end{pmatrix} e^{st}. \tag{7.36}$$

which leads to the system of equations (using $g_s = l$)

$$A \begin{pmatrix} \Delta P_0 \\ \Delta g_0 \end{pmatrix} = \begin{pmatrix} -s & 2\frac{P_s}{T_R} \\ -\frac{T_R}{E_{sat}2\tau_p} & -\frac{1}{\tau_{stim}} - s \end{pmatrix} \begin{pmatrix} \Delta P_0 \\ \Delta g_0 \end{pmatrix} = 0. \tag{7.37}$$

There is only a solution, if the determinante of the coefficient matrix vanishes, i.e.

$$s\left(\frac{1}{\tau_{stim}} + s\right) + \frac{P_s}{E_{sat}\tau_p} = 0, \tag{7.38}$$

which determines the relaxation rates or eigen frequencies of the linearized system

$$s_{1/2} = -\frac{1}{2\tau_{stim}} \pm \sqrt{\left(\frac{1}{2\tau_{stim}}\right)^2 - \frac{P_s}{E_{sat}\tau_p}}. \tag{7.39}$$

Introducing the pump parameter $r = 1 + \frac{P_s}{P_{sat}}$, which tells us how often we pump the laser over threshold, the eigen frequencies can be rewritten as

$$s_{1/2} = -\frac{1}{2\tau_{stim}} \left(1 \pm j \sqrt{\frac{4\,(r-1)}{r} \frac{\tau_{stim}}{\tau_p} - 1} \right), \tag{7.40}$$

$$= -\frac{r}{2\tau_L} \pm j \sqrt{\frac{(r-1)}{\tau_L \tau_p} - \left(\frac{r}{2\tau_L} \right)^2} \tag{7.41}$$

There are several conclusions to draw:

- (i): The stationary state $(0, g_0)$ for $g_0 < l$ and (P_s, g_s) for $g_0 > l$ are always stable, i.e. $\mathrm{Re}\{s_i\} < 0$.
- (ii): For lasers pumped above threshold, $r > 1$, and long upper state lifetimes, i.e. $\frac{r}{4\tau_L} < \frac{1}{\tau_p}$,

  the relaxation rate becomes complex, i.e. there are relaxation oscillations

  $$s_{1/2} = -\frac{1}{2\tau_{stim}} \pm j\omega_R. \tag{7.42}$$

  with a frequency ω_R approximately equal to the geometric mean of inverse stimulated lifetime and photon life time

  $$\omega_R \approx \sqrt{\frac{1}{\tau_{stim}\tau_p}}. \tag{7.43}$$

- If the laser can be pumped strong enough, i.e. r can be made large enough so that the stimulated lifetime becomes as short as the cavity decay time, relaxation oscillations vanish.

The physical reason for relaxation oscillations and instabilities related to it is, that the gain reacts to slow on the light field, i.e. the stimulated lifetime is long in comparison with the cavity decay time.

Example: diode-pumped Nd:YAG-Laser

$$\lambda_0 = 1064 \text{ nm}, \sigma = 4 \cdot 10^{-20} cm^2,\ A_{eff} = \pi\,(100\mu m \times 150\mu m)\,, r = 50$$
$$\tau_L = 1.2\ ms,\ l = 1\%,\ T_R = 10ns$$

From Eq.(7.16) we obtain:

$$\begin{aligned} I_{sat} &= \frac{hf_L}{\sigma\tau_L} = 3.9\frac{kW}{cm^2},\ P_{sat} = I_{sat}A_{eff} = 1.8\ W,\ P_s = 91.5W \\ \tau_{stim} &= \frac{\tau_L}{r} = 24\mu s,\ \tau_p = 1\mu s, \omega_R = \sqrt{\frac{1}{\tau_{stim}\tau_p}} = 2\cdot 10^5 s^{-1}. \end{aligned}$$

Figure 7.25 shows the typically observed fluctuations of the output of a solid-

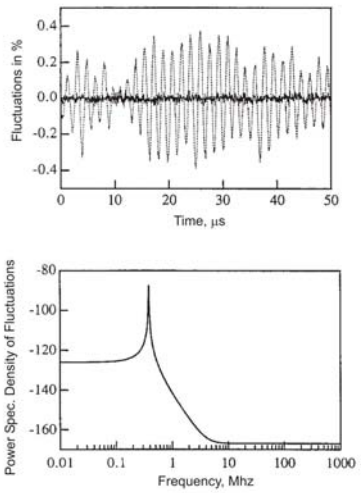

Figure 7.25: Relaxation oscillations in the time and frequency domain.

state laser in the time and frequency domain. Note, that this laser has a long upperstate lifetime of several 100 μs

One can also define a quality factor for the relaxation oscillations by the ratio of the imaginary to the real part of the complex eigen frequencies 7.41

$$Q = \sqrt{\frac{4\tau_L}{\tau_p}\frac{(r-1)}{r^2}}. \tag{7.44}$$

The quality factor can be as large a several thousand for solid-state lasers with long upper-state lifetimes in the millisecond range.

7.7 Laser Efficiency

An important measure for a laser is the efficiency with which pump power is converted into laser output power. To determine the efficiency we must review the important parameters of a laser and the limitations these parameters impose.

From Eq.(7.31) we found that the steady state intracavity power P_s of a laser is

$$P_s = P_{sat}\left(\frac{2g_0}{2l} - 1\right), \tag{7.45}$$

where $2g_0$ is the small signal round-trip power gain, P_{sat} the gain saturation power and $2l$ is the power loss per round-trip. Both parameters are expressed in terms of the fundamental pump parameter R_p, $\sigma\tau_L$-product and mode cross section A_{eff} of the gain medium. For this derivation it was asummed that all pumped atoms are in the laser mode with constant intensity over the beam cross section

$$2g_0 = 2^* \frac{R_p}{A_{eff}}\sigma\tau_L, \tag{7.46}$$

$$P_{sat} = \frac{hf_L}{2^*\sigma\tau_L}A_{eff} \tag{7.47}$$

The power losses of lasers are due to the internal losses $2l_{int}$ and the transmission T through the output coupling mirror. The internal losses can be a significant fraction of the total losses. The output power of the laser is

$$P_{out} = T \cdot P_{sat}\left(\frac{2g_0}{2l_{int} + T} - 1\right) \tag{7.48}$$

The pump power of a laser is minimized given

$$P_p = R_p h f_P, \tag{7.49}$$

where hf_P is the energy of the pump photons. In discussing the efficiency of a laser, we consider the overall efficiency

$$\eta = \frac{P_{out}}{P_p} \tag{7.50}$$

which approaches the differential efficiency η_D if the laser is pumped many times over threshold, i.e. $r = 2g_0/2l \to \infty$

$$\eta_D = \frac{\partial P_{out}}{\partial P_p} = \eta(r \to \infty) \tag{7.51}$$

$$= \frac{T}{2l_{int} + T} P_{sat} \frac{2^*}{A_{eff} h f_P} \sigma \tau_L \tag{7.52}$$

$$= \frac{T}{2l_{int} + T} \cdot \frac{h f_L}{h f_P}. \tag{7.53}$$

Thus the efficiency of a laser is fundamentally limited by the ratio of output coupling to total losses and the quantum defect in pumping. Therefore, one would expect that the optimum output coupling is achieved with the largest output coupler, however, this is not true as we considered the case of operating many times above threshold.

7.8 "Thresholdless" Lasing

So far we neglected the spontaneous emission into the laser mode. This is justified for large lasers where the density of radiation modes in the laser medium is essentially the free space mode density and effects very close to threshold are not of interest. For lasers with small mode volume, or a laser operating very close to threshold, the spontaneous emission into the laser mode can no longer be neglected and we should use the full rate equations (7.21) and (7.22)

$$\frac{d}{dt} g = -\frac{g - g_0}{\tau_L} - \frac{gP}{E_{sat}} \tag{7.54}$$

$$\frac{d}{dt} P = -\frac{1}{\tau_p} P + \frac{2g}{T_R} \left(P + P_{vac} \right), \tag{7.55}$$

where P_{vac} is the power of a single photon in the mode. The steady state conditions are

$$g_s = \frac{g_0}{(1 + P_s/P_{sat})}, \tag{7.56}$$

$$0 = \left(2g_s - 2l \right) P + 2 g_s P_{vac}. \tag{7.57}$$

Substitution of the saturated gain condition (7.56) into (7.57) and using the pump parameter $r = 2g_0/2l$, leads to a quadratic equation for the normalized intracavity steady state power $p{=}P_s/P_{sat}$ in terms of normalized vacuum power $p_v = P_{vac}/P_{sat} = \sigma\tau_L v_g/V$. This equation has the solutions

$$p = \frac{r-1+rp_v}{2} \pm \sqrt{\left(\frac{r-1+rp_v}{2}\right)^2 + (rp_v)^2}. \tag{7.58}$$

where only the solution with the plus sign is of physical significance. Note, the typical value for the $\sigma\tau_L$-product of the laser materials in table 7.1 is $\sigma\tau_L = 10^{-23} cm^2 s$. If the volume is measured in units of wavelength cubed we obtain $p_v = 0.3/\beta$ for$\lambda = 1\mu m, V = \beta\lambda^3$ and $v_g = c$. Figure 7.26 shows the behavior of the intracavity power as a function of the pump parameter for various values of the normalized vacuum power.

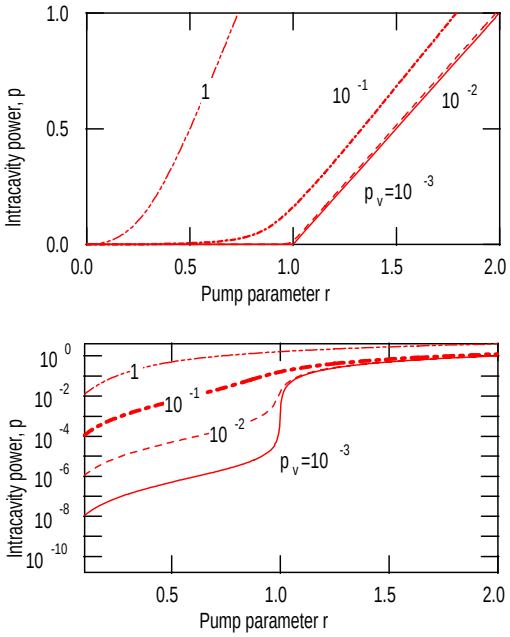

Figure 7.26: Intracavity power as a function of pump parameter r on a linear scale (a) and a logarithmic scale (b) for various values of the normalized vacuum power.

Figure 7.26 shows that for lasers with small mode volumes, i.e. mode volumes of the size of the wavelength cubed, the threshold is no longer well defined.

7.9 Short pulse generation by Q-Switching

The energy stored in the laser medium can be released suddenly by increasing the Q-value of the cavity so that the laser reaches threshold. This can be done actively, for example by quickly moving one of the resonator mirrors in place or passively by placing a saturable absorber in the resonator [**?**, 8]. Hellwarth first suggest this method only one year after the invention of the laser. As a rough orientation for a solid-state laser the following relation for the relevant time scales is generally valid

$$\tau_L \gg T_R \gg \tau_p. \tag{7.59}$$

7.9.1 Active Q-Switching

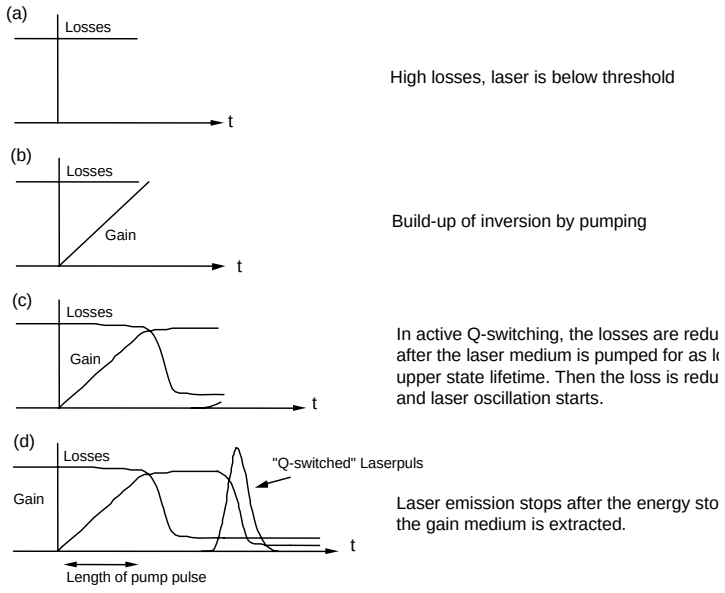

Figure 7.27: Gain and loss dynamics of an actively Q-switched laser.

Fig. 7.27 shows the principle dynamics of an actively Q-switched laser. The laser is pumped by a pump pulse with a length on the order of the upper-state lifetime, while the intracavity losses are kept high enough so that the laser can not reach threshold. At this point, the laser medium acts as energy storage with the energy slowly relaxing by spontaneous and nonradiative transitions. Intracavity loss is suddenly reduced, for example by a rotating cavity mirror. The laser is pumped way above threshold and the light field builts exponentially up until the pulse energy comes close to the saturation energy of the gain medium. The gain saturates and its energy is extracted, causing the laser to be shut off by the pulse itself.

A typical actively Q-switched pulse is asymmetric: The rise time is proportional to the net gain after the Q-value of the cavity is actively switched to a high value. The light intensity grows proportional to $2g_0/T_R$. When the gain is depleted, the fall time mostly depends on the cavity decay time τ_p, see Figure 7.28. For short Q-switched pulses a short cavity length, high gain and a large change in the cavity Q is necessary. If the Q-switch is not fast, the pulse width may be limited by the speed of the switch. Typical time scales for electro-optical and acousto-optical switches are 10 ns and 50 ns, respectively

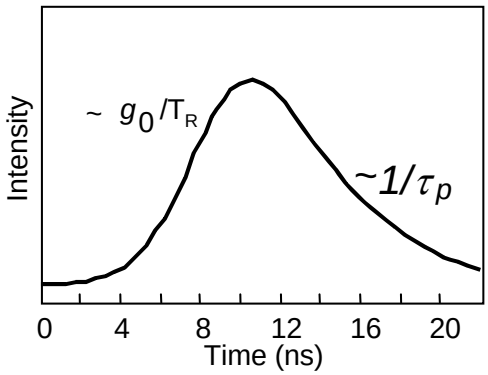

Figure 7.28: Asymmetric actively Q-switched pulse.

For example, with a diode-pumped Nd:YAG microchip laser [15] using an electro-optical switch based on $LiTaO_3$ Q-switched pulses as short as 270 ps at repetition rates of 5 kHz, peak powers of 25 kW at an average power of 34 mW, and pulse energy of 6.8 μJ have been generated (Figure 7.29).

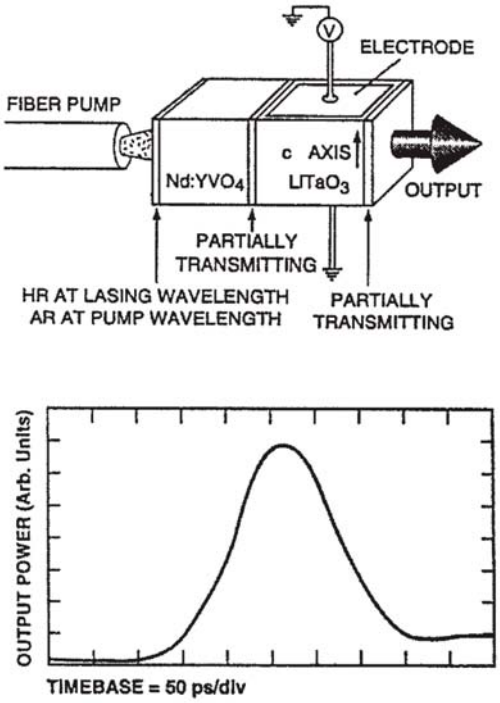

Figure 7.29: Q-switched microchip laser using an electro-optic switch. The pulse is measured with a sampling scope [15]

7.9.2 Passive Q-Switching

In the case of passive Q-switching, the intracavity loss modulation is performed by a saturable absorber, which introduces large losses for low intensities of light and small losses for high intensity.

Relaxation oscillations are due to a periodic exchange of energy stored in the laser medium by the inversion and the light field. Without the saturable absorber these oscillations are damped. If for some reason there is t0o much gain in the system, the light field can build up quickly. Especially for a low gain cross section the back action of the growing laser field on the inversion is weak and it can grow further. This growth is favored in the presence of loss that saturates with the intensity of the light. The laser becomes unstable and the field intensity growth as long as the gain does not saturate below the net loss, see Fig.7.30.

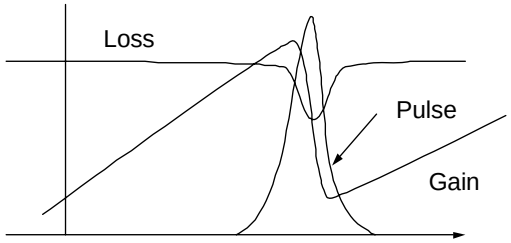

Figure 7.30: Gain and loss dynamics of a passively Q-switched laser

The saturable absorber leads to a destabilization of the relaxation oscillations resulting in the giant laser pulses.

7.10 Short pulse generation by mode locking

Q-switching is a single mode pheonomenon, i.e. the pulse build-up and decay occurs over many round-trips via build-up of energy and decay in a single (or a few) longitudinal mode. If one could get several longitudinal modes lasing in a phase coherent fashion with resepct to each other, these modes would be the Fourier components of a periodic pulse train emitted from the laser. Then a single pulse is traveling inside the laser cavity. This is called mode locking and the pulses generated are much shorter than the cavity round-trip time due to interference of the fields from many modes. The electric field can be written as a superpostion of the longitudinal modes. We neglect polarization for the moment.

$$E(z,t) = \Re\left[\sum_m \hat{E}_m e^{\mathrm{j}(\omega_m t - k_m z + \phi_m)}\right], \tag{7.60a}$$

$$\omega_m = \omega_0 + m\Delta\omega = \omega_0 + \frac{m\pi c}{\ell}, \tag{7.60b}$$

$$k_m = \frac{\omega_m}{c}. \tag{7.60c}$$

Equation (7.60a) can be rewritten as

$$E(z,t) = \Re\left\{ e^{\mathrm{j}\omega_0(t-z/c)} \sum_m \hat{E}_m e^{\mathrm{j}(m\Delta\omega(t-z/c)+\phi_m)} \right\} \tag{7.61a}$$

$$= \Re\left[A(t-z/c) e^{\mathrm{j}\omega_0(t-z/c)} \right] \tag{7.61b}$$

with the complex envelope

$$A\left(t-\frac{z}{c}\right) = \sum_m E_m e^{\mathrm{j}(m\Delta\omega(t-z/c)+\phi_m)} = \text{complex envelope (slowly varying).} \tag{7.62}$$

$e^{\mathrm{j}\omega_0(t-z/c)}$ is the carrier wave (fast oscillation). Here, both the carrier and the envelope travel with the same speed (no dispersion assumed). The envelope function is periodic with period

$$T = \frac{2\pi}{\Delta\omega} = \frac{2\ell}{c} = \frac{L}{c}, \tag{7.63}$$

where L is the optical round-trip length in the cavity. If we assume that N modes with equal amplitudes $E_m = E_0$ and equal phases $\phi_m = 0$ are lasing, the envelope is given by

$$A(z,t) = E_0 \sum_{m=-(N-1)/2}^{(N-1)/2} e^{\mathrm{j}(m\Delta\omega(t-z/c))}. \tag{7.64}$$

With

$$\sum_{m=0}^{q-1} a^m = \frac{1-a^q}{1-a}, \tag{7.65}$$

we obtain

$$A(z,t) = E_0 \frac{\sin\left[\frac{N\Delta\omega}{2}\left(t-\frac{z}{c}\right)\right]}{\sin\left[\frac{\Delta\omega}{2}\left(t-\frac{z}{c}\right)\right]}. \tag{7.66}$$

The laser intensity I is proportional to $E(z,t)^2$ averaged over one optical cycle: $I \sim |A(z,t)|^2$. At $z = 0$, we obtain

$$I(t) \sim |E_0|^2 \frac{\sin^2\left(\frac{N\Delta\omega t}{2}\right)}{\sin^2\left(\frac{\Delta\omega t}{2}\right)}. \tag{7.67}$$

The periodic pulses given by Eq. (7.67) have

- the period: $T = 1/\Delta f = L/c$
- pulse duration: $\Delta t = \frac{2\pi}{N\Delta\omega} = \frac{1}{N\Delta f}$
- peak intensity $\sim N^2|E_0|^2$
  - average intensity $\sim N|E_0|^2 \Rightarrow$ peak intensity is enhanced by a factor N.

If the phases of the modes are not locked, i.e. ϕ_m is a random sequence then the intensity fluctuates randomly about its average value ($\sim N|E_0|^2$), which is the same as in the mode-locked case. Figure 7.31 shows the intensity of a modelocked laser versus time,if the relative phases of the modes to each other is (a) constant and (b) random

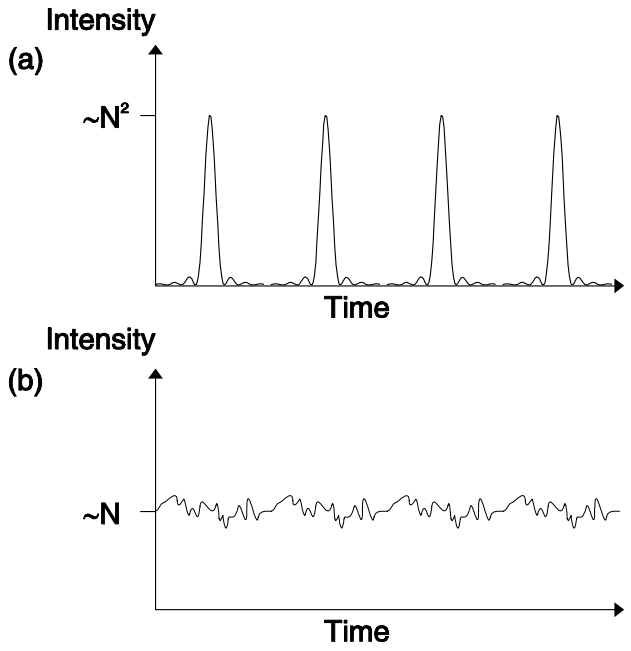

Figure 7.31: Laser intensity versus time from a mode-locked laser with (a) perfectly locked phases and (b) random phases.

We are distinguishing between active modelocking, where an external loss modulator is inserted in the cavity to generate modelocked pulses, and passive modelocking, where the pulse is modulating the intracavity itself via a saturable absorber.

7.10.1 Active Mode Locking

Active mode locking was first investigated in 1970 by Kuizenga and Siegman [16] and later by Haus [17] . In the approach by Haus, modelocking is treated as a pulse propagation problem.

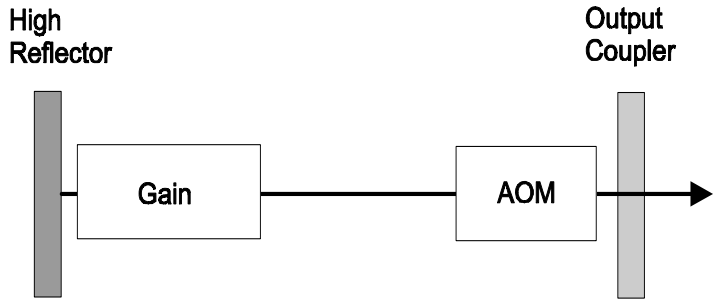

Figure 7.32: Actively modelocked laser with an amplitude modulator (Acousto-Optic-Modulator).

The pulse is shaped in the resonator by the finite bandwidth of the gain and the loss modulator, which periodically varies the intracavity loss according to $q(t) = M\left(1 - \cos(\omega_M t)\right)$. The modulation frequency has to be precisely tuned to the resonator round-trip time, $\omega_M = 2\pi/T_R$, see Fig.7.32. The mode locking process is then described by the master equation for the slowly varying pulse envelope

$$T_R \frac{\partial A}{\partial T} = \left[g(T) + D_g \frac{\partial^2}{\partial t^2} - l - M\left(1 - \cos(\omega_M t)\right) \right] A. \tag{7.68}$$

This equation can be interpreted as the total pulse shaping due to gain, loss and modulator within one roundtrip, see Fig.7.33.

If we fix the gain in Eq. (7.68) at its stationary value, what ever it might be, Eq.(7.68) is a linear p.d.e, which can be solved by separation of variables. The pulses, we expect, will have a width much shorter than the round-trip time T_R. They will be located in the minimum of the loss modulation where the cosine-function can be approximated by a parabola and we obtain

$$T_R \frac{\partial A}{\partial T} = \left[g - l + D_g \frac{\partial^2}{\partial t^2} - M_s t^2 \right] A. \tag{7.69}$$

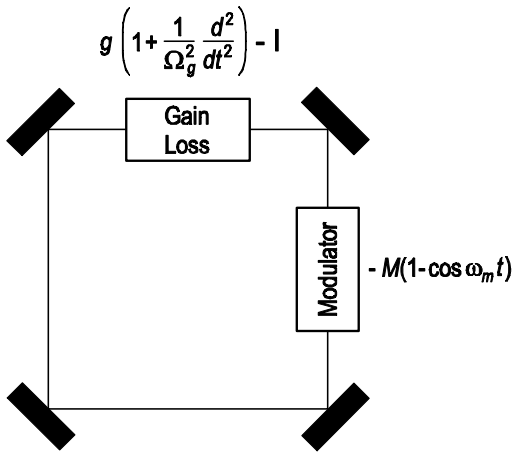

Figure 7.33: Schematic representation of the master equation for an actively mode-locked laser.

M_s is the modulation strength, and corresponds to the curvature of the loss modulation in the time domain at the minimum loss point

$$D_g = \frac{g}{\Omega_g^2}, \tag{7.70}$$

$$M_s = \frac{M\omega_M^2}{2}. \tag{7.71}$$

The differential operator on the right side of (7.69) corresponds to the Schrödinger-Operator of the harmonic oscillator problem. Therefore, the eigen functions of this operator are the Hermite-Gaussians

$$A_n(T,t) = A_n(t)e^{\lambda_n T/T_R}, \tag{7.72}$$

$$A_n(t) = \sqrt{\frac{W_n}{2^n\sqrt{\pi}n!\tau_a}}H_n(t/\tau_a)e^{-\frac{t^2}{2\tau_a^2}}, \tag{7.73}$$

where τ_a determines the pulse width of the Gaussian pulse. The width is given by the fourth root of the ratio between gain dispersion and modulator strength

$$\tau_a = \sqrt[4]{D_g/M_s}. \tag{7.74}$$

Note, from Eq. (7.72) we see that the gain per round-trip of each eigenmode

is given by λ_n (or in general the real part of λ_n), which is given by

$$\lambda_n = g_n - l - 2M_s\tau_a^2(n + \frac{1}{2}). \tag{7.75}$$

The corresponding saturated gain for each eigen solution is given by

$$g_n = \frac{1}{1 + \frac{W_n}{P_L T_R}}, \tag{7.76}$$

where W_n is the energy of the corresponding solution and $P_L = E_L/\tau_L$ the saturation power of the gain. Eq. (7.75) shows that for a given g the eigen solution with $n = 0$, the ground mode, has the largest gain per roundtrip. Thus, if there is initially a field distribution which is a superpostion of all eigen solutions, the ground mode will grow fastest and will saturate the gain to a value

$$g_s = l + M_s\tau_a^2. \tag{7.77}$$

such that $\lambda_0 = 0$ and consequently all other modes will decay since $\lambda_n < 0$ for $n \geq 1$. This also proves the stability of the ground mode solution [17]. Thus active modelocking without detuning between resonator round-trip time and modulator period leads to Gaussian steady state pulses with a FWHM pulse width

$$\Delta t_{FWHM} = 2\ln 2\tau_a = 1.66\tau_a. \tag{7.78}$$

The spectrum of the Gaussian pulse is given by

$$\tilde{A}_0(\omega) = \int_{-\infty}^{\infty} A_0(t)e^{i\omega t}dt \tag{7.79}$$

$$= \sqrt{\sqrt{\pi}W_n\tau_a}e^{-\frac{(\omega\tau_a)^2}{2}}, \tag{7.80}$$

and its FWHM is

$$\Delta f_{FWHM} = \frac{1.66}{2\pi\tau_a}. \tag{7.81}$$

Therfore, the time-bandwidth product of the Gaussian is

$$\Delta t_{FWHM} \cdot \Delta f_{FWHM} = 0.44. \tag{7.82}$$

The stationary pulse shape of the modelocked laser is due to the parabolic loss modulation (pulse shortening) in the time domain and the parabolic filtering (pulse stretching) due to the gain in the frequency domain, see Figs.

7.34 and 7.35. The stationary pulse is achieved when both effects balance. Since external modulation is limited to electronic speed and the pulse width does only scale with the inverse square root of the gain bandwidth actively modelocking typically only results in pulse width in the range of 10-100ps.

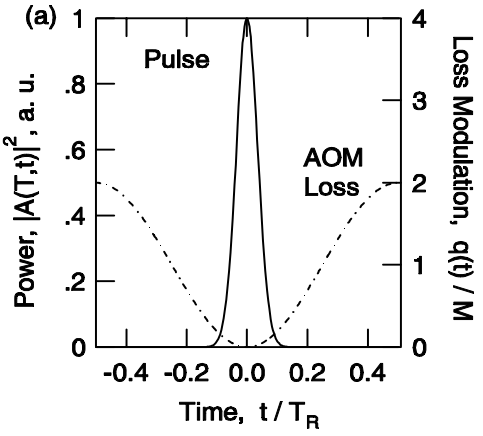

Figure 7.34: Loss modulation leads to pulse shortening in each roundtrip

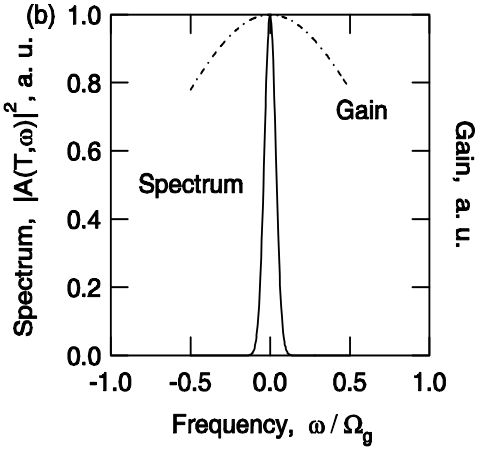

Figure 7.35: The finite gain bandwidth broadens the pulse in each roundtrip. For a certain pulse width there is balance between the two processes.

For example: Nd:YAG; $2l = 2g = 10\%$, $\Omega_g = \pi\Delta f_{FWHM} = 0.65$ THz,

$M = 0.2$, $f_m = 100$ MHz, $D_g = 0.24$ ps^2, $M_s = 4 \cdot 10^{16} s^{-1}$, $\tau_p \approx 99$ ps.

With the pulse width (7.74), Eq.(7.77) can be rewritten in several ways

$$g_s = l + M_s \tau_a^2 = l + \frac{D_g}{\tau_a^2} = l + \frac{1}{2} M_s \tau_a^2 + \frac{1}{2} \frac{D_g}{\tau_a^2}, \tag{7.83}$$

which means that in steady state the saturated gain is lifted above the loss level l, so that many modes in the laser are maintained above threshold. There is additional gain necessary to overcome the loss of the modulator due to the finite temporal width of the pulse and the gain filter due to the finite bandwidth of the pulse. Usually

$$\frac{g_s - l}{l} = \frac{M_s \tau_a^2}{l} \ll 1, \tag{7.84}$$

since the pulses are much shorter than the round-trip time. The stationary pulse energy can therefore be computed from

$$g_s = \frac{1}{1 + \frac{W_s}{P_L T_R}} = l. \tag{7.85}$$

The name modelocking originates from studying this pulse formation process in the frequency domain. Note, the term

$$-M \left[1 - \cos(\omega_M t)\right] A$$

generates sidebands on each cavity mode present according to

$$\begin{aligned} & -M \left[1 - \cos(\omega_M t)\right] \exp(j\omega_{n_0} t) \\ = \; & -M \left[\exp(j\omega_{n_0} t) - \frac{1}{2} \exp(j(\omega_{n_0} t - \omega_M t)) - \frac{1}{2} \exp(j(\omega_{n_0} t + \omega_M t))\right] \\ = \; & M \left[-\exp(j\omega_{n_0} t) + \frac{1}{2} \exp(j\omega_{n_0 - 1} t) + \frac{1}{2} \exp(j\omega_{n_0 + 1} t)\right] \end{aligned}$$

if the modulation frequency is the same as the cavity round-trip frequency. The sidebands generated from each running mode is injected into the neighboring modes which leads to synchronisation and locking of neighboring modes, i.e. mode-locking, see Fig.7.36

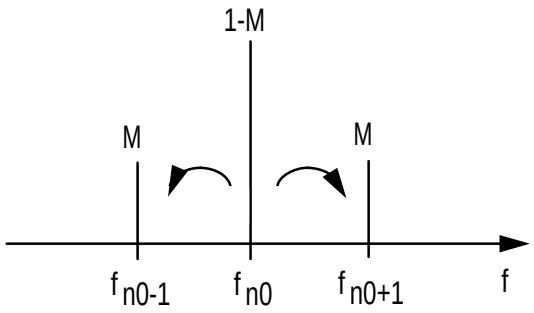

Figure 7.36: Modelocking in the frequency domain: The modulator transvers energy from each mode to its neighboring mode, thereby redistributing energy from the center to the wings of the spectrum. This process seeds and injection locks neighboring modes.

7.10.2 Passive Mode Locking

Electronic loss modulation is limited to electronic speeds. Therefore, the curvature of the loss modulation that determines the pulse length is limited. It is desirable that the pulse itself modulates the loss. The shorter the pulse the sharper the loss modulation and eventually much shorter pulses can be reached.

The dynamics of a laser modelocked with a fast saturable absorber can also be easily understood using the master equation (7.68) and replacing the loss due to the modulator with the loss from a saturable absorber. For a fast saturable absorber, the losses q react instantaneously on the intensity or power $P(t) = |A(t)|^2$ of the field

$$q(A) = \frac{q_0}{1 + \frac{|A|^2}{P_A}}, \tag{7.86}$$

where P_A is the saturation power of the absorber. Such absorbers can be made out of semiconductor materials or nonlinear optical effects can be used to create artificial saturable absorption, such as in Kerr lens mode locking.

There is no analytic solution of the master equation (7.68) when the loss modulation is replaced by the absorber response (7.86). We can however make expansions on the absorber response to get analytic insight. If the absorber is not saturated, we can expand the response (7.86) for small

intensities

$$q(A) = q_0 - \gamma|A|^2, \tag{7.87}$$

with the saturable absorber modulation coefficient $\gamma = q_0/P_A$. The constant nonsaturated loss q_0 can be absorbed in the losses $l_0 = l + q_0$. The resulting master equation is

$$T_R \frac{\partial A(T,t)}{\partial T} = \left[g - l_0 + D_f \frac{\partial^2}{\partial t^2} + \gamma|A|^2 \right] A(T,t). \tag{7.88}$$

Up to the imaginary unit, this equation is similar to a type of nonlinear Schroedinger Equation, i.e. the potential depends on the wave function itself. One finds as a possible stationary solution

$$A_s(T,t) = A_s(t) = A_0 \text{sech}\left(\frac{t}{\tau}\right). \tag{7.89}$$

Note, there is

$$\frac{d}{dx}\text{sech}x = -\tanh x \, \text{sech}x, \tag{7.90}$$

$$\begin{aligned} \frac{d^2}{dx^2}\text{sech}x &= \tanh^2 x \, \text{sech}x - \text{sech}^3 x, \\ &= \left(\text{sech}x - 2\,\text{sech}^3 x\right). \end{aligned} \tag{7.91}$$

Substitution of the solution (7.89) into the master equation (7.88), and assuming steady state, results in

$$\begin{aligned} 0 = &\left[(g - l_0) + \frac{D_f}{\tau^2}\left[1 - 2\text{sech}^2\left(\frac{t}{\tau}\right)\right]\right. \\ &\left. + \gamma|A_0|^2 \text{sech}^2\left(\frac{t}{\tau}\right)\right] \cdot A_0 \text{sech}\left(\frac{t}{\tau}\right). \end{aligned} \tag{7.92}$$

Comparison of the coefficients with the sech- and sech^3-expressions leads to a condition for the peak pulse intensity and pulse width, τ, and for the saturated gain

$$\frac{D_f}{\tau^2} = \frac{1}{2}\gamma|A_0|^2, \tag{7.93}$$

$$g = l_0 - \frac{D_f}{\tau^2}. \tag{7.94}$$

From Eq.(7.93) and with the pulse energy of a sech pulse

$$W = \int_{-\infty}^{+\infty} 2|A_s(t)|^2 dt = 2|A_0|^2\tau, \tag{7.95}$$

follows

$$\tau = \frac{4D_f}{\gamma W}. \tag{7.96}$$

$$g_s(W) = \frac{g_0}{1 + \frac{W}{P_L T_R}} \tag{7.97}$$

Equation (7.94) together with (7.96) determines the pulse energy

$$\begin{aligned} g_s(W) &= \frac{g_0}{1 + \frac{W}{P_L T_R}} = l_0 - \frac{D_f}{\tau^2} \\ &= l_0 - \frac{(\gamma W)^2}{16 D_g} \end{aligned} \tag{7.98}$$

Figure 7.37 shows the time dependent variation of gain and loss in a laser modelocked with a fast saturable absorber on a normalized time scale.

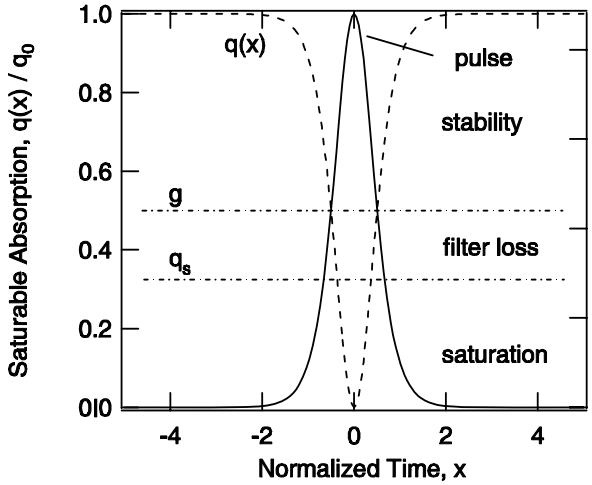

Figure 7.37: Gain and loss in a passively modelocked laser using a fast saturable absorber on a normalized time scale $x = t/\tau$. The absorber is assumed to saturate linearly with intensity according to $q(A) = q_0\left(1 - \frac{|A|^2}{A_0^2}\right)$.

Here, we assumed that the absorber saturates linearly with intensity up to a maximum value $q_0 = \gamma A_0^2$. If this maximum saturable absorption is completely exploited, see Figure 7.38.

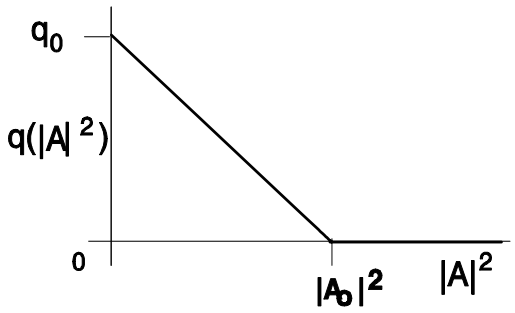

Figure 7.38: Saturation characteristic of an ideal saturable absorber

The minimum pulse width achievable with a given saturable absorption q_0 results from Eq.(7.93)

$$\frac{D_f}{\tau^2} = \frac{q_0}{2}, \tag{7.99}$$

to be

$$\tau = \sqrt{\frac{2}{q_0}} \frac{1}{\Omega_f}. \tag{7.100}$$

Note that in contrast to active modelocking the achievable pulse width is now scaling with the inverse gain bandwidth. This gives much shorter pulses. Figure 7.37 can be interpreted as follows: In steady state, the saturated gain is below loss, by about one half of the exploited saturable loss before and after the pulse. This means that there is net loss outside the pulse, which keeps the pulse stable against growth of instabilities at the leading and trailing edge of the pulse. If there is stable mode-locked operation, there must always be net loss far away from the pulse, otherwise, a continuous wave signal running at the peak of the gain would experience more gain than the pulse and would break through. From Eq.(7.93) it follows, that one third of the exploited saturable loss is used up during saturation of the aborber and actually only one sixth is used to overcome the filter losses due to the finite gain bandwidth. Note, there is a limit to the mimium pulse width. This limit is due to the saturated gain (7.94), $g_s = l + \frac{1}{2}q_0$. Therefore, from

Eq.(7.100), if we assume that the finite bandwidth of the laser is set by the gain, i.e. $D_f = D_g = \frac{g}{\Omega_g^2}$, we obtain for $q_0 \gg l$

$$\tau_{\min} = \frac{1}{\Omega_g} \tag{7.101}$$

for the linearly saturating absorber model. This corresponds to mode locking over the full bandwidth of the gain medium, as for a sech-shaped pulse, the time-bandwidth product is 0.315, and therefore,

$$\Delta f_{FWHM} = \frac{0.315}{1.76 \cdot \tau_{\min}} = \frac{\Omega_g}{1.76 \cdot \pi}. \tag{7.102}$$

As an example, for the Ti:sapphire laser this corresponds to $\Omega_g = 240$ THz, $\tau_{\min} = 3.7$ fs, $\tau_{FWHM} = 6.5$ fs, which is in good agreement with experimentally observed results [19].

This concludes the introdcution into ultrashort pulse generation by mode locking.

Bibliography

[1] R. W. Hellwarth, Eds., Advances in Quantum Electronics, Columbia Press, New York (1961).

[2] A. E. Siegman, "Lasers," University Science Books, Mill Valley, California (1986).

[3] O. Svelto, "Principles of Lasers," Plenum Press, NY 1998.

[4] http://www.llnl.gov/nif/library/aboutlasers/how.html

[5] T. H. Maimann, "Stimulated optical radiation in ruby", Nature **187**, 493-494, (1960).

[6] B.E.A. Saleh and M.C. Teich, "Fundamentals of Photonics," John Wiley and Sons, Inc., 1991.

[7] M. J. Weber, "Handbook of Lasers", CRC Press, 2000.

[8] A. Javan, W. R. Bennett and D. H. Herriott, "Population Inversion and Continuous Optical Maser Oscillation in a Gas Discharge Containing a He-Ne Mixture," Phys. Rev. Lett. **6,** (1961).

[9] W. R. Bennett, Applied Optics, Supplement 1, Optical Masers, 24, (1962).

[10] W. Koechner, "Solid-State Lasers," Springer Verlag (1990).

[11] R F Kazarinov and R A Suris, "Amplification of electromagnetic waves in a semiconductor superlattice," Sov. Phys. Semicond. **5**, p. 707 (1971).

[12] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, A. Y. Cho, "Quantum cascade laser," Science **264**, p. 553 (1994).

[13] B. S. Williams, H. Callebaut, S. Kumar, Q. Hu and J. L. Reno, "3.4-THz quantum cascade laser based on longitudinal-optical-phonon scattering for depopulation," Appl. Phys. Lett. **82**, p. 1015-1017 (2003).

[14] D. Kopf, F. X. Kärtner, K. J. Weingarten, M. Kamp, and U. Keller, "Diode-pumped mode-locked Nd:glass lasers with an antiresonant Fabry-Perot saturable absorber," Optics Letters **20**, p. 1169 (1995).

[15] J. J. Zayhowski, C. Dill, "Diode-pumped passively Q-switched picosecond microchip lasers," Opt. Lett. **19**, pp. 1427 – 1429 (1994).

[16] D. J. Kuizenga and A. E. Siegman, "FM and AM modelocking of the homogeneous laser - part I: theory," IEEE J. Qunat. Electron. **6**, pp. 694 – 701 (1970).

[17] H. A. Haus, "A Theory of Forced Mode Locking", IEEE Journal of Quantum Electronics **QE-11**, pp. 323 - 330 (1975).

[18] H. A. Haus, "Theory of modelocking with a fast saturable absorber," J. Appl. Phys. **46**, pp. 3049 – 3058 (1975).

[19] R. Ell, U. Morgner, F.X. Kärtner, J.G. Fujimoto, E.P. Ippen, V. Scheuer, G. Angelow, T. Tschudi: Generation of 5-fs pulses and octave-spanning spectra directly from a Ti:Sappire laser, Opt. Lett. **26**, 373-375 (2001)