

Fundamentals of Modern Optics

Winter Term 2022/2023

Prof. Thomas Pertsch
 Abbe School of Photonics, Friedrich-Schiller-Universität Jena

Table of content

1. Introduction	4
2. Optical fields in dispersive and isotropic media	8
2.1 Maxwell's equations	8
2.1.1 Adaption to optics	8
2.1.2 Temporal dependence of the fields	12
2.1.3 Maxwell's equations in Fourier domain	13
2.1.4 From Maxwell's equations to the wave equation.....	13
2.1.5 Decoupling of the vectorial wave equation	15
2.2 Optical properties of matter.....	16
2.2.1 Basics.....	16
2.2.2 Types of considered light-matter interactions (covered by lecture Structure of Matter).....	19
2.2.3 Dielectric polarization and susceptibility (covered by lecture Structure of Matter)	20
2.2.4 Conductive current and conductivity (covered by lecture Structure of Matter) ...	22
2.2.5 Generalized complex dielectric function.....	23
2.2.6 Material models in time domain	27
2.3 Poynting vector and energy balance.....	29
2.3.1 Time averaged Poynting vector.....	29
2.3.2 Time averaged energy balance	31
2.4 Kramers-Kronig relation (covered by lecture Structure of Matter).....	33
3. Solution of wave equations for homogeneous isotropic media	37
3.1 Normal modes in homogeneous isotropic media.....	37
3.1.1 Transverse waves (nonzero epsilon)	38
3.1.2 Longitudinal waves (epsilon equal zero)	40
3.1.3 Plane wave solutions in different frequency regimes	40
3.1.4 Time averaged Poynting vector of plane waves	47
3.2 Beams and pulses as well as the analogy of diffraction and dispersion.....	48
3.3 Diffraction of monochromatic beams in homogeneous isotropic media	49
3.3.1 Arbitrarily narrow beams (general case).....	50
3.3.2 Fresnel- (paraxial) approximation	58
3.3.3 Paraxial wave equation.....	62
3.4 Propagation of Gaussian beams.....	64
3.4.1 Propagation in paraxial approximation.....	65
3.4.2 Propagation of Gaussian beams with q-parameter formalism.....	71
3.4.3 Gaussian optics	72
3.4.4 Gaussian modes in a resonator	75
3.5 Dispersion of pulses in homogeneous isotropic media	81
3.5.1 Pulses with finite transverse width (pulsed beams)	81
3.5.2 Pulses with infinite transverse extension (pulse propagation).....	89
3.5.3 Analogy of diffraction and dispersion	90
3.5.4 Propagation of a Gaussian pulse without chirp	91

3.5.5	Propagation of a chirped Gaussian pulse	95
4.	Diffraction theory	100
4.1	Interaction with plane masks	100
4.2	Propagation using different approximations	101
4.2.1	General case - small aperture	101
4.2.2	Fresnel approximation (paraxial approximation).....	102
4.2.3	Paraxial Fraunhofer approximation (far field approximation).....	102
4.2.4	Non-paraxial Fraunhofer approximation	104
4.3	Paraxial Fraunhofer diffraction at plane masks	104
4.4	Remarks on Fresnel diffraction	109
5.	Fourier optics - optical filtering	111
5.1	Imaging of arbitrary optical fields with a thin lens.....	111
5.1.1	Transfer function of a thin lens	111
5.1.2	Optical imaging using the 2f-setup	112
5.2	Optical filtering and image processing.....	114
5.2.1	4f-setup.....	114
5.2.2	Examples of aperture functions	117
5.2.3	Optical resolution	118
6.	The polarization of electromagnetic waves.....	121
6.1	Polarization of normal modes in isotropic media	121
6.2	Polarization states	122
7.	Principles of optics in crystals.....	124
7.1	Susceptibility and dielectric tensor	124
7.2	Optical classification of crystals.....	126
7.3	Index ellipsoid	127
7.4	Normal modes in anisotropic media	128
7.4.1	Normal modes propagating in principal directions	129
7.4.2	Normal modes for arbitrary propagation direction	130
7.4.3	Normal surfaces of normal modes.....	136
7.4.4	Special case: uniaxial crystals	137
8.	Optical fields in isotropic, dispersive and piecewise homogeneous media	141
8.1	Basics.....	141
8.1.1	Definition of the problem.....	141
8.1.2	Decoupling of the vectorial wave equation	142
8.1.3	Interfaces and symmetries	143
8.1.4	Transition conditions	143
8.2	Fields in a layer system → matrix method.....	144
8.2.1	Fields in one homogeneous layer.....	144
8.2.2	Fields in a system of layers	147
8.3	Reflection – transmission problem for layer systems.....	149
8.3.1	General layer systems	149
8.3.2	Single interface	155
8.3.3	Periodic multi-layer systems – Bragg mirrors – 1D photonic crystals	164
8.3.4	Fabry-Perot resonators.....	171
8.4	Guided waves in layer systems.....	177
8.4.1	Field structure of guided waves	177
8.4.2	Dispersion relation for guided waves.....	179
8.4.3	Guided waves at interface - surface polariton.....	181
8.4.4	Guided waves in a layer – film waveguide	184
8.4.5	Excitation of guided waves	187

9. Ray optics - geometrical optics (covered by lecture Introduction to Optical Modeling).....	190
9.1 Introduction	190
9.2 Postulates.....	190
9.3 Simple rules for propagation of light.....	191
9.4 Simple optical components	191
9.5 Ray tracing in inhomogeneous media (graded-index - GRIN optics).....	195
9.5.1 Ray equation	195
9.5.2 Eikonal equation	197
9.6 Matrix optics	198
9.6.1 Ray-transfer-matrix.....	198
9.6.2 Matrices of optical elements.....	198
9.6.3 Cascaded elements.....	199

This script originates from the lecture series “Theoretische Optik” given by Falk LEDERER in the physics program at the Friedrich Schiller University Jena (Germany) for many years between 1990 and 2012. Later the script was adapted by Stefan SKUPIN and Thomas PERTSCH for the international education program in photonics at the Abbe School of Photonics, Friedrich Schiller University Jena (Germany).

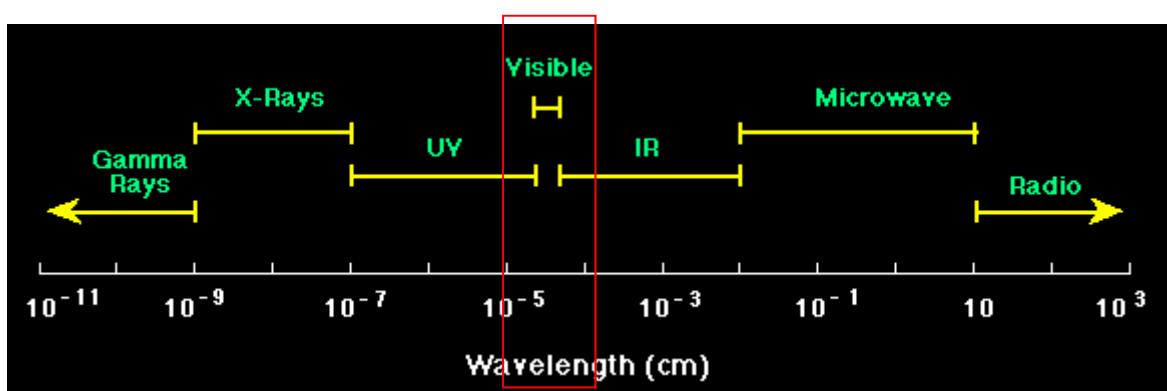
1. Introduction

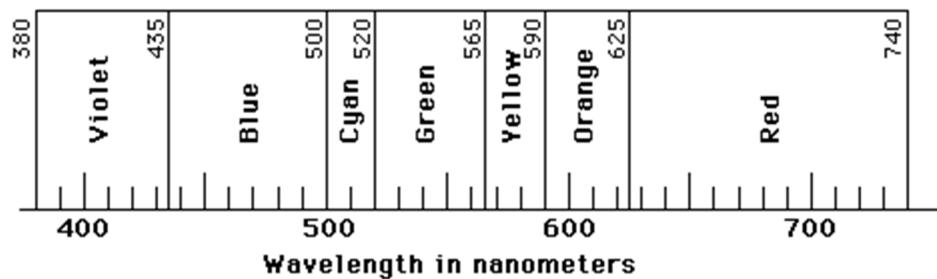
- 'optique' (Greek) → lore of light → 'what is light'?
- Is light a wave or a particle (photon)?
- Light is one of the requirements for life → photosynthesis
- 90% of information we receive is visual

A) What is light?

- electromagnetic wave propagating with the speed of $c = 3 \times 10^8 \text{ m / s}$
- wave = evolution of
 - amplitude and phase → complex description
 - polarization → vectorial field description
 - coherence → statistical description

Spectrum of Electromagnetic Radiation				
Region	Wavelength [nm]	Wavelength [m] ($\text{nm} = 10^{-9} \text{ m}$)	Frequency [Hz] ($\text{THz} = 10^{12} \text{ Hz}$)	Energy [eV]
Radio	$> 10^8$	$> 10^{-1}$	$< 3 \times 10^9$	$< 10^{-5}$
Microwave	$10^8 - 10^5$	$10^{-1} - 10^{-4}$	$3 \times 10^9 - 3 \times 10^{12}$	$10^{-5} - 0.01$
Infrared	$10^5 - 700$	$10^{-4} - 7 \times 10^{-7}$	$3 \times 10^{12} - 4.3 \times 10^{14}$	$0.01 - 2$
Visible	$700 - 400$	$7 \times 10^{-7} - 4 \times 10^{-7}$	$4.3 \times 10^{14} - 7.5 \times 10^{14}$	$2 - 3$
Ultraviolet	$400 - 1$	$4 \times 10^{-7} - 10^{-9}$	$7.5 \times 10^{14} - 3 \times 10^{17}$	$3 - 10^3$
X-Rays	$1 - 0.01$	$10^{-9} - 10^{-11}$	$3 \times 10^{17} - 3 \times 10^{19}$	$10^3 - 10^5$
Gamma Rays	< 0.01	$< 10^{-11}$	$> 3 \times 10^{19}$	$> 10^5$





B) Origin of light

- atomic system → determines original properties of light (e.g. statistics, frequency, line width)
- laser → control of emission properties of matter at the origin of emission
 - laser: artificial light source with new and unmatched properties (e.g. **coherent, directed, focused, monochromatic**)
 - invention in 1958 (A. L. Schawlow & C. H. Townes, Phys. Rev.: very important development)
- optical system → modifies properties of light (e.g. intensity, duration, ...)

C) Propagation of light through matter

- light-matter interaction (G: Licht-Materie-Wechselwirkung)

effect	dispersion	diffraction	absorption	scattering
governed by	↓ frequency spectrum	↓ spatial frequency	↓ center of frequency	↓ wavelength spectrum

- matter is the medium of propagation → the properties of the medium (natural or artificial) determine the propagation of light
- light provides a powerful technique for studying matter (spectroscopy) → measurement methods (interferometer)
- design media with the desired properties: glasses, polymers, semiconductors, compounded media (effective media, photonic crystals, metamaterials)

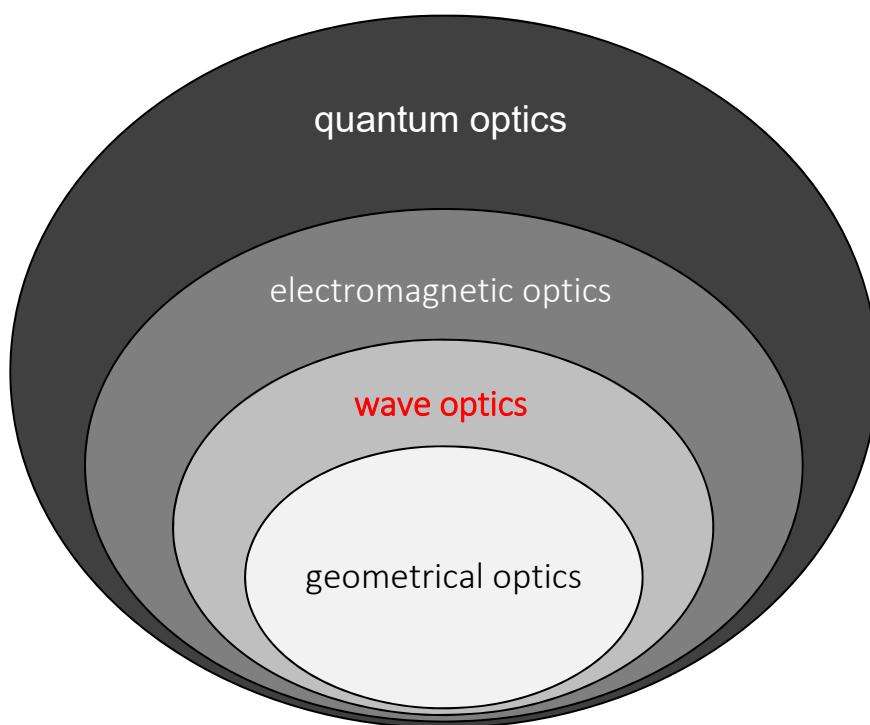
D) Light can modify matter

- light induces physical, chemical and biological processes
- used for lithography, material processing, or modification of biological objects (bio-photonics)

E) Prominent fields of application

- Optical telecommunication
 - transmitting data (Terabit/s in one fiber) over transatlantic distances
- Optics in medicine and life sciences
- Light sensors and light sources
 - new light sources to reduce energy consumption
 - new projection techniques
- Micro- and nano-optics
 - ultra-small camera
- Relativistic optics

J) Schematic of optics



- geometrical optics
 - $\lambda \ll$ size of objects → daily experiences
 - optical instruments, optical imaging
 - intensity, direction, coherence, phase, polarization, photons
(G: Intensität, Richtung, Kohärenz, Phase, Polarisation, Photon)
- wave optics
 - $\lambda \approx$ size of objects → interference, diffraction, dispersion, coherence
 - laser, holography, resolution, pulse propagation
 - intensity, direction, coherence, phase, polarization, photons

- electromagnetic optics
 - reflection, transmission, guided waves, resonators
 - laser, integrated optics, photonic crystals, Bragg mirrors ...
 - intensity, direction, coherence, phase, polarization, photons
- quantum optics
 - small number of photons, fluctuations, light-matter interaction
 - quantum cryptography, quantum computation, teleportation ...
 - intensity, direction, coherence, phase, polarization, photons
- in this lecture
 - electromagnetic optics and wave optics
 - no quantum optics → subject of advanced lectures

K) Literature

- Fundamental
 1. Saleh, Teich, 'Fundamentals of Photonics', Wiley (1992)
in German: "Grundlagen der Photonik" Wiley (2008)
 2. Hecht, 'Optic', Addison-Wesley (2001)
in German: "Optik", Oldenbourg (2005)
 3. Mansuripur, 'Classical Optics and its Applications', Cambridge (2002)
 4. Menzel, 'Photonics', Springer (2000)
 5. Lipson, Lipson, Tannhäuser, 'Optik'; Springer (1997)
 6. Born, Wolf, 'Principles of Optics', Pergamon
 7. Sommerfeld, 'Optik'
- Advanced
 1. W. Silvast, 'Laser Fundamentals',
 2. Agrawal, 'Fiber-Optic Communication Systems', Wiley
 3. Band, 'Light and Matter', Wiley, 2006
 4. Karthe, Müller, 'Integrierte Optik', Teubner
 5. Diels, Rudolph, 'Ultrashort Laser Pulse Phenomena', Academic
 6. Yariv, 'Optical Electronics in modern Communications', Oxford
 7. Snyder, Love, 'Optical Waveguide Theory', Chapman&Hall
 8. Römer, 'Theoretical Optics', Wiley, 2005.

2. Optical fields in dispersive and isotropic media

2.1 Maxwell's equations

Our general starting point is the set of Maxwell's equations. They are the basis of the electromagnetic approach to optics, which is developed in this lecture.

2.1.1 Adaption to optics

The notation of Maxwell's equations is different for different disciplines of science and engineering, which rely on these equations to describe electromagnetic phenomena at different frequency ranges. Even though Maxwell's equations are valid for all frequencies, the physics of light-matter interaction is different for different frequencies. Since light-matter interaction must be included in the Maxwell's equations to solve them consistently, different ways have been established on how to express Maxwell's equations in different frequency ranges. Here we follow a notation which was established for convenience at frequencies close to that of visible light.

Maxwell's equations (macroscopic)

In a rigorous way the electromagnetic theory is developed starting from the properties of electromagnetic fields in vacuum. In vacuum one could write down Maxwell's equations in their so-called pure microscopic form, which includes the interaction with any kind of matter based on the consideration of point charges. Obviously this is inadequate for the description of light in condensed matter, since the number of point charges, which would need to be taken into account to describe a macroscopic object, would exceed all imaginable computational resources.

To solve this problem one uses an averaging procedure which summarizes the influence of many point charges on the electromagnetic field in a homogeneously distributed response of the solid state on the excitation by light. In turn, the electromagnetic fields are also averaged over some adequate volume. For optics, this procedure is justified since any kind of available experimental detector could not resolve the very fine spatial details of the fields between the point charges, e.g. ions or electrons, which are lost in this averaging procedure.

These averaged electromagnetic equations have been rigorously derived in a number of fundamental textbooks on electrodynamic theory. Here we will not redo this derivation. We will instead start directly from the averaged Maxwell's equations.

$\text{rot } \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{B}(\mathbf{r}, t)}{\partial t}$	$\text{div } \mathbf{D}(\mathbf{r}, t) = \rho_{\text{ext}}(\mathbf{r}, t)$
$\text{rot } \mathbf{H}(\mathbf{r}, t) = \mathbf{j}_{\text{makr}}(\mathbf{r}, t) + \frac{\partial \mathbf{D}(\mathbf{r}, t)}{\partial t}$	$\text{div } \mathbf{B}(\mathbf{r}, t) = 0$

– electric field (G ¹ : elektrisches Feld)	E(r,t)	[V/m]
– magnetic flux density or magnetic induction (G: magnetische Flussdichte oder magnetische Induktion)	B(r,t)	[Vs/m ²] or [tesla]
– electric flux density or electric displacement field (G: elektrische Flussdichte oder dielektrische Verschiebung)	D(r,t)	[As/m ²]
– magnetic field (G: magnetisches Feld)	H(r,t)	[A/m]
– external charge density	$\rho_{\text{ext}}(\mathbf{r},t)$	[As/m ³]
– macroscopic current density	j_{makr}(r,t)	[A/m ²]

Auxiliary fields

The "cost" of introducing these macroscopic Maxwell's equations is the occurrence of two additional fields, the dielectric flux density **D(r,t)** and the magnetic field **H(r,t)**. These two fields are related to the electric field **E(r,t)** and magnetic flux density **B(r,t)** by two other new fields, **P(r,t)** and **M(r,t)** as

$$\boxed{\begin{aligned}\mathbf{D}(\mathbf{r},t) &= \epsilon_0 \mathbf{E}(\mathbf{r},t) + \mathbf{P}(\mathbf{r},t) \\ \mathbf{H}(\mathbf{r},t) &= \frac{1}{\mu_0} [\mathbf{B}(\mathbf{r},t) - \mathbf{M}(\mathbf{r},t)]\end{aligned}}$$

- dielectric polarization
(G: dielektrische Polarisation) **P(r,t)** [As/m²],
- magnetic polarization
or magnetization
(G: Magnetisierung) **M(r,t)** [Vs/m²]
- electric constant
or vacuum permittivity
(G: Vakuumpermittivität) $\epsilon_0 \approx 8.854 \times 10^{-12}$ As/Vm
- magnetic constant
or vacuum permeability
(G: Vakuumpermeabilität) $\mu_0 = 4\pi \times 10^{-7}$ Vs/Am
- Here the electric constant and the magnetic constant are connected to the speed of light in vacuum c by

$$\epsilon_0 = \frac{1}{\mu_0 c^2}$$

¹ Marked by "G:", the German translation of important terms will be given.

Light matter interaction

In order to solve this set of equations, i.e. Maxwell's equations and auxiliary field equations, one needs to relate the dielectric flux density $\mathbf{D}(\mathbf{r},t)$ and the magnetic field $\mathbf{H}(\mathbf{r},t)$ to the electric field $\mathbf{E}(\mathbf{r},t)$ and the magnetic flux density $\mathbf{B}(\mathbf{r},t)$. This is achieved by modeling the material properties by introducing the material equations.

- The **effect on the medium** gives rise to polarization $\mathbf{P}(\mathbf{r},t) = f[\mathbf{E}]$ and magnetization $\mathbf{M}(\mathbf{r},t) = f[\mathbf{B}]$. → In order to solve Maxwell's equations we need material models, which describe these quantities.
- In optics at visible wavelength, we generally deal with non-magnetizable media. Hence we can assume $\mathbf{M}(\mathbf{r},t) = 0$. Exceptions to this general property are metamaterials, which might possess some artificial effective magnetization properties resulting in $\mathbf{M}(\mathbf{r},t) \neq 0$.

Furthermore we need to introduce sources of the fields into our model. This is achieved by the so-called source terms, which are inhomogeneities and hence they define the unique solutions of Maxwell's equations.

- **free** charge density (G: Dichte freier Ladungsträger)

$$\rho_{\text{ext}}(\mathbf{r},t) \quad [\text{As}/\text{m}^3]$$

- **macroscopic** current density (G: makroskopische Stromdichte) consisting of two contributions

$$\mathbf{j}_{\text{makr}}(\mathbf{r},t) = \mathbf{j}_{\text{cond}}(\mathbf{r},t) + \mathbf{j}_{\text{conv}}(\mathbf{r},t) \quad [\text{A}/\text{m}^2]$$

- *conductive current density* (G: Konduktionsstromdichte)

$$\mathbf{j}_{\text{cond}}(\mathbf{r},t) = f[\mathbf{E}]$$

- *convective current density* (G: Konvektionsstromdichte)

$$\mathbf{j}_{\text{conv}}(\mathbf{r},t) = \rho_{\text{ext}}(\mathbf{r},t) \mathbf{v}(\mathbf{r},t)$$

- In optics, we generally have no free charges which change their local density in time at rates comparable to the oscillation frequency of the electromagnetic fields of light (several Terahertz):

$$\frac{\partial \rho_{\text{ext}}(\mathbf{r},t)}{\partial t} \approx 0 \rightarrow \mathbf{j}_{\text{conv}}(\mathbf{r},t) = 0$$

- With the above simplifications, we can formulate Maxwell's equations in the context of optics:

$\text{rot } \mathbf{E}(\mathbf{r},t) = -\mu_0 \frac{\partial \mathbf{H}(\mathbf{r},t)}{\partial t}$	$\epsilon_0 \text{div } \mathbf{E}(\mathbf{r},t) = -\text{div } \mathbf{P}(\mathbf{r},t)$
$\text{rot } \mathbf{H}(\mathbf{r},t) = \mathbf{j}(\mathbf{r},t) + \frac{\partial \mathbf{P}(\mathbf{r},t)}{\partial t} + \epsilon_0 \frac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t}$	$\text{div } \mathbf{H}(\mathbf{r},t) = 0$

- In optics, the medium (or more precisely the mathematical material model) determines the dependence of the induced polarization on the electric field $\mathbf{P}(\mathbf{E})$ and the dependence of the induced (conductive) current density on the electric field $\mathbf{j}(\mathbf{E})$.
- Once we have specified these relations, we can solve Maxwell's equations consistently.

Example:

- In vacuum, both polarization \mathbf{P} and current density \mathbf{j} are zero (simplest material model). Hence we can solve Maxwell's equations directly.

Remarks:

- for linear response of matter: principle of superposition applies → electromagnetic effects at different frequencies coexist without influencing each other
- Even though the Maxwell's equations, in the way they have been written above, are derived to describe light, i.e. electromagnetic fields at optical frequencies, they are simultaneously valid for other frequency ranges as well. Furthermore, since Maxwell's equations are linear equations as long as the material does not introduce any nonlinearity, the principle of superposition holds. Hence we can decompose the total electromagnetic fields into components of different frequency ranges. In turn this means that we do not have to take care of any slow electromagnetic phenomena, e.g. electrostatics or radio wave, in our formulation of Maxwell's equations since they can be split off from our optical problem and can be treated separately.
- We can define a bound charge density (G: Dichte gebundener Ladungsträger) as the source of spatially varying polarization \mathbf{P} .

$$\rho_b(\mathbf{r}, t) = -\mathbf{div} \mathbf{P}(\mathbf{r}, t)$$

- Analogously, we can define a bound current density (G: Stromdichte gebundener Ladungsträger) as the source of the temporal variation in the polarization \mathbf{P} .

$$\mathbf{j}_b(\mathbf{r}, t) = \frac{\partial \mathbf{P}(\mathbf{r}, t)}{\partial t}$$

- This essentially means that we can describe the same physics in two different ways since currents are in principle moving charges. Thus we can use either $\mathbf{j}_b(\mathbf{r}, t)$ or $\partial \mathbf{P}(\mathbf{r}, t) / \partial t$ (see generalized complex dielectric function below).

Complex field formalism (G: komplexer Feld-Formalismus)

- Maxwell's equations are also valid for complex fields and are easier to solve this way.

- This fact can be exploited to simplify calculations, because it is easier to deal with complex exponential functions ($\exp(ix)$) than with trigonometric functions $\cos(x)$ and $\sin(x)$.
- Hence we use the following convention in this lecture to distinguish between the two types of fields.

real physical field: $\mathbf{E}_r(\mathbf{r}, t)$

complex mathematical representation: $\mathbf{E}(\mathbf{r}, t)$

- In this lecture we define their relation as

$$\mathbf{E}_r(\mathbf{r}, t) = \frac{1}{2} [\mathbf{E}(\mathbf{r}, t) + \mathbf{E}^*(\mathbf{r}, t)] = \text{Re}[\mathbf{E}(\mathbf{r}, t)]$$

However, this relation can be defined differently in different textbooks.

- This means in general: for calculations we use the complex fields [$\mathbf{E}(\mathbf{r}, t)$] and for physical results we go back to real fields by simply omitting the imaginary part. This works because Maxwell's equations are linear and no multiplications of fields occurs.
- Therefore, be careful when multiplications of fields are required in the description of the material response or the field dynamics. In this case you would have to go back to real quantities before you perform such multiplications. This becomes relevant in e.g. the calculation of the Poynting vector, as can be seen in the following chapters.

2.1.2 Temporal dependence of the fields

When it comes to time dependence of the electromagnetic field, we can distinguish two different types of light:

A) monochromatic light (stationary fields)

- harmonic dependence on temporal coordinate
- $\sim \exp(-i\omega t) \rightarrow$ phase is fixed \rightarrow coherent, infinite wave train e.g.:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}) \exp(-i\omega t)$$

- Monochromatic light approximates very well the typical output of a continuous wave (CW) laser. Once we know the frequency we only have to compute the spatial dependence of the (stationary) fields.

B) polychromatic light (non-stationary fields)

- finite wave train
- With the help of Fourier transformation we can decompose the fields into infinite wave trains and use all the results from case A) (see also next section). Based on this concept, arbitrary temporal dynamics of the electromagnetic fields can be described by decomposing it into monochromatic light fields (so-

called Fourier decomposition), which can later be recomposed to obtain again the entire dynamics.

$$\mathbf{E}(\mathbf{r},t) = \int_{-\infty}^{\infty} \bar{\mathbf{E}}(\mathbf{r},\omega) \exp(-i\omega t) d\omega$$

$$\bar{\mathbf{E}}(\mathbf{r},\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r},t) \exp(i\omega t) dt$$

Remark: The position of the sign in the exponent and the factor $1/2\pi$ can be defined differently in different textbooks. The bar and the ω in $\bar{\mathbf{E}}(\mathbf{r},\omega)$ indicate the frequency domain fields.

2.1.3 Maxwell's equations in Fourier domain

In order to solve Maxwell's equations more easily, we would like to introduce a Fourier decomposition of the fields directly in the Maxwell's equations. The complexity of Maxwell's equations arises from the mixing of different derivatives with respect to space and time. The Fourier decomposition of the fields allows us to perform some or all of these derivatives analytically and hence to considerably simplify the remaining equations.

For this purpose, we need to find out how a derivative of a dynamic variable can be calculated in Fourier space. Here we will do this for the temporal derivatives but later in the course we will also apply this concept to spatial derivatives.

A simple rule for the transformation of a time derivative into Fourier space can be obtained using integration by parts:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} dt \left[\frac{\partial}{\partial t} \mathbf{E}(\mathbf{r},t) \right] \exp(i\omega t) = -i\omega \frac{1}{2\pi} \int_{-\infty}^{+\infty} dt \mathbf{E}(\mathbf{r},t) \exp(i\omega t) = -i\omega \bar{\mathbf{E}}(\mathbf{r},\omega)$$

Thus, a time derivative in real space transforms into a simple product with $-i\omega$ in Fourier space.

$$\rightarrow \text{rule: } \frac{\partial}{\partial t} \xrightarrow{\text{FT}} -i\omega \quad \text{under the condition } \int_{-\infty}^{+\infty} |\mathbf{E}(\mathbf{r},t)|^2 dt < \infty$$

Now we can write Maxwell's equations in the Fourier domain:

$$\begin{aligned} \text{rot } \bar{\mathbf{E}}(\mathbf{r},\omega) &= i\omega \mu_0 \bar{\mathbf{H}}(\mathbf{r},\omega) & \varepsilon_0 \text{div } \bar{\mathbf{E}}(\mathbf{r},\omega) &= -\text{div } \bar{\mathbf{P}}(\mathbf{r},\omega) \\ \text{rot } \bar{\mathbf{H}}(\mathbf{r},\omega) &= \bar{\mathbf{j}}(\mathbf{r},\omega) - i\omega \bar{\mathbf{P}}(\mathbf{r},\omega) - i\omega \varepsilon_0 \bar{\mathbf{E}}(\mathbf{r},\omega) & \text{div } \bar{\mathbf{H}}(\mathbf{r},\omega) &= 0 \end{aligned}$$

2.1.4 From Maxwell's equations to the wave equation

Maxwell's equations provide the basis for deriving all possible mathematical solutions to electromagnetic problems. However, very often we are just interested in the radiation fields, which can be described more easily by an adapted equation, the so-called wave equation. From Maxwell's equations, it is straightforward to derive the wave equation by using the two curl equations.

A) Time domain derivation

We start by applying the curl operator (rot) a second time on $\text{rot E}(\mathbf{r}, t) = \dots$ and substitute rot H with the other curl equation from Maxwell's equations.

$$\text{rot rot E}(\mathbf{r}, t) = -\mu_0 \text{rot} \frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} = -\mu_0 \frac{\partial}{\partial t} \left[\mathbf{j}(\mathbf{r}, t) + \frac{\partial \mathbf{P}(\mathbf{r}, t)}{\partial t} + \epsilon_0 \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} \right]$$

We find the wave equation for the electric field as

$$\text{rot rot E}(\mathbf{r}, t) + \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \mathbf{E}(\mathbf{r}, t) = -\mu_0 \frac{\partial \mathbf{j}(\mathbf{r}, t)}{\partial t} - \mu_0 \frac{\partial^2 \mathbf{P}(\mathbf{r}, t)}{\partial t^2}$$

The blue terms on the right hand side require knowledge of the material model. Additionally, we have to make sure that all other Maxwell's equations are fulfilled, in particular that for the divergence of the electric field:

$$\text{div} [\epsilon_0 \mathbf{E}(\mathbf{r}, t) + \mathbf{P}(\mathbf{r}, t)] = 0$$

Once we have solved the wave equation, we know the electric field. From that, we can easily compute the magnetic field:

$$\frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} = -\frac{1}{\mu_0} \text{rot E}(\mathbf{r}, t)$$

Remarks:

- An analogous procedure is also possible for the magnetic field \mathbf{H} , i.e. we can derive a wave equation for the magnetic field as well.
- Normally, the wave equation for the electric field \mathbf{E} is more convenient because the material model defines $\mathbf{P}(\mathbf{E})$.
- However, for inhomogeneous media the wave equation for \mathbf{H} can sometimes be the better choice for the numerical solution of the partial differential equation since it forms a Hermitian operator.
- analogous procedure possible for \mathbf{H} instead of \mathbf{E}
- generally, wave equation for \mathbf{E} is more convenient, because $\mathbf{P}(\mathbf{E})$ given
- for inhomogeneous media \mathbf{H} can sometimes be a better choice

B) Frequency domain derivation

We can apply the same procedure to also derive the wave equation directly in the Fourier domain and find

$$\text{rot rot } \bar{\mathbf{E}}(\mathbf{r}, \omega) - \frac{\omega^2}{c^2} \bar{\mathbf{E}}(\mathbf{r}, \omega) = i\omega \mu_0 \bar{\mathbf{j}}(\mathbf{r}, \omega) + \mu_0 \omega^2 \bar{\mathbf{P}}(\mathbf{r}, \omega)$$

and

$$\text{div} [\epsilon_0 \bar{\mathbf{E}}(\mathbf{r}, \omega) + \bar{\mathbf{P}}(\mathbf{r}, \omega)] = 0$$

- magnetic field:

$$\boxed{\bar{\mathbf{H}}(\mathbf{r}, \omega) = -\frac{\mathbf{i}}{\omega \mu_0} \operatorname{rot} \bar{\mathbf{E}}(\mathbf{r}, \omega)}$$

- transferring the results from the Fourier domain to the time domain:

- for stationary fields: take solution and multiply by $e^{-i\omega t}$.
- for non-stationary fields (and linear media) → inverse Fourier transformation (principle of superposition)

$$\mathbf{E}(\mathbf{r}, t) = \int_{-\infty}^{\infty} \bar{\mathbf{E}}(\mathbf{r}, \omega) \exp(-i\omega t) d\omega$$

2.1.5 Decoupling of the vectorial wave equation

So far we have seen that for the general problem of electromagnetic waves, all 3 vectorial field components of the electric or the magnetic field are coupled. Hence we have to solve a vectorial wave equation for the general problem. However, it would be desirable to instead express problems by a scalar equation since they are much easier to solve. For problems with **translational invariance** in at least one direction, e.g. for homogeneous infinite media, layers or interfaces, this can be achieved since the vectorial components of the fields can be decoupled.

Let's assume invariance in the y -direction and propagation only in the x - z -plane. All spatial derivatives along the y -direction must then disappear ($\partial/\partial y = 0$) and the operators in the wave equation simplify.

$$\operatorname{rot} \operatorname{rot} \bar{\mathbf{E}} = \operatorname{grad} \operatorname{div} \bar{\mathbf{E}} - \Delta \bar{\mathbf{E}} = \begin{bmatrix} \frac{\partial}{\partial x} \left(\frac{\partial \bar{E}_x}{\partial x} + \frac{\partial \bar{E}_z}{\partial z} \right) \\ 0 \\ \frac{\partial}{\partial z} \left(\frac{\partial \bar{E}_x}{\partial x} + \frac{\partial \bar{E}_z}{\partial z} \right) \end{bmatrix} - \begin{bmatrix} \Delta^{(2)} \bar{E}_x \\ \Delta^{(2)} \bar{E}_y \\ \Delta^{(2)} \bar{E}_z \end{bmatrix}$$

The decoupling becomes visible when the three components of the general vectorial field are decomposed in the following way:

- decomposition of electric field

$$\bar{\mathbf{E}} = \bar{\mathbf{E}}_{\perp} + \bar{\mathbf{E}}_{\parallel} \quad \rightarrow \quad \bar{\mathbf{E}}_{\perp} = \begin{pmatrix} 0 \\ \bar{E}_y \\ 0 \end{pmatrix}, \quad \bar{\mathbf{E}}_{\parallel} = \begin{pmatrix} \bar{E}_x \\ 0 \\ \bar{E}_z \end{pmatrix}$$

$$\text{with nabla operator } \nabla^{(2)} = \begin{pmatrix} \partial/\partial x \\ 0 \\ \partial/\partial z \end{pmatrix}, \text{ and Laplacian } \Delta^{(2)} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}$$

Hence we obtain two wave equations for the $\bar{\mathbf{E}}_{\perp}$ and $\bar{\mathbf{E}}_{\parallel}$ fields.

- gives two decoupled wave equations

$$\Delta^{(2)} \bar{\mathbf{E}}_{\perp}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \bar{\mathbf{E}}_{\perp}(\mathbf{r}, \omega) = -i\omega\mu_0 \bar{\mathbf{j}}_{\perp}(\mathbf{r}, \omega) - \mu_0\omega^2 \bar{\mathbf{P}}_{\perp}(\mathbf{r}, \omega)$$

$$\Delta^{(2)} \bar{\mathbf{E}}_{\parallel}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \bar{\mathbf{E}}_{\parallel}(\mathbf{r}, \omega) - \mathbf{grad}^{(2)} \cdot \mathbf{div}^{(2)} \bar{\mathbf{E}}_{\parallel} = -i\omega\mu_0 \bar{\mathbf{j}}_{\parallel}(\mathbf{r}, \omega) - \mu_0\omega^2 \bar{\mathbf{P}}_{\parallel}(\mathbf{r}, \omega)$$

These two wave equations are independent as long as the material response, which is expressed by \mathbf{j} and \mathbf{P} , does not couple the respective field components by some anisotropic response.

Properties

- propagation of perpendicularly polarized fields $\bar{\mathbf{E}}_{\perp}$ and $\bar{\mathbf{E}}_{\parallel}$ can be treated separately
- propagation of $\bar{\mathbf{E}}_{\perp}$ is described by scalar equation
- similarly the other two E-field components can be described by a scalar equation for $\bar{\mathbf{H}}_{\perp}$, from which $\bar{\mathbf{E}}_{\parallel}$ can be computed
- alternative notations used in some books:
 - $\perp \rightarrow s \rightarrow \text{TE}$ (transverse electric)
 - $\parallel \rightarrow p \rightarrow \text{TM}$ (transverse magnetic)

2.2 Optical properties of matter

In this chapter, we will derive a simple material model for the polarization and the current density. The basic idea is to write down an equation of motion for a single charged particle and assume that all other particles of the same type behave the same way. More precisely, we will use a driven harmonic oscillator model to describe the motion of bound charges giving rise to a polarization of the medium. For free charges we will use the same model but without the restoring force, leading to a current density. In the literature, this simple approach is often called the Drude-Lorentz model (named after Paul Drude and Hendrik Antoon Lorentz).

2.2.1 Basics

We are looking for $\mathbf{P}(\mathbf{E})$ and $\mathbf{j}(\mathbf{E})$. In general, this leads to a many body problem in solid state theory, which is rather complex. However, in many cases phenomenological models are sufficient to describe the necessary phenomena. As already pointed out above, we use the simplest approach, the so-called **Drude-Lorentz model** for free or bound charge carriers (electrons).

- assume an ensemble of uncoupled, driven, and damped harmonic oscillators
- **free** charge carriers: metals and excited semiconductors (intraband transitions)
- **bound** charge carriers: dielectric media and semiconductors (interband transitions)

- The Drude-Lorentz model creates a link between cause (electric field) and effect (induced polarization or current). Because the resulting relations $\mathbf{P}(\mathbf{E})$ and $\mathbf{j}(\mathbf{E})$ are linear (no \mathbf{E}^2 dependence etc.), we can use **linear response theory**.

Basics of linear response theory

For the polarization $\mathbf{P}(\mathbf{E})$ (very similar for $\mathbf{j}(\mathbf{E})$):

- description in both time and frequency domain possible
- In time domain: we introduce the **response function** (*G: Responsfunktion*)
 $\mathbf{E}(\mathbf{r}, t) \rightarrow$ acts on medium (response function) \rightarrow medium reacts by being polarized $\mathbf{P}(\mathbf{r}, t)$

$$P_i(\mathbf{r}, t) = \epsilon_0 \sum_j \int_{-\infty}^t R_{ij}(\mathbf{r}, t - t') E_j(\mathbf{r}, t') dt' \quad (\text{convolution integral})$$

with $\hat{\mathbf{R}}$ in the convolution integral being a 2nd rank tensor
 $i = x, y, z$ and summing over $j = x, y, z$

- In frequency domain: we introduce the **susceptibility** (*G: Suszeptibilität*) as the transfer function
 $\bar{\mathbf{E}}(\mathbf{r}, \omega) \rightarrow$ medium (transfer function / susceptibility) $\rightarrow \bar{\mathbf{P}}(\mathbf{r}, \omega)$

$$\bar{P}_i(\mathbf{r}, \omega) = \epsilon_0 \sum_j \chi_{ij}(\mathbf{r}, \omega) \bar{E}_j(\mathbf{r}, \omega)$$

- response function and transfer function (susceptibility) are linked via Fourier transform (convolution theorem)

$$R_{ij}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi_{ij}(\omega) \exp(-i\omega t) d\omega$$

Note the factor of $1/2\pi$ in addition to our usual definition of the Fourier transform arises due to the definitions of the response function, susceptibility and the convolution theorem.

- Obviously, things look more manageable in frequency domain. Using the wave equation from before and assuming that there are no currents ($\mathbf{j}=0$), we find

$$\text{rotrot } \bar{\mathbf{E}}(\mathbf{r}, \omega) - \frac{\omega^2}{c^2} \bar{\mathbf{E}}(\mathbf{r}, \omega) = \mu_0 \omega^2 \bar{\mathbf{P}}(\mathbf{r}, \omega)$$

or

$$\Delta \bar{\mathbf{E}}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \bar{\mathbf{E}}(\mathbf{r}, \omega) - \mathbf{grad} \operatorname{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = -\mu_0 \omega^2 \bar{\mathbf{P}}(\mathbf{r}, \omega)$$

- and for auxiliary fields

$$\bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \bar{\mathbf{E}}(\mathbf{r}, \omega) + \bar{\mathbf{P}}(\mathbf{r}, \omega)$$

The general response function and the corresponding susceptibility given above simplifies for certain properties of the medium:

Simplification of the wave equation for different types of media

A) linear, homogenous, isotropic, non-dispersive media (simplest but unphysical case)

- homogeneous (position-independent) $\rightarrow \chi_{ij}(\mathbf{r}, \omega) = \chi_{ij}(\omega)$
- isotropic (direction-independent)
 $\rightarrow \chi_{ij}(\mathbf{r}, \omega) = \chi(\mathbf{r}, \omega) \delta_{ij}$ (proportional to identity matrix/tensor)
- non-dispersive (frequency-independent) $\rightarrow \chi_{ij}(\mathbf{r}, \omega) = \chi_{ij}(\mathbf{r})$
 \rightarrow instantaneous response (by Fourier transform): $R_{ij}(\mathbf{r}, t) = \chi_{ij}(\mathbf{r}) \delta(t)$
(Attention: This is unphysical!)
- $\rightarrow \chi_{ij}(\mathbf{r}, \omega) \rightarrow \chi$ is a scalar constant

frequency domain	time domain description
$\bar{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \chi \bar{\mathbf{E}}(\mathbf{r}, \omega)$	$\leftrightarrow \mathbf{P}(\mathbf{r}, t) = \epsilon_0 \chi \mathbf{E}(\mathbf{r}, t)$ (unphysical!)
$\bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \epsilon \bar{\mathbf{E}}(\mathbf{r}, \omega)$	$\leftrightarrow \mathbf{D}(\mathbf{r}, t) = \epsilon_0 \epsilon \mathbf{E}(\mathbf{r}, t)$
with $\epsilon = 1 + \chi$	

Maxwell: $\mathbf{div} \bar{\mathbf{D}} = 0 \rightarrow \mathbf{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0$ for $\epsilon(\omega) \neq 0$

$$\boxed{\Delta \bar{\mathbf{E}}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \epsilon \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0} \leftrightarrow \boxed{\Delta \mathbf{E}(\mathbf{r}, t) - \frac{\epsilon}{c^2} \frac{\partial^2}{\partial t^2} \mathbf{E}(\mathbf{r}, t) = 0}$$

- approximation is valid only for a certain frequency range, because all media are dispersive
- based on an unphysical material model

B) linear, homogeneous, isotropic, *dispersive* media $\rightarrow \chi(\omega)$

$$\left. \begin{array}{l} \bar{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \chi(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) \\ \bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) \end{array} \right\} \epsilon(\omega) = 1 + \chi(\omega)$$

$\mathbf{div} \bar{\mathbf{D}}(\mathbf{r}, \omega) = 0 \curvearrowright \mathbf{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0$ for $\epsilon(\omega) \neq 0$

$$\rightarrow \boxed{\Delta \bar{\mathbf{E}}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \epsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0} \quad \text{Helmholtz equation}$$

- This description is sufficient for many materials.

C) linear, *inhomogeneous*, isotropic, dispersive media $\rightarrow \chi(\mathbf{r}, \omega)$

$$\bar{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \chi(\mathbf{r}, \omega) \bar{\mathbf{E}}(\mathbf{r}, \omega),$$

$$\bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\mathbf{r}, \omega) \bar{\mathbf{E}}(\mathbf{r}, \omega).$$

$$\operatorname{div} \bar{\mathbf{D}}(\mathbf{r}, \omega) = 0$$

$$\operatorname{div} \bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\mathbf{r}, \omega) \operatorname{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) + \epsilon_0 \bar{\mathbf{E}}(\mathbf{r}, \omega) \cdot \operatorname{grad} \epsilon(\mathbf{r}, \omega) = 0,$$

$$\rightarrow \operatorname{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = -\frac{\operatorname{grad} \epsilon(\mathbf{r}, \omega)}{\epsilon(\mathbf{r}, \omega)} \cdot \bar{\mathbf{E}}(\mathbf{r}, \omega).$$

$$\Delta \bar{\mathbf{E}}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \epsilon(\mathbf{r}, \omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) = -\operatorname{grad} \left\{ \frac{\operatorname{grad} \epsilon(\mathbf{r}, \omega) \cdot \bar{\mathbf{E}}(\mathbf{r}, \omega)}{\epsilon(\mathbf{r}, \omega)} \right\}$$

- All field components couple.
- This equation is also valid for case B) with $\operatorname{grad} \epsilon(\mathbf{r}, \omega) = 0$.

D) linear, homogeneous, **anisotropic**, dispersive media $\rightarrow \chi_{ij}(\omega)$

$$\bar{P}_i(\mathbf{r}, \omega) = \epsilon_0 \sum_j \chi_{ij}(\omega) \bar{E}_j(\mathbf{r}, \omega)$$

$$\bar{D}_i(\mathbf{r}, \omega) = \epsilon_0 \sum_j \epsilon_{ij}(\omega) \bar{E}_j(\mathbf{r}, \omega).$$

- This is the most complicated case for a medium with linear response.
- See chapter on crystal optics.

A similar classification of material responses could be made based on the conductivity $\sigma(\mathbf{r}, \omega)$ with the current density $\mathbf{j}(\mathbf{r}, \omega) = \sigma(\mathbf{r}, \omega) \mathbf{E}(\mathbf{r}, \omega)$.

Before we start writing down actual material model equations, let us define which types of light matter interactions we would like to consider.

2.2.2 Types of considered light-matter interactions (covered by lecture Structure of Matter)

I) Interaction of light with bound electrons and the lattice

The contributions of **bound electrons** and **lattice vibrations** in dielectrics and semiconductors give rise to the polarization \mathbf{P} . The **lattice vibrations (phonons)** are the ionic part of the material model. Because of the large mass of the ions ($10^3 \times$ mass of electron), the resulting oscillation frequencies will be small. Generally speaking, phonons are responsible for thermal properties of the medium. However, some phonon modes may contribute to optical properties, but they have weak dispersion (weak dependence on frequency ω).

Fully understanding the **electronic transitions** of bound electrons requires quantum theoretical treatment, which allows for an accurate computation of the transition frequencies. However, a classical (phenomenological) treatment of the oscillation of bound electrons is possible and useful.

II) Interaction of light with free electrons

The contribution of **free electrons** in metals and excited semiconductors gives rise to a current density \mathbf{j} . We assume a so-called (interaction-)free electron gas, where the electron charges are neutralized by the background ions. Only collisions with ions and related damping of electronic motion will be considered.

We will look at the contributions from bound electrons / lattice vibrations and free electrons separately. Later we will join the results in a generalized material model which holds for many common optical materials..

2.2.3 Dielectric polarization and susceptibility (covered by lecture Structure of Matter)

Let us first focus on bound charges (ions, electrons). In the **Drude-Lorentz model**, the electric field $\mathbf{E}(\mathbf{r},t)$ gives rise to a displacement $\mathbf{s}(\mathbf{r},t)$ of charged particles from their equilibrium positions. In the simplest approach, this can be modeled by a **driven harmonic oscillator**:

$$\frac{\partial^2}{\partial t^2}\mathbf{s}(\mathbf{r},t) + g\frac{\partial}{\partial t}\mathbf{s}(\mathbf{r},t) + \omega_0^2\mathbf{s}(\mathbf{r},t) = \frac{q}{m}\mathbf{E}(\mathbf{r},t)$$

- **resonance frequency** (electronic transition) $\rightarrow \omega_0$
- **damping coefficient** $\rightarrow g$
- **charge** $\rightarrow q$
- **mass** $\rightarrow m$

The induced electric dipole moment due to the displacement of charged particles is given by

$$\mathbf{p}(\mathbf{r},t) = q\mathbf{s}(\mathbf{r},t),$$

We further assume that all bound charges of the same type behave identically, i.e., we treat an ensemble of uncoupled, driven, and damped harmonic oscillators. The dipole density (polarization) is then given by multiplying with the number density of dipoles

$$\mathbf{P}(\mathbf{r},t) = N\mathbf{p}(\mathbf{r},t) = Nq\mathbf{s}(\mathbf{r},t)$$

Hence, the governing equation for the polarization $\mathbf{P}(\mathbf{r},t)$ reads

$$\frac{\partial^2}{\partial t^2}\mathbf{P}(\mathbf{r},t) + g\frac{\partial}{\partial t}\mathbf{P}(\mathbf{r},t) + \omega_0^2\mathbf{P}(\mathbf{r},t) = \frac{q^2N}{m}\mathbf{E}(\mathbf{r},t) = \epsilon_0 f \mathbf{E}(\mathbf{r},t)$$

with oscillator strength $f = \frac{1}{\epsilon_0} \frac{e^2 N}{m}$, for $q=-e$ (electrons)

This equation is easy to solve in Fourier domain:

$$-\omega^2 \bar{\mathbf{P}}(\mathbf{r},\omega) - ig\omega \bar{\mathbf{P}}(\mathbf{r},\omega) + \omega_0^2 \bar{\mathbf{P}}(\mathbf{r},\omega) = \epsilon_0 f \bar{\mathbf{E}}(\mathbf{r},\omega)$$

$$\rightarrow \bar{\mathbf{P}}(\mathbf{r}, \omega) = \frac{\epsilon_0 f}{(\omega_0^2 - \omega^2) - i g \omega} \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

$$\text{with } \bar{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \chi(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) \Rightarrow \chi(\omega) = \frac{f}{(\omega_0^2 - \omega^2) - i g \omega}$$

In general we have several different types of oscillators in a medium, i.e., several different resonance frequencies. Nevertheless, since to a good approximation they do not influence each other, all these different oscillators contribute individually to the polarization. Hence the model can be constructed by simply summing up all contributions.

- several resonance frequencies

$$\bar{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \sum_j \left\{ \frac{f_j}{(\omega_{0j}^2 - \omega^2) - i g_j \omega} \right\} \bar{\mathbf{E}}(\mathbf{r}, \omega) = \epsilon_0 \chi(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

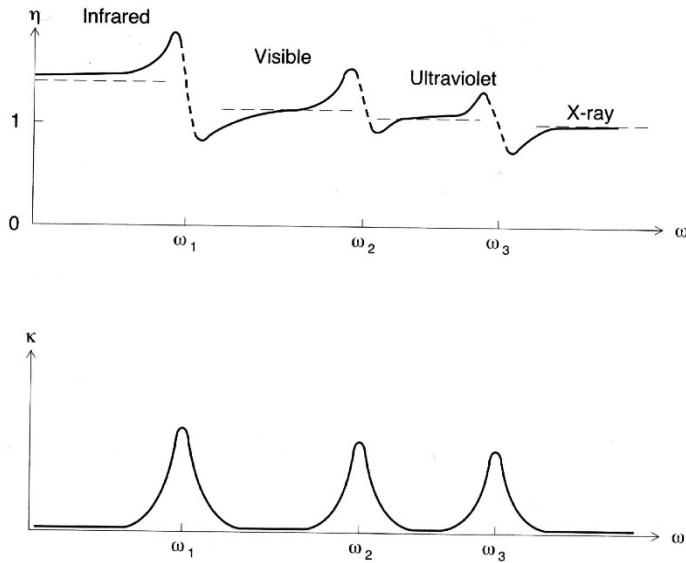
$$\chi(\omega) = \sum_j \left\{ \frac{f_j}{(\omega_{0j}^2 - \omega^2) - i g_j \omega} \right\}$$

- $\chi(\omega)$ is the complex, frequency dependent susceptibility

$$\bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \bar{\mathbf{E}}(\mathbf{r}, \omega) + \epsilon_0 \chi(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

- $\epsilon(\omega)$ is the complex, frequency dependent dielectric function

Example: (plotted for eta and kappa with $\epsilon(\omega) = [\eta(\omega) + i\kappa(\omega)]^2$)



2.2.4 Conductive current and conductivity (covered by lecture Structure of Matter)

Let us now describe the response of a free electron gas which undergoes no interaction with the positively charged background. Again we use the model of a driven harmonic oscillator, but this time with resonance frequency $\omega_0 = 0$. This corresponds to the case of zero restoring force.

$$\frac{\partial^2}{\partial t^2} \mathbf{s}(\mathbf{r}, t) + g \frac{\partial}{\partial t} \mathbf{s}(\mathbf{r}, t) = -\frac{e}{m} \mathbf{E}(\mathbf{r}, t),$$

The resulting induced current density is given by multiplying with the charge density

$$\mathbf{j}(\mathbf{r}, t) = -Ne \frac{\partial}{\partial t} \mathbf{s}(\mathbf{r}, t)$$

and the governing dynamic equation reads

$$\frac{\partial}{\partial t} \mathbf{j}(\mathbf{r}, t) + g \mathbf{j}(\mathbf{r}, t) = \frac{e^2 N}{m} \mathbf{E}(\mathbf{r}, t) = \epsilon_0 \omega_p^2 \mathbf{E}(\mathbf{r}, t)$$

$$\text{with plasma frequency } \omega_p^2 = f = \frac{1}{\epsilon_0} \frac{e^2 N}{m}$$

Again we solve this equation in Fourier domain:

$$-i\omega \bar{\mathbf{j}}(\mathbf{r}, \omega) + g \bar{\mathbf{j}}(\mathbf{r}, \omega) = \epsilon_0 \omega_p^2 \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

$$\rightarrow \bar{\mathbf{j}}(\mathbf{r}, \omega) = \frac{\epsilon_0 \omega_p^2}{g - i\omega} \bar{\mathbf{E}}(\mathbf{r}, \omega) = \sigma(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega).$$

Here we introduced the complex frequency dependent conductivity

$$\sigma(\omega) = \frac{\epsilon_0 \omega_p^2}{g - i\omega} = -i \frac{\epsilon_0 \omega \omega_p^2}{-\omega^2 - ig\omega}.$$

Remarks on plasma frequency

We consider a cloud of electrons and positive ions described by their total charge density ρ and their self-consistent field \mathbf{E} . Then we find according to Maxwell:

$$\epsilon_0 \mathbf{div} \mathbf{E}(\mathbf{r}, t) = \rho(\mathbf{r}, t)$$

For cold electrons, because the total charge is zero, we can use our damped oscillator model from before to describe the current density (only electrons move):

$$\frac{\partial}{\partial t} \mathbf{j} + g\mathbf{j} = \epsilon_0 \omega_p^2 \mathbf{E}(\mathbf{r}, t)$$

Now we apply the divergence operator and substitute from above (red terms):

$$\mathbf{div} \frac{\partial}{\partial t} \mathbf{j} + g \mathbf{div} \mathbf{j} = \epsilon_0 \omega_p^2 \mathbf{div} \mathbf{E}(\mathbf{r}, t) = \omega_p^2 \rho(\mathbf{r}, t)$$

With the continuity equation for the charge density (from Maxwell's equations)

$$\frac{\partial}{\partial t} \rho + \mathbf{div} \mathbf{j} = 0,$$

We can substitute the divergence of the current density and find:

$$-\frac{\partial^2}{\partial t^2} \rho - g \frac{\partial}{\partial t} \rho = \omega_p^2 \rho$$

$$\frac{\partial^2}{\partial t^2} \rho + g \frac{\partial}{\partial t} \rho + \omega_p^2 \rho = 0 \rightarrow \text{harmonic oscillator equation}$$

Hence, the plasma frequency ω_p is the eigen-frequency of such a charge density.

2.2.5 Generalized complex dielectric function

In the above sections, we have derived expressions for both polarization (bound charges) and conductive current density (free charges). Let us now plug our $\bar{\mathbf{j}}(\mathbf{r}, \omega)$ and $\bar{\mathbf{P}}(\mathbf{r}, \omega)$ into the wave equation (in Fourier domain)

$$\text{rotrot } \bar{\mathbf{E}}(\mathbf{r}, \omega) - \frac{\omega^2}{c^2} \bar{\mathbf{E}}(\mathbf{r}, \omega) = \mu_0 \omega^2 \bar{\mathbf{P}}(\mathbf{r}, \omega) + i\omega \mu_0 \bar{\mathbf{j}}(\mathbf{r}, \omega)$$

$$= [\mu_0 \epsilon_0 \omega^2 \chi(\omega) + i\omega \mu_0 \sigma(\omega)] \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

Hence we can collect all terms proportional to $\bar{\mathbf{E}}(\mathbf{r}, \omega)$ and write

$$\text{rotrot } \bar{\mathbf{E}}(\mathbf{r}, \omega) = \frac{\omega^2}{c^2} \left\{ 1 + \chi(\omega) + \frac{i}{\omega \epsilon_0} \sigma(\omega) \right\} \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

$$\text{rotrot } \bar{\mathbf{E}}(\mathbf{r}, \omega) = \frac{\omega^2}{c^2} \varepsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

Here we have introduced the generalized **complex dielectric function**

$$\varepsilon(\omega) = 1 + \chi(\omega) + \frac{\mathbf{i}}{\omega \varepsilon_0} \sigma(\omega) = \varepsilon'(\omega) + \mathbf{i} \varepsilon''(\omega)$$

So, in general we have

$$\varepsilon(\omega) = 1 + \sum_j \left\{ \frac{f_j}{(\omega_{0j}^2 - \omega^2) - \mathbf{i}g_j\omega} \right\} + \frac{\omega_p^2}{-\omega^2 - \mathbf{i}g\omega},$$

since (from before)

$$\chi(\omega) = \sum_j \left\{ \frac{f_j}{(\omega_{0j}^2 - \omega^2) - \mathbf{i}g_j\omega} \right\}, \quad \sigma(\omega) = -\mathbf{i} \frac{\varepsilon_0 \omega \omega_p^2}{-\omega^2 - \mathbf{i}g\omega}.$$

$\varepsilon(\omega)$ contains contributions from vacuum, **phonons (lattice vibrations)**, bound and **free electrons**.

Some special cases for materials in the infrared and visible spectral range:

A) Dielectrics (insulators) in the infrared (IR) spectral range near phonon resonance

If we are interested in dielectrics (insulators) near a phonon resonance in the infrared spectral range, we can simplify the dielectric function as follows:

$$\varepsilon(\omega) = 1 + \sum_j \left\{ \frac{f_j}{(\omega_{0j}^2 - \omega^2) - \mathbf{i}g_j\omega} \right\} + \frac{f}{(\omega_0^2 - \omega^2) - \mathbf{i}g\omega} \text{ with } \omega_0 \ll \omega_{0j} \text{ and } \omega \sim \omega_0$$

Close to a single isolated phonon resonance, it can be approximated by

$$\rightarrow \varepsilon(\omega) = \varepsilon_\infty + \frac{f}{(\omega_0^2 - \omega^2) - \mathbf{i}g\omega}$$

This function has the following parameters:

- resonance frequency: ω_0
- resonance strength: f
- width of resonance peak: g
- epsilon from vacuum and electronic transitions: ε_∞

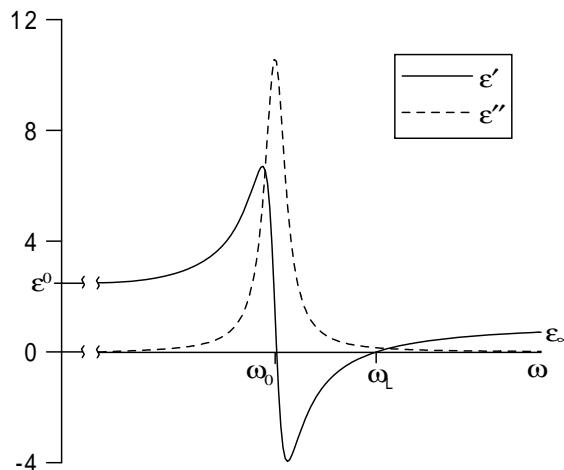
The contribution from electronic transitions shows almost no frequency dependence (dispersion) in this frequency range far away from the electronic resonances. Hence it can be expressed together with the vacuum contribution as a constant ε_∞ .

Let us study the real and the imaginary part of the resulting $\varepsilon(\omega)$ separately:

$$\rightarrow \varepsilon(\omega) = \Re \varepsilon(\omega) + i \Im \varepsilon(\omega) = \varepsilon'(\omega) + i \varepsilon''(\omega)$$

$$\varepsilon'(\omega) = \varepsilon_\infty + \frac{f (\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + g^2 \omega^2},$$

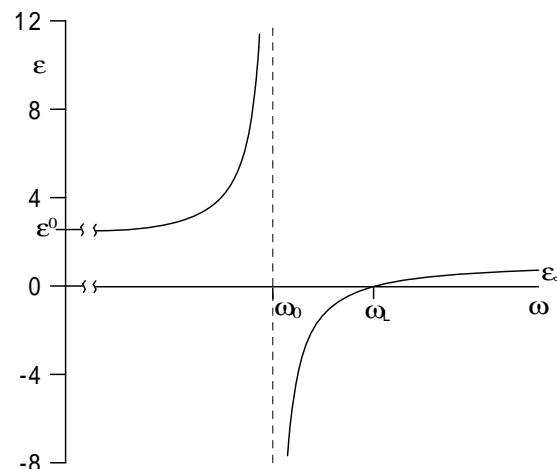
$$\varepsilon''(\omega) = \frac{g f \omega}{(\omega_0^2 - \omega^2)^2 + g^2 \omega^2}. \rightarrow \text{Lorentzian curve}$$



Properties:

- at the so-called **longitudinal** frequency ω_L : $\varepsilon'(\omega = \omega_L) = 0$
- $\varepsilon''(\omega) \neq 0$: absorption and dispersion appear always together
- frequency range with normal dispersion: $\partial \varepsilon'(\omega) / \partial \omega > 0$
- frequency range with anomalous dispersion: $\partial \varepsilon'(\omega) / \partial \omega < 0$
- static dielectric constant in the limit $\omega \rightarrow 0$: $\varepsilon^0 = \varepsilon_\infty + f / \omega_0^2$
- near resonance we find $\varepsilon'(\omega) < 0$ (damping, i.e. decay of field, without absorption if $\varepsilon'' \approx 0$)

Simplified example: sharp resonance for undamped oscillator $g \rightarrow 0$



- relation between resonance frequency ω_0 and longitudinal frequency ω_L
- ($\varepsilon'(\omega = \omega_L) = 0$) (Lyddane-Sachs-Teller relation)

$$\epsilon'(\omega_L) = \epsilon_\infty + \frac{f}{(\omega_0^2 - \omega_L^2)} = 0, f = (\epsilon^0 - \epsilon_\infty) \omega_0^2 \text{ (from above)}$$

$$\Rightarrow \boxed{\omega_L = \omega_0 \sqrt{\frac{\epsilon^0}{\epsilon_\infty}}}$$

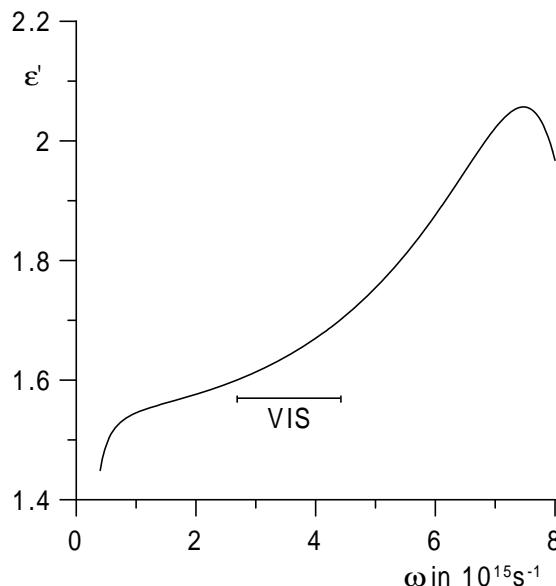
B) Dielectrics in the visible (VIS) spectral range

Dielectric media in the visible (VIS) spectral range can be described by a so-called double resonance model, where a phonon resonance exists in the infrared (IR) at ω_{0p} and an electronic transition exists in the ultraviolet (UV) at ω_{0e} .

$$\boxed{\epsilon(\omega) = \epsilon_\infty + \frac{f_{\text{phonon}}}{(\omega_{0\text{phonon}}^2 - \omega^2) - i g_{\text{phonon}} \omega} + \frac{f_{\text{electron}}}{(\omega_{0\text{electron}}^2 - \omega^2) - i g_{\text{electron}} \omega}},$$

with $\omega_{0\text{phonon}} \ll \omega \ll \omega_{0\text{electron}}$

$\epsilon_\infty \rightarrow$ contribution of vacuum and other (far away) high frequency resonances



The generalization of this approach in the transparent spectral range leads to the so-called **Sellmeier formula**.

$$\boxed{\epsilon'(\omega) - 1 = \sum_j \frac{\bar{f}_j \omega_{0j}^2}{(\omega_{0j}^2 - \omega^2)}}$$

- with j being the number of resonances taken into account
- describes many media very well (dispersion of absorption is neglected)
- oscillator strengths and resonance frequencies are often fit parameters to match experimental data

C) Metals in the visible spectral range

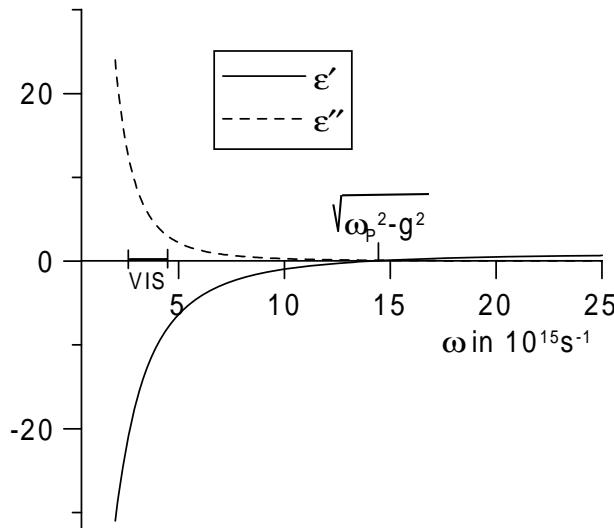
If we want to describe metals in the visible spectral range we find

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + ig\omega} \quad \text{with } \omega_p \gg \omega$$

where ω_p is the so-called plasma frequency of the metal (usually in the UV spectral range). Hence the real and imaginary part of $\epsilon(\omega)$ of a metal is

$$\epsilon'(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + g^2}, \quad \epsilon''(\omega) = \frac{g\omega_p^2}{\omega(\omega^2 + g^2)}.$$

Metals show a large negative real part of the dielectric function $\epsilon'(\omega)$, which gives rise to decay of the fields. This results in reflection of light at metallic surfaces.



2.2.6 Material models in time domain

Let us now transform the results of our material models back into the time domain. In Fourier domain we found for homogeneous and isotropic media:

$$\bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

$$\bar{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \chi(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega).$$

The response function (or Green's function) $R(t)$ in the time domain is then given by

$$R(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(\omega) \exp(-i\omega t) d\omega \quad \chi(\omega) = \int_{-\infty}^{\infty} R(t) \exp(i\omega t) dt$$

and the polarization is determined by the convolution integral of the driving electric field with the response function as

- with the convolution integral to determine the polarization

$$\mathbf{P}(\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^{\infty} R(t-t') \mathbf{E}(\mathbf{r}, t') dt'$$

To prove this, we can use the convolution theorem

$$\begin{aligned}\mathbf{P}(\mathbf{r}, t) &= \int_{-\infty}^{\infty} \bar{\mathbf{P}}(\mathbf{r}, \omega) \exp(-i\omega t) d\omega = \epsilon_0 \int_{-\infty}^{\infty} \chi(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) \exp(-i\omega t) d\omega \\ &= \epsilon_0 \int_{-\infty}^{\infty} \chi(\omega) \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{E}(\mathbf{r}, t') \exp(i\omega t') dt' \exp(-i\omega t) d\omega\end{aligned}$$

Now we switch the order of integration and identify the response function R (red terms):

$$\begin{aligned}\mathbf{P}(\mathbf{r}, t) &= \epsilon_0 \int_{-\infty}^{\infty} \underbrace{\frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(\omega) \exp(-i\omega(t-t')) d\omega}_{R(t-t')} \mathbf{E}(\mathbf{r}, t') dt' \\ &= \epsilon_0 \int_{-\infty}^{\infty} R(t-t') \mathbf{E}(\mathbf{r}, t') dt'\end{aligned}$$

For a “delta function” excitation in the electric field, we find the response to be given by the response function (Greens function - GF) itself:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{e} \delta(t - t_0) \rightarrow \mathbf{P}(\mathbf{r}, t) = \epsilon_0 R(t - t_0) \mathbf{e} \rightarrow \text{response function}$$

Examples

A) instantaneous media (unphysical simplification)

- For instantaneous (or non-dispersive) media, which cannot really exist in nature, we would find:

$$R(t) = \chi \delta(t) \quad \rightarrow \quad \mathbf{P}(\mathbf{r}, t) = \epsilon_0 \chi \mathbf{E}(\mathbf{r}, t) \quad (\text{unphysical!})$$

B) dielectrics

$$R_p(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(\omega) \exp(-i\omega t) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{f}{\omega_0^2 - \omega^2 - ig\omega} \exp(-i\omega t) d\omega,$$

- Using the residual theorem we find:

$$R(t) = \begin{cases} \frac{f}{\Omega} \exp\left(-\frac{g}{2}t\right) \sin \Omega t & t \geq 0 \\ 0 & t < 0 \end{cases} \quad \text{with} \quad \Omega = \sqrt{\omega_0^2 - \frac{g^2}{4}}$$

- In other words, the response to a delta function excitation in dielectrics is that of a damped sinusoidal wave, as would be expected for the **Drude-Lorentz model** of bound charges as damped harmonic oscillators.

$$\mathbf{P}(\mathbf{r}, t) = \frac{f}{\Omega} \int_{-\infty}^t \exp\left[-\frac{g}{2}(t-t')\right] \sin[\Omega(t-t')] \mathbf{E}(\mathbf{r}, t') dt'$$

C) metals

$$R_j(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sigma(\omega) \exp(-i\omega t) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\epsilon_0 \omega_p^2}{g - i\omega} \exp(-i\omega t) d\omega,$$

- Using again the residual theorem we find:

$$R(t) = \begin{cases} \frac{\omega_p^2}{g} \exp(-gt) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

- In other words, the response to a delta function excitation in metals is simply that of a exponentially decaying wave, as would be expected for the **Drude-Lorentz model** of free charges with no restoring forces.

$$\mathbf{j}(\mathbf{r}, t) = \epsilon_0 \omega_p^2 \int_{-\infty}^t \exp[-g(t-t')] \mathbf{E}(\mathbf{r}, t') dt'$$

2.3 Poynting vector and energy balance

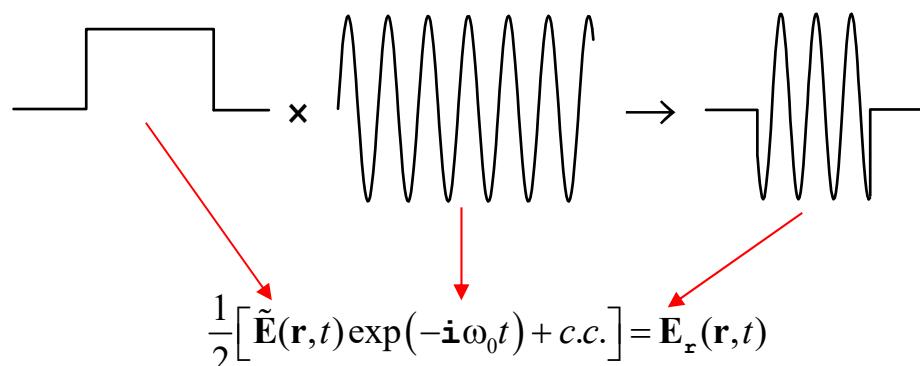
2.3.1 Time averaged Poynting vector

The energy flux of the electromagnetic field is given by the Poynting vector \mathbf{S} (named after its discoverer John Henry Poynting 1884). In practice, we always measure the energy flux through a surface (detector), $\mathbf{S} \cdot \mathbf{n}$, where \mathbf{n} is the normal vector of the surface. To be more precise, the Poynting vector $\mathbf{S}(\mathbf{r}, t) = \mathbf{E}_x(\mathbf{r}, t) \times \mathbf{H}_x(\mathbf{r}, t)$ gives the instantaneous energy flux. Note that we have to use the real electric and magnetic fields, since a product of fields occurs.

In optics, we must consider the following time scales:

- optical cycle: $T_0 = 2\pi/\omega_0 \leq 10^{-14} \text{ s}$
- pulse duration: T_{pulse} in general $T_{\text{pulse}} \gg T_0$
- duration of measurement: T_m in general $T_m \gg T_0$

Hence, in general the detector does not resolve the fast oscillations of the optical field $\sim e^{-i\omega_0 t}$ (optical cycles) and only measures a **time averaged value**. For the situation described above, the electromagnetic fields factorize into slowly varying envelopes and fast carrier oscillations:



The tilde in $\tilde{\mathbf{E}}(\mathbf{r}, t)$ will be used throughout the script to indicates that this is the slowly varying part of the field.

For such pulses, the instantaneous Poynting vector reads:

$$\begin{aligned}
 \mathbf{S}(\mathbf{r}, t) &= \mathbf{E}_\text{r}(\mathbf{r}, t) \times \mathbf{H}_\text{r}(\mathbf{r}, t) \\
 &= \frac{1}{4} \left[\tilde{\mathbf{E}}(\mathbf{r}, t) \times \tilde{\mathbf{H}}^*(\mathbf{r}, t) + \tilde{\mathbf{E}}^*(\mathbf{r}, t) \times \tilde{\mathbf{H}}(\mathbf{r}, t) \right] \\
 &\quad + \frac{1}{4} \left[\tilde{\mathbf{E}}(\mathbf{r}, t) \times \tilde{\mathbf{H}}(\mathbf{r}, t) \exp(-2i\omega_0 t) + \tilde{\mathbf{E}}^*(\mathbf{r}, t) \times \tilde{\mathbf{H}}^*(\mathbf{r}, t) \exp(2i\omega_0 t) \right] \\
 &= \underbrace{\frac{1}{2} \Re \left[\tilde{\mathbf{E}}(\mathbf{r}, t) \times \tilde{\mathbf{H}}^*(\mathbf{r}, t) \right]}_{\text{slow}} + \underbrace{\frac{1}{2} \Re \left[\tilde{\mathbf{E}}(\mathbf{r}, t) \times \tilde{\mathbf{H}}(\mathbf{r}, t) \right] \cos(2\omega_0 t)}_{\text{fast}} \\
 &\quad + \underbrace{\frac{1}{2} \Im \left[\tilde{\mathbf{E}}^*(\mathbf{r}, t) \times \tilde{\mathbf{H}}^*(\mathbf{r}, t) \right] \sin(2\omega_0 t)}_{\text{fast}}.
 \end{aligned}$$

We find that the instantaneous Poynting vector has **slow contributions** which change over time scales of the pulse envelope T_p and **fast contributions** $\sim \cos(2\omega_0 t), \sim \sin(2\omega_0 t)$ changing over time scales of the optical cycle T_0 . Now, a measurement of the Poynting vector over a time interval T_m leads to a time average of $\mathbf{S}(\mathbf{r}, t)$.

$$\langle \mathbf{S}(\mathbf{r}, t) \rangle = \frac{1}{T_m} \int_{t-T_m/2}^{t+T_m/2} \mathbf{S}(\mathbf{r}, t') dt'$$

Since the slowly varying pulse envelope does not change significantly over one optical cycle, the **fast oscillating terms** $\sim \cos 2\omega_0 t$ and $\sim \sin 2\omega_0 t$ cancel in the integration/average. Hence, only the **slow term** contributes to the average Poynting vector.

$$\langle \mathbf{S}(\mathbf{r}, t) \rangle = \frac{1}{2} \frac{1}{T_m} \int_{t-T_m/2}^{t+T_m/2} \Re \left[\tilde{\mathbf{E}}(\mathbf{r}, t') \times \tilde{\mathbf{H}}^*(\mathbf{r}, t') \right] dt'$$

Let us now have a look at the special (but important) case of stationary (monochromatic) fields. Then, the pulse envelope does not depend on time at all (infinitely long pulses).

$$\tilde{\mathbf{E}}(\mathbf{r}, t') = \mathbf{E}(\mathbf{r}), \quad \tilde{\mathbf{H}}(\mathbf{r}, t') = \mathbf{H}(\mathbf{r})$$

$$\boxed{\langle \mathbf{S}(\mathbf{r}, t) \rangle = \frac{1}{2} \Re \left[\mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r}) \right].}$$

This is the definition for the optical intensity $I = |\langle \mathbf{S}(\mathbf{r}, t) \rangle|$. We see that an intensity measurement destroys information on the phase of the electromagnetic fields, which constitutes an interesting parallel to quantum mechanical wavefunctions.

2.3.2 Time averaged energy balance

Let us further motivate the concept of the Poynting vector. Some interesting insights on the energy flow of light and hence also on the transport of information can be obtained from the Poynting theorem, which is the equation for the energy conservation of the electromagnetic field. The Poynting theorem can be derived directly from Maxwell's equations. We multiply the two curl equations by \mathbf{H}_r or \mathbf{E}_r (note that we must use real fields):

$$\begin{aligned}\mathbf{H}_r \cdot \text{rot} \mathbf{E}_r + \mu_0 \mathbf{H}_r \cdot \frac{\partial}{\partial t} \mathbf{H}_r &= 0 \\ \mathbf{E}_r \cdot \text{rot} \mathbf{H}_r - \epsilon_0 \mathbf{E}_r \cdot \frac{\partial}{\partial t} \mathbf{E}_r &= \mathbf{E}_r \cdot \left(\mathbf{j}_r + \frac{\partial}{\partial t} \mathbf{P}_r \right)\end{aligned}$$

Next, we subtract the two equations to obtain

$$\mathbf{H}_r \cdot \text{rot} \mathbf{E}_r - \mathbf{E}_r \cdot \text{rot} \mathbf{H}_r + \epsilon_0 \mathbf{E}_r \cdot \frac{\partial}{\partial t} \mathbf{E}_r + \mu_0 \mathbf{H}_r \cdot \frac{\partial}{\partial t} \mathbf{H}_r = -\mathbf{E}_r \cdot \left(\mathbf{j}_r + \frac{\partial}{\partial t} \mathbf{P}_r \right).$$

This equation can be simplified by using the following vector identity:

$$\mathbf{H}_r \cdot \text{rot} \mathbf{E}_r - \mathbf{E}_r \cdot \text{rot} \mathbf{H}_r = \text{div}(\mathbf{E}_r \times \mathbf{H}_r)$$

Finally, with the substitution $\mathbf{E}_r \cdot \partial \mathbf{E}_r / \partial t = \frac{1}{2} \partial \mathbf{E}_r^2 / \partial t$, we find Poynting's theorem

$$\underbrace{\frac{1}{2} \epsilon_0 \frac{\partial}{\partial t} \mathbf{E}_r^2 + \frac{1}{2} \mu_0 \frac{\partial}{\partial t} \mathbf{H}_r^2}_{\text{change of vacuum energy density}} + \underbrace{\text{div}(\mathbf{E}_r \times \mathbf{H}_r)}_{\text{change of Poynting vector}} = -\mathbf{E}_r \cdot \left(\mathbf{j}_r + \frac{\partial}{\partial t} \mathbf{P}_r \right) (*)$$

This equation has the general form of a balance equation. Here it represents the energy balance. Apart from the divergence of the Poynting vector (energy flux), we can identify the vacuum electromagnetic energy density $u = \frac{1}{2} \epsilon_0 \mathbf{E}_r^2 + \frac{1}{2} \mu_0 \mathbf{H}_r^2$. The right-hand-side of the Poynting's theorem contains the so-called source terms due to electrical work performed on free and bound charges, respectively.

$$\text{where } u = \frac{1}{2} \epsilon_0 \mathbf{E}_r^2 + \frac{1}{2} \mu_0 \mathbf{H}_r^2 \rightarrow \text{vacuum energy density}$$

In the case of stationary fields and isotropic media, which is a simple but important case,

$$\begin{aligned}\mathbf{E}_r(\mathbf{r}, t) &= \frac{1}{2} [\mathbf{E}(\mathbf{r}) \exp(-i\omega_0 t) + c.c.] \\ \mathbf{H}_r(\mathbf{r}, t) &= \frac{1}{2} [\mathbf{H}(\mathbf{r}) \exp(-i\omega_0 t) + c.c.]\end{aligned}$$

Time averaging of the left hand side of Poynting's theorem (*) yields:

$$\left\langle \frac{1}{2}\varepsilon_0 \frac{\partial}{\partial t} \mathbf{E}_r^2(\mathbf{r},t) + \frac{1}{2}\mu_0 \frac{\partial}{\partial t} \mathbf{H}_r^2(\mathbf{r},t) + \mathbf{div}[\mathbf{E}_r(\mathbf{r},t) \times \mathbf{H}_r(\mathbf{r},t)] \right\rangle = \frac{1}{2} \mathbf{div} \left\{ \Re [\mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r})] \right\} \\ = \mathbf{div} \langle \mathbf{S}(\mathbf{r},t) \rangle.$$

Note that the time derivatives remove stationary terms in $\mathbf{E}_r^2(\mathbf{r},t)$ and $\mathbf{H}_r^2(\mathbf{r},t)$.

Time averaging of the right hand side of Poynting's theorem yields (source terms):

$$-\left\langle \left[\mathbf{j}_r(\mathbf{r},t) + \frac{\partial}{\partial t} \mathbf{P}_r(\mathbf{r},t) \right] \mathbf{E}_r(\mathbf{r},t) \right\rangle \\ = -\frac{1}{4} \left\langle \left[\sigma(\omega_0) \mathbf{E}(\mathbf{r}) e^{-i\omega_0 t} - i\omega_0 \varepsilon_0 \chi(\omega_0) \mathbf{E}(\mathbf{r}) e^{-i\omega_0 t} + c.c. \right] \left[\mathbf{E}(\mathbf{r}) e^{-i\omega_0 t} + c.c. \right] \right\rangle$$

Now we use our generalized dielectric function:

$$= -\frac{1}{4} \left\langle \left[-i\omega_0 \varepsilon_0 \left(\chi(\omega_0) + i \frac{\sigma(\omega_0)}{\omega_0 \varepsilon_0} \right) \mathbf{E}(\mathbf{r}) \exp(-i\omega_0 t) + c.c. \right] \left[\mathbf{E}(\mathbf{r}) \exp(-i\omega_0 t) + c.c. \right] \right\rangle \\ = \frac{1}{4} i\omega_0 \varepsilon_0 \left[\varepsilon(\omega_0) - 1 \right] \mathbf{E}(\mathbf{r}) \mathbf{E}(\mathbf{r})^* + c.c.$$

$$\text{with } \varepsilon(\omega) = 1 + \chi(\omega) + \frac{i}{\omega \varepsilon_0} \sigma(\omega)$$

Again, all fast oscillating terms $\sim \exp(\pm 2i\omega_0 t)$ cancel due to the time average.

Finally, splitting $\varepsilon(\omega_0)$ into real and imaginary part yields

$$= \frac{1}{4} i\omega_0 \varepsilon_0 \left[\varepsilon'(\omega_0) - 1 + i\varepsilon''(\omega_0) \right] \mathbf{E}(\mathbf{r}) \mathbf{E}(\mathbf{r})^* + c.c. = -\frac{1}{2} \omega_0 \varepsilon_0 \varepsilon''(\omega_0) \mathbf{E}(\mathbf{r}) \mathbf{E}^*(\mathbf{r}).$$

where the imaginary terms cancel with their complex conjugates. Hence, the divergence of the time averaged Poynting vector is related to the imaginary part of the generalized dielectric function:

$$\rightarrow \mathbf{div} \langle \mathbf{S} \rangle = -\frac{1}{2} \omega_0 \varepsilon_0 \varepsilon''(\omega_0) \mathbf{E}(\mathbf{r}) \mathbf{E}^*(\mathbf{r}).$$

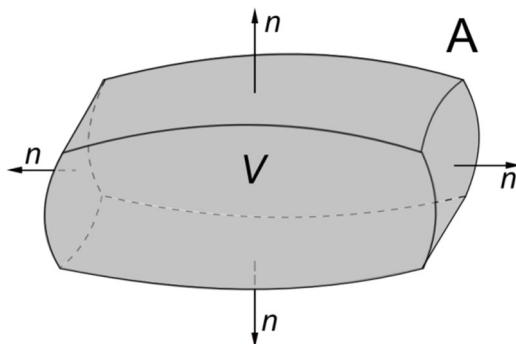
This shows that a nonzero imaginary part of epsilon ($\varepsilon''(\omega) \neq 0$) causes a **drain of energy flux**. In particular, we always have $\varepsilon''(\omega) > 0$, as otherwise there would be gain of energy, which we will not consider in this course as it would require a treatment of active media. Near resonances we have $\varepsilon''(\omega) \neq 0$ and therefore **absorption occurs**.

Further insight into the meaning of $\mathbf{div} \langle \mathbf{S} \rangle$ gives the so-called divergence theorem. If the energy of the electromagnetic field is flowing through some volume, and we wish

to know how much energy flows out of a certain region within that volume, then we need to add up the sources inside the region and subtract the sinks. The energy flux is represented by the (time averaged) Poynting vector, and its divergence at a given point describes the strength of the source or sink there. Thus, integrating the divergence of the Poynting vector over the interior of the region equals the surface integral of the Poynting vector flux over the region's boundary.

$$\int_V \operatorname{div} \langle \mathbf{S} \rangle dV = \int_A \langle \mathbf{S} \rangle \cdot \mathbf{n} dA$$

Here A is the surface of the volume V .



2.4 Kramers-Kronig relation (covered by lecture Structure of Matter)

In the previous sections we have assumed a very simple model for the description of the material's response to excitation by an electromagnetic field. This model was based on quite strong assumptions, such a single charge attached to a rigid lattice. Hence, one could imagine that more complex matter could give rise to arbitrarily complex response functions if adequate models were used for its description. However, we can show from basic laws of physics, that several properties are common to all possible response functions, as long as a linear response to the excitation is assumed.

These fundamental properties of the response function are formulated mathematically by the Kramers-Kronig relation. It is a general relation between $\epsilon'(\omega)$ (dispersion) and $\epsilon''(\omega)$ (absorption). This means in practice that we can compute $\epsilon'(\omega)$ from $\epsilon''(\omega)$ and vice versa. For example, if we have access to the absorption spectrum of a medium, we could calculate the dispersion.

The Kramers-Kronig relation follows from reality and causality of the response function R of a linear system. The real-valued nature of the response function is a direct consequence of Maxwell's equations which are real valued as well. Causality is also a fundamental property, since the present polarization must not depend on some future electric field. As we have seen in the previous sections, the polarization and the electric field are related in the time domain :

$$\mathbf{P}_r(\mathbf{r}, t) = \epsilon_0 \int_{-\infty}^t R(t-t') \mathbf{E}_r(\mathbf{r}, t') dt' \leftrightarrow \mathbf{P}_r(\mathbf{r}, t) = \epsilon_0 \int_0^\infty R(\tau) \mathbf{E}_r(\mathbf{r}, t-\tau) d\tau$$

where a simple variable substitution has been performed in the integral.

Reality of the response function implies:

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \chi(\omega) e^{-i\omega\tau} = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \chi^*(\omega) e^{i\omega\tau} \rightarrow \chi(\omega) = \chi^*(-\omega)$$

which again follows from a simple variable substitution in the integral.

Causality of the response function implies:

$$R(\tau) = \theta(\tau) y(\tau) \text{ with } \theta(\tau) = \begin{cases} 1 & \text{for } \tau > 0 \\ \frac{1}{2} & \text{for } \tau = 0 \\ 0 & \text{for } \tau < 0 \end{cases} \rightarrow \text{Heaviside distribution}$$

In the following, we will make use of the Fourier transform of the Heaviside distribution:

$$2\pi \bar{\theta}(\omega) = \int_{-\infty}^{\infty} dt \theta(t) e^{i\omega t} = P \frac{i}{\omega} + \pi \delta(\omega)$$

\rightarrow only defined under an integral

This Fourier transform of the Heaviside distribution consists of the Dirac delta distribution

$$\int_{-\infty}^{\infty} d\omega \delta(\omega - \omega_0) f(\omega) = f(\omega_0) \rightarrow \text{Dirac delta distribution}$$

and the expression $P(i/\omega)$ involving a Cauchy principal value:

$$P \int_{-\infty}^{\infty} d\omega \frac{i}{\omega} f(\omega) = \lim_{\alpha \rightarrow 0} \left[\int_{-\infty}^{-\alpha} d\omega \frac{i}{\omega} f(\omega) + \int_{\alpha}^{\infty} d\omega \frac{i}{\omega} f(\omega) \right] \rightarrow \text{Cauchy principle value}$$

which defines how to treat the pole at zero.

As we have seen above, causality implies that the response function must contain the Heaviside distribution. Hence, in Fourier space (susceptibility) we expect a convolution:

$$\begin{aligned} \chi(\omega) &= \int_{-\infty}^{\infty} d\tau R(\tau) e^{i\omega\tau} = \int_{-\infty}^{\infty} d\tau \theta(\tau) y(\tau) e^{i\omega\tau} \\ \rightarrow & \quad \bar{\theta}(\omega) = \frac{1}{2\pi} P \frac{i}{\omega} + \frac{1}{2} \delta(\omega) \\ &= \int_{-\infty}^{\infty} d\bar{\omega} \bar{\theta}(\omega - \bar{\omega}) \bar{y}(\bar{\omega}) \end{aligned}$$

$$\chi(\omega) = \frac{1}{2\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} + \frac{\bar{y}(\omega)}{2} \quad (*)$$

In order to derive the Kramers-Kronig relation, we can use a small trick which saves us from having to use complex integration in the derivation. Because of the Heaviside distribution which vanishes for $\tau < 0$, we can **arbitrarily** choose the behaviour of the function $y(\tau)$ for $\tau < 0$ **without altering the susceptibility!** In particular, we can choose:

- a) $y(-\tau) = y(\tau)$ even function
- b) $y(-\tau) = -y(\tau)$ odd function

a) $y(-\tau) = y(\tau)$

In this case $y(-\tau) = y(\tau)$ is a real-valued, even function. We can exploit this property and show that

$$\rightarrow \bar{y}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau y(\tau) e^{i\omega\tau} = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau y(\tau) e^{-i\omega\tau} = \bar{y}^*(\omega) \text{ is real as well}$$

Hence, we can conclude from equation (*) above that

$$\rightarrow \chi^*(\omega) = -\frac{1}{2\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} + \frac{\bar{y}(\omega)}{2}$$

Here $P \int$ is the so-called principal value integral (G: Hauptwertintegral).

Now that we have expressions for $\chi(\omega), \chi^*(\omega)$, we can compute the real and imaginary parts of the susceptibility:

$$\begin{aligned} \chi(\omega) + \chi^*(\omega) &= \frac{1}{2\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} + \frac{\bar{y}(\omega)}{2} - \frac{1}{2\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} + \frac{\bar{y}(\omega)}{2} = \bar{y}(\omega) \\ \chi(\omega) - \chi^*(\omega) &= \dots = \frac{1}{\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} \end{aligned}$$

Plugging the last two equations together, we find the first Kramers-Kronig relation:

$$\rightarrow \Im \chi(\omega) = -\frac{1}{\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{\Re \chi(\bar{\omega})}{\bar{\omega} - \omega} \quad 1. \text{ K-K relation}$$

Knowledge of the real part of the susceptibility (dispersion) allows us to compute the imaginary part (absorption).

b) $y(-\tau) = -y(\tau)$

The second K-K relation can be found by a similar procedure if we assume that $y(-\tau) = -y(\tau)$ is a real odd function. We can show that in this case

$$\rightarrow \bar{y}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau y(\tau) e^{i\omega\tau} = -\frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau y(\tau) e^{-i\omega\tau} = -\bar{y}^*(\omega) \text{ is purely imaginary}$$

With equation (*), we then find that

$$\rightarrow \chi^*(\omega) = \frac{1}{2\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} - \frac{\bar{y}(\omega)}{2} \quad (\text{see } *) \quad \text{and}$$

Again we can then compute real and imaginary part of the susceptibility

$$\chi(\omega) - \chi^*(\omega) = \frac{1}{2\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} + \frac{\bar{y}(\omega)}{2} - \frac{1}{2\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} + \frac{\bar{y}(\omega)}{2} = \bar{y}(\omega)$$

$$\chi(\omega) + \chi^*(\omega) = \dots = \frac{1}{\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{i \bar{y}(\bar{\omega})}{\omega - \bar{\omega}}$$

and finally obtain

$$\rightarrow \Re \chi(\omega) = \frac{1}{\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \frac{\Im \chi(\bar{\omega})}{\bar{\omega} - \omega}$$

2. K-K relation

The second Kramers-Kronig relation allows us to compute the real part of the susceptibility (dispersion) when we know its imaginary part (absorption).

The Kramers-Kronig relation can also be rewritten in terms of the dielectric function, where one applies also the symmetry relation for $\chi(\omega)$ (in turn derived from the real-valued nature of the response function):

K-K relation for ϵ :

- $\chi(\omega) = \chi^*(-\omega) \rightarrow \chi'(\omega) = \chi'(-\omega) \quad \chi''(\omega) = -\chi''(-\omega)$ and
 $\chi(\omega) = \epsilon(\omega) - 1 = [\epsilon'(\omega) - 1] + i\epsilon''(\omega)$

$$\boxed{\begin{aligned} \epsilon'(\omega) - 1 &= \frac{2}{\pi} P \int_0^{\infty} \frac{\bar{\omega} \epsilon''(\bar{\omega})}{\bar{\omega}^2 - \omega^2} d\bar{\omega}, \\ \epsilon''(\omega) &= -\frac{2}{\pi} \omega P \int_0^{\infty} \frac{[\epsilon'(\bar{\omega}) - 1]}{\bar{\omega}^2 - \omega^2} d\bar{\omega}. \end{aligned}}$$

- dispersion and absorption are linked, e.g., we can measure absorption and compute dispersion

Example:

$$\epsilon''(\omega) \sim \delta(\omega - \omega_0) \rightarrow \epsilon'(\omega) - 1 \sim \frac{\omega_0}{\omega_0^2 - \omega^2} \rightarrow \text{Drude-Lorentz model}$$

3. Solution of wave equations for homogeneous isotropic media

3.1 Normal modes in homogeneous isotropic media

Using the linear material models which we discussed in the previous chapters, we can now look for self-consistent solutions to the wave equation including the material response.

It is convenient to use the generalized complex dielectric function to derive solutions to the wave equation

$$\epsilon(\omega) = 1 + \chi(\omega) + \frac{\mathbf{i}}{\omega\epsilon_0}\sigma(\omega) = \epsilon'(\omega) + \mathbf{i}\epsilon''(\omega)$$

We will conduct our analysis in Fourier domain. In particular, we will focus on the most simple solutions to the wave equation in Fourier domain, the so-called normal modes. These normal modes are the stationary solutions of the wave equation. Hence they usually correspond to waves which possess a trivial spatio-temporal dynamics and have infinite extent in space and time. However, as we will see later, by the principle of superposition, we can construct general solutions by superimposing multiple normal modes. In this context, one should understand that by superimposing infinitely many infinitely extended normal modes, it is possible to form all sorts of localized/transient, i.e. finitely extended, waves.

We start from the wave equation in Fourier domain, which reads as

$$\text{rot rot } \bar{\mathbf{E}}(\mathbf{r}, \omega) = \frac{\omega^2}{c^2} \epsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

According to Maxwell the solutions have to fulfill additionally the divergence equation of the **D** field, which for homogeneous media can be transformed to:

$$\epsilon_0 [1 + \chi(\omega)] \mathbf{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0$$

In general, the additional condition $1 + \chi(\omega) \neq 0$ implies that the electric field itself is free of divergence:

$$\text{for the usual case of } 1 + \chi(\omega) \neq 0 \rightarrow \mathbf{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0$$

Since we already know from past research in this field that plane waves constitute the normal mode solutions of interest, we will use them as a 'smart' guess, i.e. an ansatz. Thus, here in the lecture we will restrict ourselves to showing the validity of this ansatz, i.e. we are showing that this ansatz correctly fulfills the Maxwell's equations.

We will start from plane wave solutions of the following form in the frequency domain:

$$\bar{\mathbf{E}}(\mathbf{r}, \omega) = \bar{\mathbf{E}}(\mathbf{k}, \omega) \exp(\mathbf{i}\mathbf{k}\mathbf{r}),$$

k = unknown complex wavevector

The corresponding stationary field in time domain is given by:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{k}, \omega) \exp[\mathbf{i}(\mathbf{k}\mathbf{r} - \omega t)]$$

→ monochromatic plane wave → normal mode

This is a monochromatic plane wave, the simplest solution we can expect, a so-called normal mode.

We insert this ansatz into Maxwell's equations to see if Maxwell's equations are fulfilled by this ansatz.

Starting from the divergence equation for the \mathbf{E} field (derived from the divergence equation for the \mathbf{D} field in homogeneous media), we see that for $\epsilon(\omega) \neq 0$ the plane wave ansatz fulfills the divergence condition only if the so-called transversality condition $\mathbf{k} \perp \bar{\mathbf{E}}(\omega)$ is fulfilled. Accordingly, these waves are called transverse waves.

$$0 = \operatorname{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = \mathbf{i} \mathbf{k} \cdot \bar{\mathbf{E}}(\mathbf{r}, \omega) \rightarrow \mathbf{k} \perp \bar{\mathbf{E}}(\omega) \rightarrow \text{transverse wave}$$

Here transverse means that the electric field and the wavevector are oriented perpendicularly to each other.

If we split the complex wavevector into real and imaginary parts $\mathbf{k} = \mathbf{k}' + \mathbf{i} \mathbf{k}''$, we can define:

- planes of **constant phase** $\mathbf{k}'\mathbf{r} = \text{const}$
- planes of **constant amplitude** $\mathbf{k}''\mathbf{r} = \text{const}$

In the following we will call the solutions

- A) **homogeneous waves** → if these two planes are identical (parallel)
- B) **evanescent waves** → if these two planes are perpendicular
- C) **inhomogeneous waves** → otherwise

We will see that in dielectrics ($\sigma(\omega) = 0$), we can find another exotic type of wave solutions, since at the frequency $\omega = \omega_L$ where $\epsilon(\omega_L) = 0$, so-called **longitudinal waves** $\mathbf{k} \parallel \bar{\mathbf{E}}(\omega)$ can occur.

Before we have a closer look at the detailed properties of these individual types of solutions, let us consider the two principal forms of plane waves, which are the transverse and longitudinal waves.

3.1.1 Transverse waves (nonzero epsilon)

Let us first take a closer look at the nature of transverse fields. As pointed out above, for $\omega \neq \omega_L$ the electric field becomes divergence-free:

$$\epsilon_0 \epsilon(\omega) \operatorname{div} \mathbf{E}(\mathbf{r}, \omega) = 0 \rightarrow \operatorname{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0$$

Thus, the wave equation reduces to the Helmholtz equation:

$$\Delta \bar{\mathbf{E}}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \epsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0.$$

Hence, we have three scalar equations for $\bar{\mathbf{E}}(\mathbf{r}, \omega)$ (from Helmholtz). Together with the above divergence condition, the original six electric and magnetic field components are reduced to just two independent field components. We will now construct solutions using the plane wave ansatz:

$$\bar{\mathbf{E}}(\mathbf{r}, \omega) = \bar{\mathbf{E}}(\omega) \exp(i\mathbf{k}\mathbf{r})$$

Immediately we see that the wave must be **transverse**:

$$0 = \mathbf{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = i\mathbf{k} \cdot \bar{\mathbf{E}}(\mathbf{r}, \omega) \rightarrow \mathbf{k} \perp \bar{\mathbf{E}}(\omega).$$

Hence, we have to solve

$$\left[-\mathbf{k}^2 + \frac{\omega^2}{c^2} \epsilon(\omega) \right] \bar{\mathbf{E}}(\omega) = 0 \quad \text{and} \quad \mathbf{k} \cdot \bar{\mathbf{E}}(\omega) = 0.$$

If $\bar{\mathbf{E}}(\omega) \neq 0$ the first term must be zero. This leads to the following **dispersion relation**

$$\mathbf{k}^2 = k^2 = k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2} \epsilon(\omega) \quad \text{with } k \text{ as the complex wavenumber}$$

We see that the so-called wavenumber $k(\omega) = \frac{\omega}{c} \sqrt{\epsilon(\omega)}$ is a function of the frequency.

We can conclude that transverse plane waves are solutions to Maxwell's equations in homogeneous, isotropic media, only if the dispersion relation for $k(\omega)$ is fulfilled.

In general, $\mathbf{k} = \mathbf{k}' + i\mathbf{k}''$ is a complex-valued vector. For cases where $\mathbf{k}' \parallel \mathbf{k}''$, the vectorial character is not important and thus it is sometimes useful to introduce the complex scalar refractive index \hat{n} as an alternative notation:

$$k(\omega) = \frac{\omega}{c} \sqrt{\epsilon(\omega)} = \frac{\omega}{c} \hat{n}(\omega) = \frac{\omega}{c} [n(\omega) + i\kappa(\omega)]$$

Remark: Instead of assuming that $\hat{n}(\omega)$ and $\sqrt{\epsilon(\omega)}$ have the same meaning, one should clearly distinguish between the two. While $\epsilon(\omega)$ is a property of the medium, $\hat{n}(\omega)$ is a property of a particular type of electromagnetic field in the medium, i.e. a property of the infinitely extended monochromatic plane wave

$$\bar{\mathbf{E}}(\mathbf{r}, \omega) = \bar{\mathbf{E}}(\omega) \exp(i\mathbf{k}\mathbf{r}).$$

Hence, the coincidence of the complex refractive index $\hat{n}(\omega)$ with $\sqrt{\epsilon(\omega)}$ holds only for homogeneous media since $\sqrt{\epsilon(\omega)}$ is a local property which can in general vary within an inhomogeneous medium while $\hat{n}(\omega)$ is a global property of a wave solution. From knowledge of the electric field, we can compute the magnetic field if desired:

$$\bar{\mathbf{H}}(\mathbf{r}, \omega) = -\frac{i}{\omega \mu_0} \mathbf{rot} \bar{\mathbf{E}}(\mathbf{r}, \omega) = \frac{1}{\omega \mu_0} [\mathbf{k} \times \bar{\mathbf{E}}(\omega)] \exp(i\mathbf{k}\mathbf{r})$$

$$\rightarrow \bar{\mathbf{H}}(\mathbf{r}, \omega) = \bar{\mathbf{H}}(\omega) \exp(i\mathbf{k}\mathbf{r}) \quad \text{with} \quad \bar{\mathbf{H}}(\omega) = \frac{1}{\omega \mu_0} [\mathbf{k} \times \bar{\mathbf{E}}(\omega)]$$

and hence $\bar{\mathbf{H}}(\omega) \perp \bar{\mathbf{E}}(\omega)$.

3.1.2 Longitudinal waves (epsilon equal zero)

Let us now take a look at the rather exotic case of longitudinal waves. These waves can only exist for $\epsilon(\omega)=0$ in dielectrics, which occurs at the longitudinal frequency $\omega=\omega_L$. In this case, we cannot conclude from MWQE that $\operatorname{div} \bar{\mathbf{E}}(\mathbf{r}, \omega)=0$, and hence the wave equation reads (the l.h.s. vanishes because $\epsilon(\omega)=0$):

$$\rightarrow \operatorname{rot} \operatorname{rot} \bar{\mathbf{E}}(\mathbf{r}, \omega_L)=0$$

As with the transversal waves, we try the plane wave ansatz and assume \mathbf{k} to be real.

$$\bar{\mathbf{E}}(\mathbf{r}, \omega)=\bar{\mathbf{E}}(\omega) \exp(i \mathbf{k} \cdot \mathbf{r})$$

With $\operatorname{rot}[\bar{\mathbf{E}}(\omega) \exp(i \mathbf{k} \cdot \mathbf{r})]=i \mathbf{k} \times \bar{\mathbf{E}}(\omega) \exp(i \mathbf{k} \cdot \mathbf{r})$, we obtain from the wave equation:

$$\mathbf{k} \times [\mathbf{k} \times \bar{\mathbf{E}}(\mathbf{r}, \omega_L)]=0$$

Now we decompose the electric field into transversal and longitudinal components with respect to the wave vector:

$$\bar{\mathbf{E}}(\mathbf{r}, \omega)=\bar{\mathbf{E}}(\omega) \exp(i \mathbf{k} \cdot \mathbf{r})=\bar{\mathbf{E}}_{\perp}(\omega) \exp(i \mathbf{k} \cdot \mathbf{r})+\bar{\mathbf{E}}_{\parallel}(\omega) \exp(i \mathbf{k} \cdot \mathbf{r})$$

$$\text { with } \bar{\mathbf{E}}_{\perp}(\omega) \perp \mathbf{k} \text { and } \bar{\mathbf{E}}_{\parallel}(\omega) \| \mathbf{k}$$

This decomposed field is inserted into the wave equation:

$$\begin{aligned} \mathbf{k} \times [\mathbf{k} \times (\bar{\mathbf{E}}_{\perp} + \bar{\mathbf{E}}_{\parallel})] \exp(i \mathbf{k} \cdot \mathbf{r}) &= 0 \\ \mathbf{k} \times [\mathbf{k} \times \bar{\mathbf{E}}_{\perp}] \exp(i \mathbf{k} \cdot \mathbf{r}) + \mathbf{k} \times [\underbrace{\mathbf{k} \times \bar{\mathbf{E}}_{\parallel}}_{=0}] \exp(i \mathbf{k} \cdot \mathbf{r}) &= 0 \end{aligned}$$

Since the cross product of \mathbf{k} with the longitudinal field $\bar{\mathbf{E}}_{\parallel}(\omega)$ is trivially zero, the remaining wave equation is (the vectors in the remaining term are mutually orthogonal):

$$k^2 \bar{\mathbf{E}}_{\perp}=0$$

Hence the transversal field $\bar{\mathbf{E}}_{\perp}$ must vanish and the only remaining field component is the longitudinal field $\bar{\mathbf{E}}_{\parallel}(\omega)$:

$$\rightarrow \boxed{\bar{\mathbf{E}}(\mathbf{r}, \omega_L)=\bar{\mathbf{E}}_{\parallel}(\omega_L) \exp(i \mathbf{k} \cdot \mathbf{r})}$$

3.1.3 Plane wave solutions in different frequency regimes

The dispersion relation $\mathbf{k}^2=k^2=k_x^2+k_y^2+k_z^2=(\omega^2/c^2)\epsilon(\omega)$ for plane wave solutions only determines the (complex scalar) wavenumber k . Thus, different solutions for the complex wavevector $\mathbf{k}=\mathbf{k}'+i\mathbf{k}''$ are possible. In addition, the generalized dielectric function $\epsilon(\omega)$ is complex. In this chapter, we will discuss possible scenarios and the corresponding plane wave solutions.

A) Positive real-valued epsilon $\epsilon(\omega) = \epsilon'(\omega) > 0$

This is the favorable regime for optics. We have transparency as the frequency of light is far from resonances of the medium. The dispersion relation gives

$$k^2 = \mathbf{k}'^2 - \mathbf{k}''^2 + 2i\mathbf{k}' \cdot \mathbf{k}'' = \frac{\omega^2}{c^2} \epsilon'(\omega) = \frac{\omega^2}{c^2} n^2(\omega) \quad \Rightarrow \quad \mathbf{k}' \cdot \mathbf{k}'' = 0$$

Note that in this case \mathbf{k}' can't be zero since $\epsilon'(\omega)$ is positive and real. There are only two possibilities for fulfilling this condition: either $\mathbf{k}'' = 0$ or $\mathbf{k}' \perp \mathbf{k}''$.

A.1) Real valued wavevector $\mathbf{k}'' = 0$

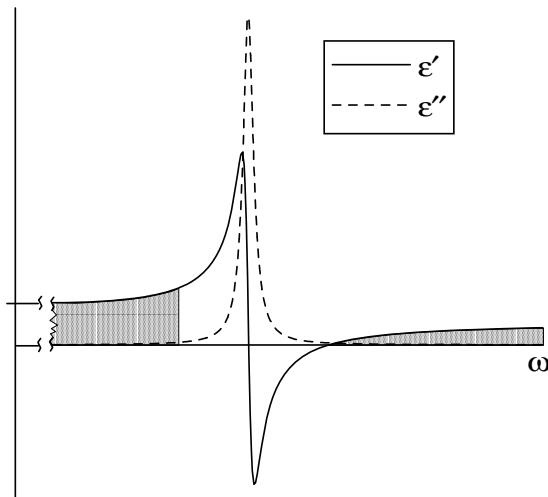
In this case the **wave vector is real** and we find the dispersion relation

$$k(\omega) = \frac{\omega}{c} n(\omega) = \frac{\omega}{c_n} = \frac{2\pi}{\lambda} n(\omega)$$

Because $\mathbf{k}'' = 0$, these waves are homogeneous, i.e. planes of constant phase are parallel to the planes of constant amplitude. This is trivial, since the amplitude is constant everywhere.

Example 1: single resonance in dielectric material

- for lattice vibrations (phonons)

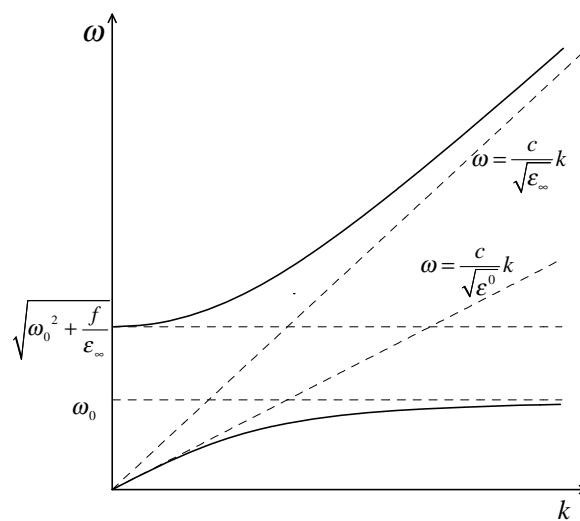


- If we neglect the imaginary part of $\epsilon(\omega)$, which mathematically corresponds to an undamped resonance, we can further simplify the material equation to

$$\epsilon(\omega) = \epsilon'(\omega) = \epsilon_{\infty} + \frac{f}{\omega_0^2 - \omega^2}$$

- We can invert the dispersion relation

$$k(\omega) = \frac{\omega}{c} \sqrt{\epsilon(\omega)} \rightarrow \omega(k)$$



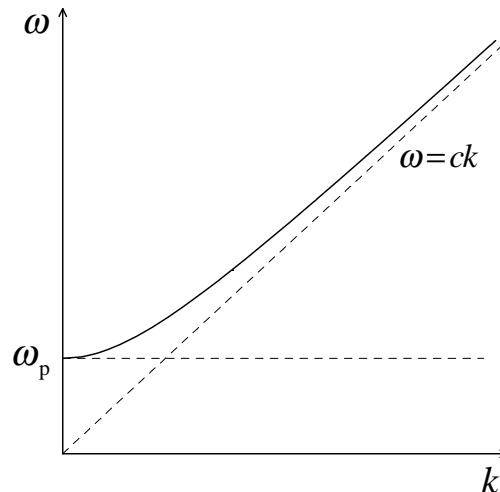
Example 2: free electrons

- for plasma and metal
- Again, the imaginary part of $\epsilon(\omega)$ is neglected

$$\epsilon(\omega) = \epsilon'(\omega) = 1 - \frac{\omega_p^2}{\omega^2}$$

- We again invert the dispersion relation

$$k(\omega) = \frac{\omega}{c} \sqrt{\epsilon(\omega)} \rightarrow \omega(k)$$



A.2) Complex valued wavevector $\mathbf{k}' \perp \mathbf{k}''$

The second possibility for fulfilling the dispersion relation leads to a complex wavevector and so-called **evanescent waves**. We find

$$k^2 = \mathbf{k}'^2 - \mathbf{k}''^2 = \frac{\omega^2}{c^2} \epsilon(\omega) \text{ and therefore } \mathbf{k}''^2 = \mathbf{k}'^2 - k^2$$

This means that

$$\rightarrow \boxed{\mathbf{k}''^2 \neq 0 \text{ and } \mathbf{k}'^2 > k^2}$$

We will discuss the importance of evanescent waves in the next chapter, where we will study the propagation of arbitrary initial field distributions. What is interesting to note here is that evanescent waves can have any arbitrary \mathbf{k}''^2 larger than the wave number k^2 , whereas the homogeneous waves of case A.1 ($\mathbf{k}''=0$) obey $\mathbf{k}''^2=k^2$. If we plug our findings into the plane wave ansatz we get for the evanescent waves:

$$\bar{\mathbf{E}}(\mathbf{r},\omega)=\mathbf{E}(\omega)\exp\left\{i[\mathbf{k}'(\omega)\mathbf{r}]\right\}\exp(-\mathbf{k}''(\omega)\mathbf{r})$$

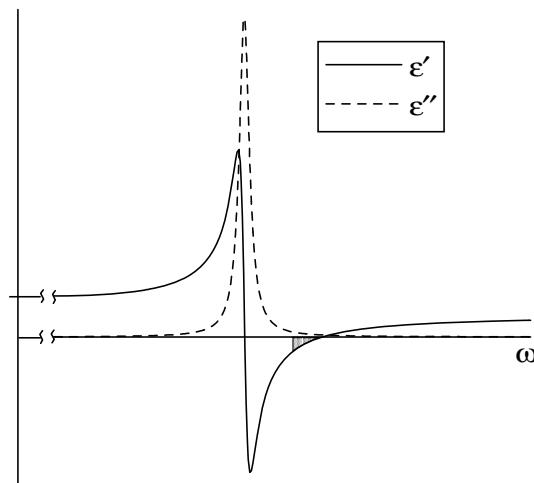
Here we see that the planes defined by the equation $\mathbf{k}''(\omega)\mathbf{r}=\text{const.}$ are the **planes of constant amplitude**, those defined by $\mathbf{k}'(\omega)\mathbf{r}=\text{const.}$ are the **planes of constant phase**. Because of $\mathbf{k}' \perp \mathbf{k}''$ these planes are **perpendicular** to each other.

The factor $\exp(-\mathbf{k}''(\omega)\mathbf{r})$ leads to **exponential growth** of evanescent waves in the direction of $-\mathbf{k}''(\omega)$ in homogeneous space. Therefore, evanescent waves can't be physical normal modes of homogeneous space and can only exist in inhomogeneous space, where the exponential growth is truncated at a finite value, e.g. at an **interface**.

B) Negative real-valued epsilon $\epsilon(\omega)=\epsilon'(\omega)<0$

This situation (negative, real $\epsilon(\omega)$) can occur in dielectric materials near resonances ($\omega_0 < \omega < \omega_L$) or in metals below the plasma frequency ($\omega < \omega_p$). Then the dispersion relation gives

$$k^2 = \mathbf{k}'^2 - \mathbf{k}''^2 + 2i\mathbf{k}' \cdot \mathbf{k}'' = \frac{\omega^2}{c^2} \epsilon'(\omega) - \frac{\omega^2}{c^2} \epsilon''(\omega) < 0$$



As in the previous case A, the vanishing imaginary term requires that $\mathbf{k}' \cdot \mathbf{k}''=0$. Again this can be achieved by two possibilities. Note that the case of $\mathbf{k}''=0$ is excluded as it would be inconsistent with a negative $\epsilon(\omega)$.

B.1) Vanishing real component of wavevector $\mathbf{k}'=0$

$$\rightarrow \mathbf{k}''^2 = \frac{\omega^2}{c^2} |\epsilon'(\omega)| \rightarrow \bar{\mathbf{E}}(\mathbf{r},\omega) \sim \exp(-\mathbf{k}''\mathbf{r}) \rightarrow \text{strong damping}$$

B.2) Complex-valued wavevector $\mathbf{k}' \cdot \mathbf{k}'' = 0$

$\rightarrow \mathbf{k}' \perp \mathbf{k}'' \rightarrow$ evanescent waves

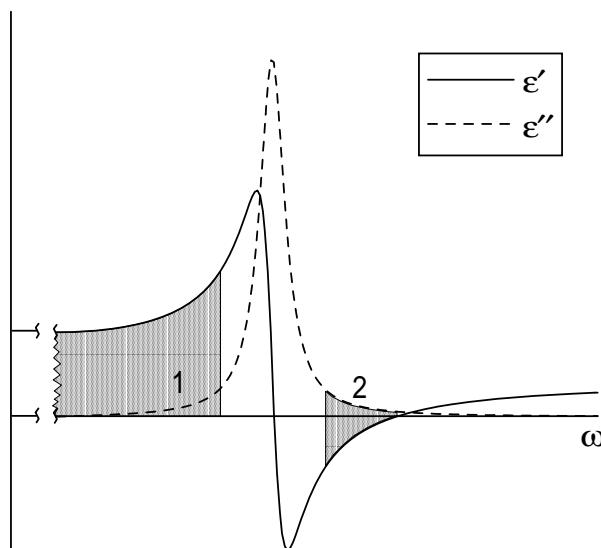
$$k^2 = \mathbf{k}'^2 - \mathbf{k}''^2 = -\frac{\omega^2}{c^2} |\epsilon'(\omega)|$$

$$\mathbf{k}''^2 = \frac{\omega^2}{c^2} |\epsilon'(\omega)| + \mathbf{k}'^2.$$

As above, these evanescent waves exist only at **interfaces** (as with $\epsilon(\omega) = \epsilon'(\omega) > 0$). The interesting point is that here we find evanescent waves for all values of \mathbf{k}' . This should be contrasted with the case A.1 for $\epsilon(\omega) = \epsilon'(\omega) > 0$, where the magnitude of the real part of the complex wavevector must exceed the complex wave number ($\mathbf{k}'^2 > k^2$). Moreover, we see that case B.1 ($\mathbf{k}' = 0$) is actually included in case B.2. Hence, we can conclude that for $\epsilon(\omega) = \epsilon'(\omega) < 0$, we find **only evanescent waves**.

C) Complex valued epsilon $\epsilon(\omega)$

This is the general case, which is particularly relevant near resonances. From our (optical) point of view, only weak absorption is interesting since otherwise the information carried by light would be lost quickly. Therefore, in the following, we will always assume $\epsilon''(\omega) \ll |\epsilon'(\omega)|$. Recall that in the treatment of passive media, we assume $\epsilon''(\omega)$ to be positive. As we can see in the following sketch, we can have the two cases $\epsilon'(\omega) > 0, \epsilon''(\omega) > 0$, or $\epsilon'(\omega) < 0, \epsilon''(\omega) > 0$.



Let us further consider only the important special case of quasi-homogeneous plane waves, i.e., \mathbf{k}' and \mathbf{k}'' are almost parallel. It is then convenient to use the complex refractive index

$$[\mathbf{k}' + i\mathbf{k}'']^2 = k^2(\omega) = \frac{\omega^2}{c^2} \epsilon(\omega) = \frac{\omega^2}{c^2} \hat{n}^2(\omega) = \frac{\omega^2}{c^2} [n(\omega) + i\kappa(\omega)]^2$$

Since \mathbf{k}' and \mathbf{k}'' are almost parallel, we can neglect their vectorial natures and express their magnitudes in terms of the complex (scalar) refractive index:

$$\rightarrow |\mathbf{k}'| = \frac{\omega}{c} n(\omega), \quad |\mathbf{k}''| = \frac{\omega}{c} \kappa(\omega)$$

The dispersion relation written in terms of the complex refractive index gives

$$\mathbf{k}^2 = k^2 = \frac{\omega^2}{c^2} \epsilon(\omega) = \frac{\omega^2}{c^2} [n(\omega) + i\kappa(\omega)]^2$$

We can find the complex refractive index as the square root of the complex dielectric constant:

$$\epsilon(\omega) = \epsilon'(\omega) + i\epsilon''(\omega) = n^2(\omega) - \kappa^2(\omega) + 2in(\omega)\kappa(\omega),$$

and therefore $\rightarrow \epsilon'(\omega) = n^2(\omega) - \kappa^2(\omega)$
 $\epsilon''(\omega) = 2n(\omega)\kappa(\omega)$

By solving these simultaneous equations for the real and imaginary parts of the complex refractive index:

$$n^2(\omega) = \frac{\epsilon'}{2} \left[\operatorname{sgn}(\epsilon') \sqrt{1 + (\epsilon''/\epsilon')^2} + 1 \right],$$

$$\kappa^2(\omega) = \frac{\epsilon'}{2} \left[\operatorname{sgn}(\epsilon') \sqrt{1 + (\epsilon''/\epsilon')^2} - 1 \right].$$

There are two important limiting cases of such quasi-homogeneous plane waves:

C.1) $\epsilon', \epsilon'' > 0, \epsilon'' \ll \epsilon'$, (dielectric media)

By keeping only the leading order terms in the Taylor expansion of the above square roots, we obtain the following approximate results

$$n(\omega) \approx \sqrt{\epsilon'(\omega)}, \quad \kappa(\omega) \approx \frac{1}{2} \frac{\epsilon''(\omega)}{\sqrt{\epsilon'(\omega)}}$$

In this regime, **propagation dominates** ($n(\omega) \gg \kappa(\omega)$) and we have weak absorption:

$$\mathbf{k}'^2 - \mathbf{k}''^2 = \frac{\omega^2}{c^2} \epsilon'(\omega), \quad 2\mathbf{k}' \cdot \mathbf{k}'' = \frac{\omega^2}{c^2} \epsilon''(\omega),$$

$$|\mathbf{k}'| = \frac{\omega}{c} n(\omega) \approx \frac{\omega}{c} \sqrt{\epsilon'(\omega)}, \quad |\mathbf{k}''| = \frac{\omega}{c} \kappa(\omega) \approx \frac{1}{2} \frac{\omega}{c} \frac{\epsilon''(\omega)}{\sqrt{\epsilon'(\omega)}}.$$

We see that

$$\rightarrow \mathbf{k}' \cdot \mathbf{k}'' \approx |\mathbf{k}'||\mathbf{k}''|$$

This is self-consistent with our initial assumption for case C that \mathbf{k}' and \mathbf{k}'' must be almost parallel, which indicates that under these conditions the plane waves are quasi-homogeneous wave.

Thus, in homogeneous, isotropic media and close to resonances, we find damped, quasi-homogeneous plane waves, with $\mathbf{k}' \parallel \mathbf{k}'' \parallel \mathbf{e}_k$ where \mathbf{e}_k is the unit vector along \mathbf{k}

$$\bar{\mathbf{E}}(\mathbf{r}, \omega) = \mathbf{E}(\omega) \exp(i\mathbf{k}\mathbf{r}) = \mathbf{E}(\omega) \exp\left\{i\left[\frac{\omega}{c}n(\omega)(\mathbf{e}_k\mathbf{r})\right]\right\} \exp\left[-\frac{\omega}{c}\kappa(\omega)(\mathbf{e}_k\mathbf{r})\right].$$

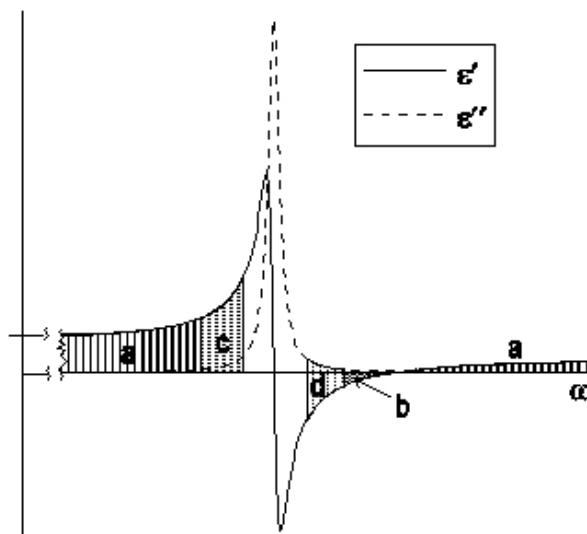
C.2) $\epsilon' < 0, \epsilon'' > 0, \epsilon'' \ll |\epsilon'|$, (metals and dielectric media in the so-called Reststrahl domain)

$$n(\omega) \approx \frac{1}{2} \frac{\epsilon''(\omega)}{\sqrt{|\epsilon'(\omega)|}}, \quad \kappa(\omega) \approx \sqrt{|\epsilon'(\omega)|},$$

In this regime, **damping dominates** ($n(\omega) \ll \kappa(\omega)$) and we find a very small refractive index. Interestingly, propagation (nonzero n) is only possible due to the imaginary part $\epsilon''(\omega)$ of the dielectric constant, which is also associated with absorption (see time averaged Poynting vector below).

Summary of normal modes

Here we summarize our previous findings about the properties of normal modes for different parameters of the material in which they exist. We do this with the example of dielectric media at frequencies close to a resonance. Here we can identify different frequency intervals in which we find the all the possible types of behavior of the material's response.



The following frequency ranges give rise to all the possible types of normal modes:

- a) Frequency far below or far above the resonance frequency, where $\epsilon'(\omega) > 0, \epsilon''(\omega) \approx 0$

Typical normal modes:

- undamped homogeneous waves → case A.1
- evanescent waves → case A.2

- b) Frequency above the resonance frequency, where $\epsilon'(\omega) < 0$, $\epsilon''(\omega) \approx 0$
 Typical normal modes:
 – evanescent waves → cases B.1 and B.2
- c) Frequency close to and below the resonance frequency, where $\epsilon'(\omega) > 0$, $\epsilon''(\omega) > 0$
 Typical normal modes:
 – weakly damped quasi-homogeneous waves → case C.1
- d) Frequency close to and above the resonance frequency, where $\epsilon'(\omega) > 0$, $\epsilon''(\omega) > 0$
 Typical normal modes:
 – strongly damped quasi-homogeneous waves → case C.2

Optical systems work mainly in regime a) since here we find light propagating undamped over long distances through space. Furthermore, we sometimes exploit regime b) when we would like to have reflection of light at interfaces, e.g. at a metallic mirror.

3.1.4 Time averaged Poynting vector of plane waves

To understand the implication of the discussed properties of plane waves on experimentally measurable observables, we should look at their energy density flow by calculating their Poynting vector, which was defined in terms of the slowly-varying field envelopes as

$$\langle \mathbf{S}(\mathbf{r}, t) \rangle = \frac{1}{2} \frac{1}{T_m} \int_{t-T_m/2}^{t+T_m/2} \Re \left[\tilde{\mathbf{E}}(\mathbf{r}, t') \times \tilde{\mathbf{H}}^*(\mathbf{r}, t') \right] dt',$$

For plane waves, we define the electric and magnetic field as

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \mathbf{E} \exp(i\mathbf{k}\mathbf{r} - i\omega t) = \mathbf{E} \exp(i\mathbf{k}'\mathbf{r} - \mathbf{k}''\mathbf{r} - i\omega t) \\ \mathbf{H}(\mathbf{r}, t) &= \frac{1}{\omega\mu_0} \mathbf{k} \times \mathbf{E}(\mathbf{r}, t) \end{aligned}$$

Since these plane waves are completely stationary, the average over measurement time in the Poynting vector definition becomes trivial:

$$\rightarrow \langle \mathbf{S}(\mathbf{r}, t) \rangle = \frac{1}{2} \frac{\mathbf{k}'}{\omega\mu_0} \exp[-2\mathbf{k}'' \cdot \mathbf{r}] |\mathbf{E}|^2 = \frac{1}{2} \cancel{n} \sqrt{\frac{\epsilon_0}{\mu_0}} \mathbf{e}_{k'} \exp \left[-2 \frac{\omega}{c} \mathbf{k} (\mathbf{e}_{k''} \cdot \mathbf{r}) \right] |\mathbf{E}|^2$$

Here $\mathbf{e}_{k'}$ is the unit vector along \mathbf{k}' and $\mathbf{e}_{k''}$ is the unit vector along \mathbf{k}'' . Note that the direction of energy density flow is along \mathbf{k}' , while \mathbf{k}'' determines the direction along which the Poynting vector decays exponentially.

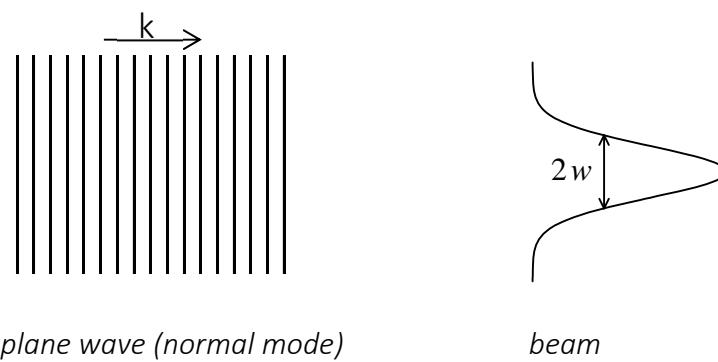
3.2 Beams and pulses as well as the analogy of diffraction and dispersion

Based on our understanding of the normal mode solutions of the wave equation, in this chapter we will analyze the propagation of light. In particular, we will answer the question how an arbitrary beam (spatial) or pulse (temporal) will change during propagation in isotropic, homogeneous, dispersive media. Relevant (linear) physical effects are diffraction and dispersion. Both phenomena can be understood very easily in the Fourier domain. Temporal effects, i.e. the dispersion of pulses, will be treated in temporal Fourier domain (temporal frequency domain). Spatial effects, i.e. the diffraction of beams, will be treated in the spatial Fourier domain (spatial frequency domain). We will see that:

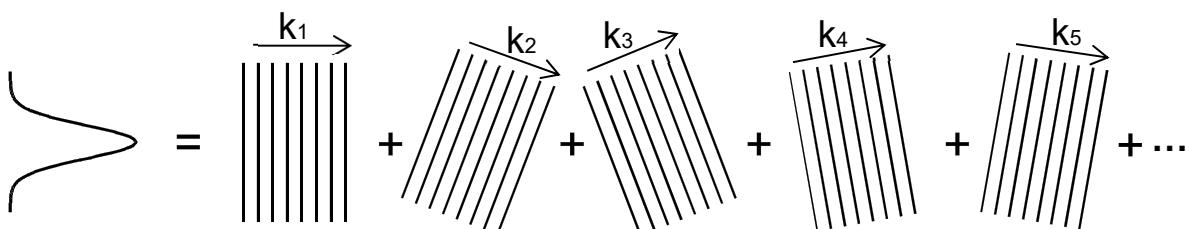
- Pulses with **finite spatial width** (i.e. pulsed beams) can be described as superpositions of normal modes in the frequency and spatial frequency domain.
- Spatio-temporally **localized** optical excitations **delocalize** during propagation because of the **different phase evolution** of the excited normal modes for different frequencies and spatial frequencies (different propagation directions of normal modes).

Let us have a look at the different possibilities (beam, pulse, pulsed beam)

A) beam → finite transverse width → diffraction

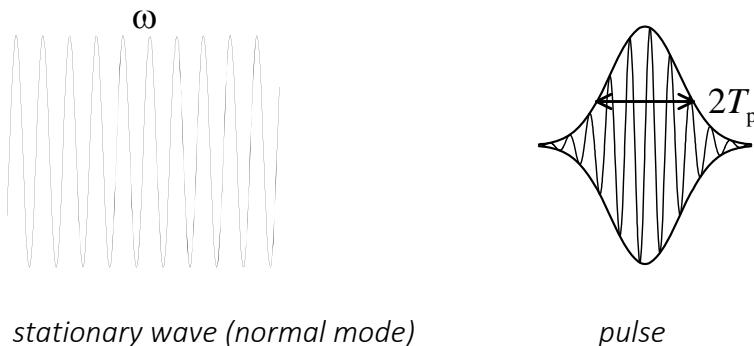


A beam is a continuous superposition of stationary plane waves (normal modes) with **different wavevectors (propagation directions)**.

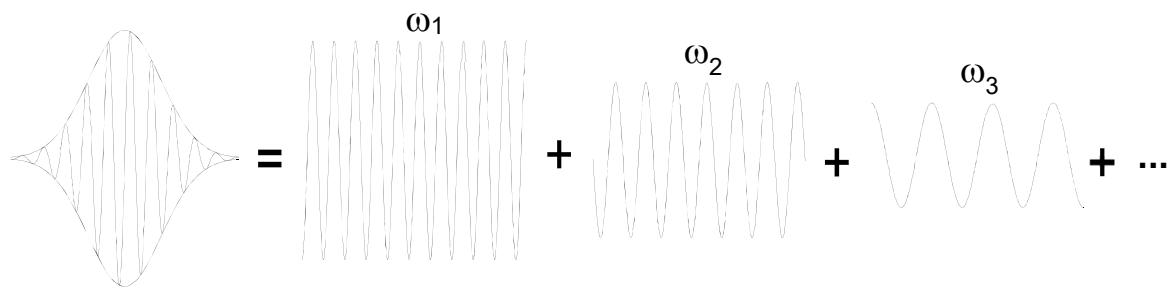


$$\mathbf{E}(\mathbf{r}, t) = \int_{-\infty}^{\infty} \bar{\mathbf{E}}(\mathbf{k}) \exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)] d^3k$$

B) pulse → finite duration → dispersion



A pulse is a continuous superposition of stationary plane waves (normal modes) with different frequencies.



$$\mathbf{E}(\mathbf{r}, t) = \int_{-\infty}^{\infty} \bar{\mathbf{E}}(\omega) \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)] d\omega.$$

C) pulsed beam → finite transverse width and finite duration → diffraction and dispersion

A pulsed beam is a continuous superposition of stationary plane waves (normal modes) with different frequencies and different wavevectors (propagation directions)

$$\mathbf{E}(\mathbf{r}, t) = \int_{-\infty}^{\infty} \bar{\mathbf{E}}(\mathbf{k}, \omega) \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)] d^3k d\omega$$

3.3 Diffraction of monochromatic beams in homogeneous isotropic media

Let us first take a look at the propagation of monochromatic beams. In this situation, we only have to deal with diffraction. We will see later that pulses and their dispersion can be treated in a very similar way. Treating diffraction in the framework of wave-optical theory (or even Maxwell's equations) allows us to rigorously describe many important optical systems and effects, i.e., optical imaging and its resolution limit as well as optical filtering, microscopy, gratings etc.

In this chapter, we assume stationary (monochromatic) fields and therefore $\omega = \text{const}$. For technical convenience and because it is sufficient for many important problems, we will make the following assumptions and approximations:

- $\epsilon(\omega) = \epsilon'(\omega) > 0$, → optically transparent regime → normal modes are stationary homogeneous and evanescent plane waves
- scalar approximation

$$\bar{E}(\mathbf{r}, \omega) \rightarrow \bar{E}_y(\mathbf{r}, \omega) \mathbf{e}_y \rightarrow \bar{E}_y(\mathbf{r}, \omega) \rightarrow u(\mathbf{r}, \omega).$$

- exact for one-dimensional beams and linear polarization (see also section 2.1.5)
- approximation in two-dimensional case

In homogeneous, isotropic media, we have to solve the Helmholtz equation

$$\Delta \bar{E}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \epsilon(\omega) \bar{E}(\mathbf{r}, \omega) = 0.$$

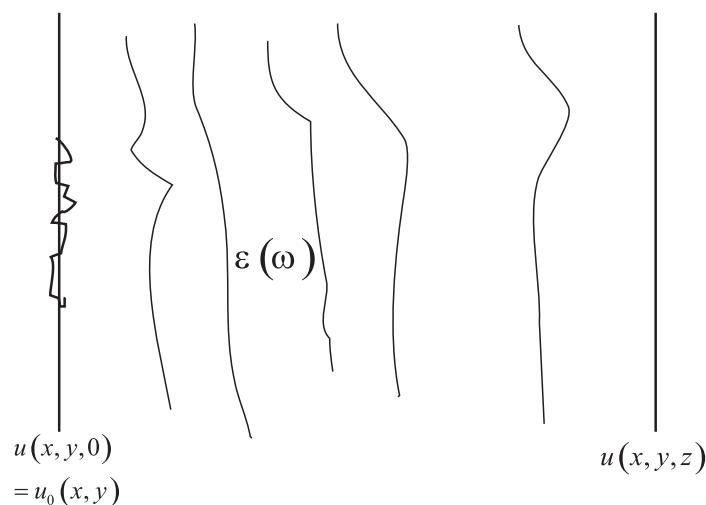
Under the scalar approximation and at fixed frequency ω , the Helmholtz equation reads

$$\begin{aligned} \Delta u(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \epsilon(\omega) u(\mathbf{r}, \omega) &= 0, && \text{scalar Helmholtz equation} \\ \Delta u(\mathbf{r}, \omega) + k^2(\omega) u(\mathbf{r}, \omega) &= 0. \end{aligned}$$

In the last step, we inserted the dispersion relation (wave number $k(\omega)$). In the following, we will often omit the argument ω in our notation since the frequency ω is fixed anyway.

3.3.1 Arbitrarily narrow beams (general case)

Let us consider the following fundamental problem. From a given field distribution $u(x, y, 0)$ in the plane $z=0$, we wish to compute the complete field $u(x, y, z)$ in the half-space $z>0$, where z is our “propagation direction”.



The governing equation describing this z -dependence of the field distribution in x and y is the scalar Helmholtz equation

$$\Delta u(\mathbf{r}, \omega) + k^2(\omega) u(\mathbf{r}, \omega) = 0.$$

To solve this equation and to calculate the dynamics of the fields, we can switch again to the Fourier domain. However, this time we switch to the spatial Fourier domain (rather than temporal).

Let's first test what happens if we take the spatial Fourier transform over all three spatial coordinates:

$$u(\mathbf{r}, \omega) = \int_{-\infty}^{\infty} U(\mathbf{k}, \omega) \exp[i\mathbf{k}(\omega)\mathbf{r}] d^3k.$$

Note the absence of the negative sign in the exponential in our definition of the spatial Fourier transform when compared with the temporal Fourier transform. This Fourier transform can be interpreted as a superposition of normal modes with different propagation directions and wavenumbers $k(\omega)$ (here the absolute value of the wavevector \mathbf{k}). Naively, we could expect that we have constructed the general solution of our problem, but the solution is incorrect because the normal modes must obey the dispersion relation

$$\mathbf{k}^2 = k^2 = k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2} \epsilon(\omega).$$

Thus, out of the three vector components of \mathbf{k} , only two can be chosen freely, e.g. k_x and k_y .

From now on we will use the following naming convention for the three wavevector components in Cartesian coordinates:

$$k_x = \alpha, \quad k_y = \beta, \quad k_z = \gamma.$$

Then, the dispersion relation reads:

$$k^2(\omega) = \alpha^2 + \beta^2 + \gamma^2.$$

Hence, our previous attempt of a Fourier transform of all three spatial coordinates x, y, z is overdetermined. This is because the dispersion relation is not fulfilled by all combinations of k_x, k_y, k_z if the temporal frequency ω is fixed as well. Thus, to solve our problem, we only need a two-dimensional Fourier transform with respect to the transverse coordinates x and y , with z as the direction of propagation:

$$u(\mathbf{r}) = \iint_{-\infty}^{\infty} U(\alpha, \beta; z) \exp[i(\alpha x + \beta y)] d\alpha d\beta.$$

It will also be a convention for this section that we indicate the fields in the Fourier space by capital letters and the fields in real space by small letters, instead of using \bar{u} for the Fourier space. Furthermore, please be aware that $U(\alpha, \beta; z)$ is a mixed representation. It is in Fourier space with respect to x and y and thus depends on α and β . But it is still in real space with respect to z .

In analogy to the frequency ω , we call α, β **spatial frequencies**.

Now we insert this expression into the scalar Helmholtz equation

$$\Delta u(\mathbf{r}) + k^2(\omega)u(\mathbf{r}) = 0.$$

This way we can transfer the Helmholtz equation in two spatial dimensions into Fourier space

$$\left(\frac{d^2}{dz^2} + k^2 - \alpha^2 - \beta^2 \right) U(\alpha, \beta; z) = 0,$$

$$\left(\frac{d^2}{dz^2} + \gamma^2 \right) U(\alpha, \beta; z) = 0.$$

where $\gamma(\alpha, \beta) = \sqrt{k^2(\omega) - \alpha^2 - \beta^2}$ is the constrained spatial frequency in the z-direction.

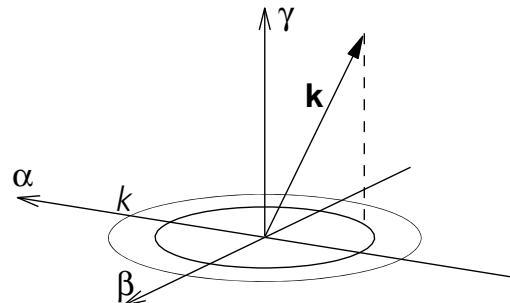
This turns the scalar Helmholtz equation from a partial differential equation into an ordinary differential equation in z , which can be easily solved to yield the general solution

$$U(\alpha, \beta; z) = U_1(\alpha, \beta) \exp[i\gamma(\alpha, \beta)z] + U_2(\alpha, \beta) \exp[-i\gamma(\alpha, \beta)z]. \quad (*)$$

Depending on α and β , we can identify two types of solutions:

A) Homogeneous waves

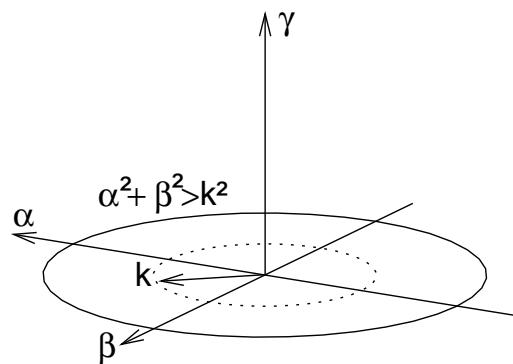
$$\gamma^2 \geq 0, \rightarrow \alpha^2 + \beta^2 \leq k^2 \rightarrow \mathbf{k} \text{ real} \rightarrow \text{homogeneous waves}$$



B) Evanescent waves

$\gamma^2 < 0, \rightarrow \alpha^2 + \beta^2 > k^2 \rightarrow \mathbf{k} \text{ complex, because } \gamma = k_z \text{ is imaginary. We then have } \mathbf{k} = \mathbf{k}' + i\mathbf{k}'', \text{ with } \mathbf{k}' = \alpha \mathbf{e}_x + \beta \mathbf{e}_y \text{ and } \mathbf{k}'' = \gamma \mathbf{e}_z.$

$\rightarrow \mathbf{k}' \perp \mathbf{k}'' \rightarrow \text{evanescent waves}$



We see immediately that in the solution (*), the second term $\sim \exp(-i\gamma z)$ grows exponentially in the half-space $z > 0$. Because this does not make physical sense, this component of the solution must vanish and hence $U_2(\alpha, \beta) = 0$. In fact, we will see later that $U_2(\alpha, \beta)$ corresponds to backward running waves, i.e., light propagating in the opposite direction. We therefore reduce the general solution (*) to

$$U(\alpha, \beta; z) = U_1(\alpha, \beta) \exp[i\gamma(\alpha, \beta)z].$$

Furthermore, the following boundary (initial) condition holds

$$U_1(\alpha, \beta) = U(\alpha, \beta; 0) = U_0(\alpha, \beta),$$

which determines uniquely the entire solution for all other positions $z \neq 0$ as well

$$\begin{aligned} U(\alpha, \beta; z) &= U(\alpha, \beta; 0) \exp[i\gamma(\alpha, \beta)z], \\ &\doteq U_0(\alpha, \beta) \exp[i\gamma(\alpha, \beta)z]. \end{aligned}$$

We can determine the optical field solution for $z > 0$ in the original real space by inverse Fourier transform:

$$u(\mathbf{r}) = \iint_{-\infty}^{\infty} U(\alpha, \beta; z) \exp[i(\alpha x + \beta y)] d\alpha d\beta.$$

$$u(\mathbf{r}) = \iint_{-\infty}^{\infty} U_0(\alpha, \beta) \exp[i\gamma(\alpha, \beta)z] \exp[i(\alpha x + \beta y)] d\alpha d\beta.$$

For homogeneous waves (real γ), the red term above causes a certain phase shift for the corresponding plane wave during propagation. Furthermore, for evanescent fields (imaginary γ), the red term causes a decay of the fields along the propagation direction. Hence, we can formulate the following result:

Diffraction is due to **different phase shifts and amplitudes** of the different excited normal modes when these modes are propagating. The individual phase shifts and amplitude reductions are determined according to different spatial frequencies α, β of the normal modes.

The initial **spatial frequency spectrum** or **angular spectrum** at $z = 0$ forms the initial condition of the initial value problem and follows from $u_0(x, y) = u(x, y, 0)$ by Fourier transform:

$$U_0(\alpha, \beta) = \left(\frac{1}{2\pi} \right)^2 \iint_{-\infty}^{\infty} u_0(x, y) \exp[-i(\alpha x + \beta y)] dx dy.$$

As mentioned above, the wave-vector components α, β are the so-called spatial frequencies. Another common terminology for the quantities $\alpha/k, \beta/k$ is “direction cosine” due to the direct connection with the propagation direction of the corresponding plane wave. For example $\alpha/k = \cos\theta_x$ gives the angle between the plane wave's propagation direction and the x -axis.

Scheme for calculation of beam diffraction

We can formulate a general scheme to describe the diffraction of beams:

1. Start from initial field $u_0(x, y)$
2. Calculate initial angular spectrum $U_0(\alpha, \beta)$ by Fourier transform
3. Describe propagation by multiplication with $\exp[i\gamma(\alpha, \beta)z]$
4. Obtain new spectrum $U(\alpha, \beta; z) = U_0(\alpha, \beta) \exp[i\gamma(\alpha, \beta)z]$
5. Calculate new field distribution $u(x, y, z)$ by Fourier back transform

This scheme allows for two interpretations:

- 1) The resulting field distribution is the Fourier transform of the propagated spectrum

$$u(\mathbf{r}) = \iint_{-\infty}^{\infty} U(\alpha, \beta; z) \exp[i(\alpha x + \beta y)] d\alpha d\beta.$$

- 2) The resulting field distribution is a superposition, i.e. interference, of homogeneous and evanescent plane waves ('plane-wave spectrum'), which obey the dispersion relation

$$u(\mathbf{r}) = \underbrace{\iint_{-\infty}^{\infty} d\alpha d\beta}_{\text{interference of eigenstates to form the field pattern after propagation}} \underbrace{U_0(\alpha, \beta)}_{\text{amplitude of the excited eigenstates}} \underbrace{\exp\{i\gamma(\alpha, \beta)z\}}_{\text{phase factor which is accumulated by the eigenstates during propagation}} \underbrace{\exp\{i[\alpha x + \beta y]\}}_{\text{shape of eigenstates (plan waves)}}.$$

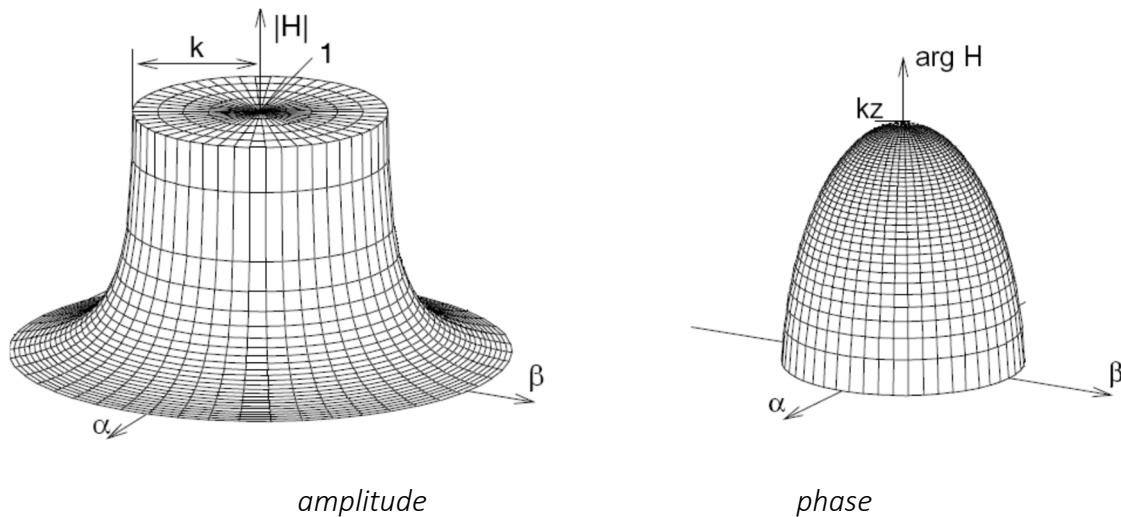
Remark: The excited eigenmodes do not exchange energy/information, they simply superimpose (interfere).

The complex transfer function for homogeneous space

To understand the diffraction of beams, let us now discuss the complex transfer function of homogeneous space, which can be extracted from the above formula as

$$\begin{aligned} H(\alpha, \beta; z) &= \exp[i\gamma(\alpha, \beta)z], \\ &= \exp\left[i\sqrt{k^2(\omega) - \alpha^2 - \beta^2}z\right]. \end{aligned}$$

It governs beam propagation in the Fourier space. For a finite propagation distance $z = \text{const.}$ it looks like:



Obviously, $H(\alpha, \beta; z) = \exp[\mathbf{i}\gamma(\alpha, \beta)z]$ acts differently on homogeneous and evanescent waves:

A) Homogeneous waves for $\alpha^2 + \beta^2 < k^2$

$$|\exp[\mathbf{i}\gamma(\alpha, \beta)z]| = 1, \quad \arg(\exp[\mathbf{i}\gamma(\alpha, \beta)z]) \neq 0$$

→ Upon propagation, the homogeneous waves are multiplied by the phase factor

$$\exp\left[\mathbf{i}\sqrt{k^2 - \alpha^2 - \beta^2}z\right].$$

→ Each homogeneous wave keeps its amplitude.

→ Homogeneous waves transport energy/information.

B) Evanescent waves for $\alpha^2 + \beta^2 \geq k^2$

$$|\exp[\mathbf{i}\gamma(\alpha, \beta)z]| = \exp\left[-\sqrt{\alpha^2 + \beta^2 - k^2}z\right], \quad \arg(\exp[\mathbf{i}\gamma(\alpha, \beta)z]) = 0$$

→ Upon propagation, the evanescent waves are multiplied by an amplitude factor ≤ 1

$$\exp\left[-\sqrt{\alpha^2 + \beta^2 - k^2}z\right] \leq 1.$$

This means that their contributions are exponentially suppressed with increasing propagation distance z .

→ Each evanescent wave keeps its phase.

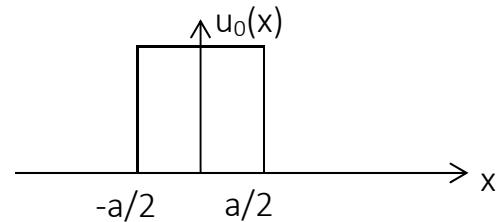
→ Evanescent waves do not transport energy/information.

Now the question is: When do we excite evanescent waves? Obviously, the answer lies in the initial condition, i.e. the boundary condition at $z = 0$. Whenever $u_0(x, y)$ yields an angular spectrum with components $U_0(\alpha, \beta) \neq 0$ for $\alpha^2 + \beta^2 > k^2$, we excite evanescent waves.

Example: Slit

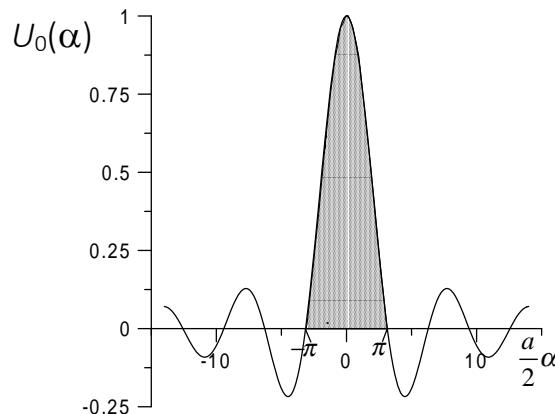
Let us consider the following one-dimensional initial condition, which corresponds to the aperture of a slit:

$$u_0(x) = \begin{cases} 1 & \text{for } |x| \leq \frac{a}{2}, \\ 0 & \text{otherwise} \end{cases}$$



We assume that the slit is infinitely extended in the y-direction. Hence, we only need to study the dynamics in the x-z-plane.

$$U_0(\alpha) = \text{FT}[u_0(x)] \sim \frac{\sin\left(\frac{a}{2}\alpha\right)}{\left(\frac{a}{2}\alpha\right)} = \text{sinc}\left(\frac{a}{2}\alpha\right)$$



- Since the sinc function has infinite support, all spatial frequencies α ($-\infty \rightarrow \infty$) are excited.
- Important spectral information is contained in the interval $|\alpha| \leq 2\pi/a$ (region between the first zeroes from zero in the above graph).
 - Most important spectral frequency for a structure with width a is $\alpha = 2\pi/a$.
- Evanescent waves appear for $|\alpha| > k$.

- To represent the most important information by homogeneous waves, the

$$\text{following condition must be fulfilled: } \frac{2\pi}{a} < k = \frac{2\pi}{\lambda} n \rightarrow a > \frac{\lambda}{n}$$

General result

We have seen in the example above that evanescent waves appear for **structures < wavelength** in the initial condition. Information about these small structures gets **lost** for propagation distances $z \gg \lambda$.

In homogeneous media, only information about structural features with length scales $|\Delta x|, |\Delta y| > \lambda/n$ are transmitted over macroscopic distances. → Homogeneous media act as **low-pass filters** for light.

Summary of the beam propagation scheme

$$u_0(x, y) \xrightarrow{\text{FT}^{-1}} U_0(\alpha, \beta) \rightarrow U(\alpha, \beta; z) = H(\alpha, \beta; z)U_0(\alpha, \beta) \xrightarrow{\text{FT}} u(x, y, z)$$

with the transfer function $H(\alpha, \beta; z) = \exp[i\gamma(\alpha, \beta)z]$

Remark: diffraction free beams

With our understanding of diffraction, it is straightforward to construct so-called diffraction free beams, i.e., beams that do not change their amplitude distribution during propagation. In Fourier space this means that all spatial frequency components must accumulate the same phase shift during propagation. This condition assures that the superposition, i.e. interference, of the excited normal modes will always be the same. The propagation then simply amounts to a phase shift in real space:

$$U(\alpha, \beta; z) = U_0(\alpha, \beta) \exp[i\gamma(\alpha, \beta)z] \equiv U_0(\alpha, \beta) \exp[iCz],$$

$$\rightarrow u(x, y, z) = \exp[iCz]u_0(x, y).$$

Since in general $\gamma(\alpha, \beta) \neq \text{const}$, the excitation $u_0(x, y)$ must have a shape such that its Fourier transform only has components where the transfer function takes the same value

$$U_0(\alpha, \beta) \neq 0 \text{ only for } \gamma(\alpha, \beta) = \sqrt{k^2 - \alpha^2 - \beta^2} = C.$$

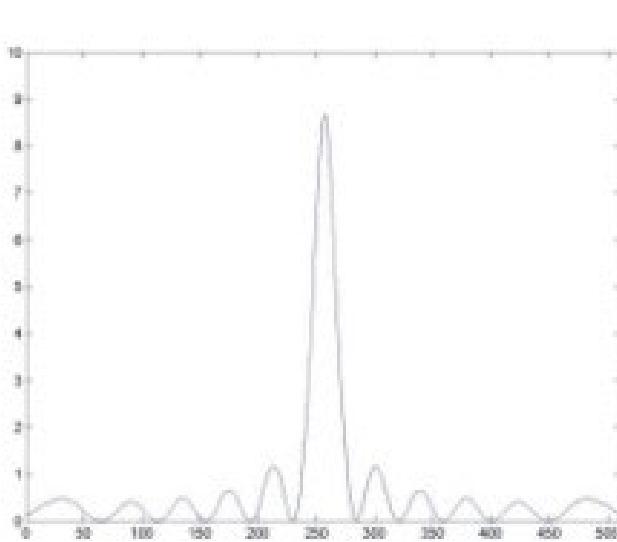
It is straightforward to see that the excited spatial frequencies must lie on a circular ring in the (α, β) plane

$$\alpha^2 + \beta^2 = \rho_0^2.$$

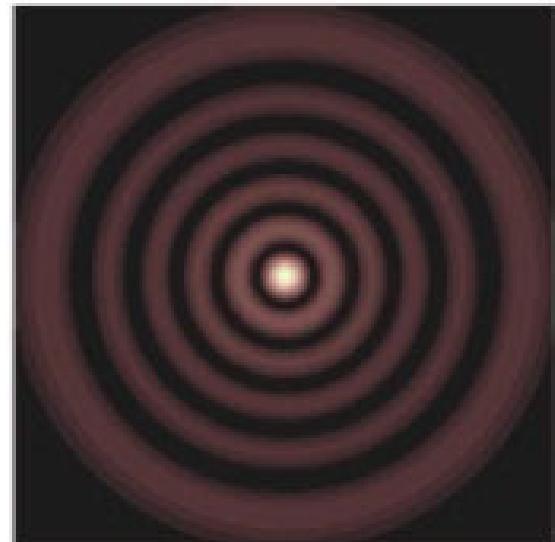
To find out which field distributions $u_0(x, y)$ have a circular Fourier transform, we can just compute the inverse Fourier transform of a circular angular spectrum distribution. For constant spectral amplitude on this ring the Fourier back-transform yields (see exercises):

$$u_0(x, y) = J_0(\rho r).$$

This corresponds to a so-called Bessel beam.



Bessel-beam (profile)



Bessel-beam

3.3.2 Fresnel- (paraxial) approximation

The beam propagation formalism developed in the previous chapter can be simplified for the important special case of a narrowband angular spectrum

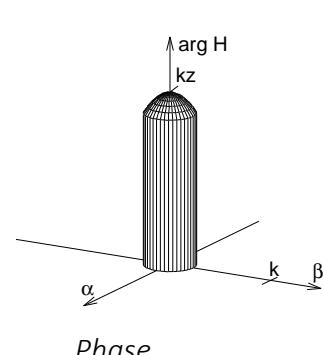
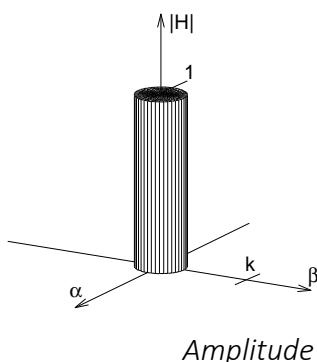
$$U_0(\alpha, \beta) \neq 0 \text{ only for } \alpha^2 + \beta^2 \ll k^2.$$

In this situation, the beam consists of plane waves having only small inclination with respect to the optical z -axis (**paraxial** (Fresnel) approximation). Then, we can simplify the expression for $\gamma(\alpha, \beta)$ by a Taylor expansion to:

$$\gamma(\alpha, \beta) = \sqrt{k^2 - \alpha^2 - \beta^2} \approx k \left(1 - \frac{\alpha^2 + \beta^2}{2k^2} \right) = k - \frac{\alpha^2 + \beta^2}{2k}$$

The resulting expression for the transfer function in Fresnel approximation reads:

$$H = \exp(i\gamma(\alpha, \beta)z) \approx \exp(ikz) \exp\left(-i\frac{\alpha^2 + \beta^2}{2k}z\right) = H_F(\alpha, \beta; z)$$



We can see that the absolute value of $H_F(\alpha, \beta; z)$ is always one. Hence, it does not account for the physics of evanescent waves. However, we must remember that the derivation of $H_F(\alpha, \beta; z)$ as an approximation of $H(\alpha, \beta; z)$ required the assumption that the spatial frequency spectrum is narrow (paraxial waves). Thus, we had already excluded the excitation of evanescent waves in justifying the paraxial approximation.

The assumption of a narrow frequency spectrum corresponds to the requirement that all structural features $|\Delta x|, |\Delta y|$ of the field distribution in the excitation plane (at $z = 0$) must be much larger than the wavelength:

$$|\Delta x|, |\Delta y| > 10 \frac{\lambda}{n} \gg \frac{\lambda}{n}$$

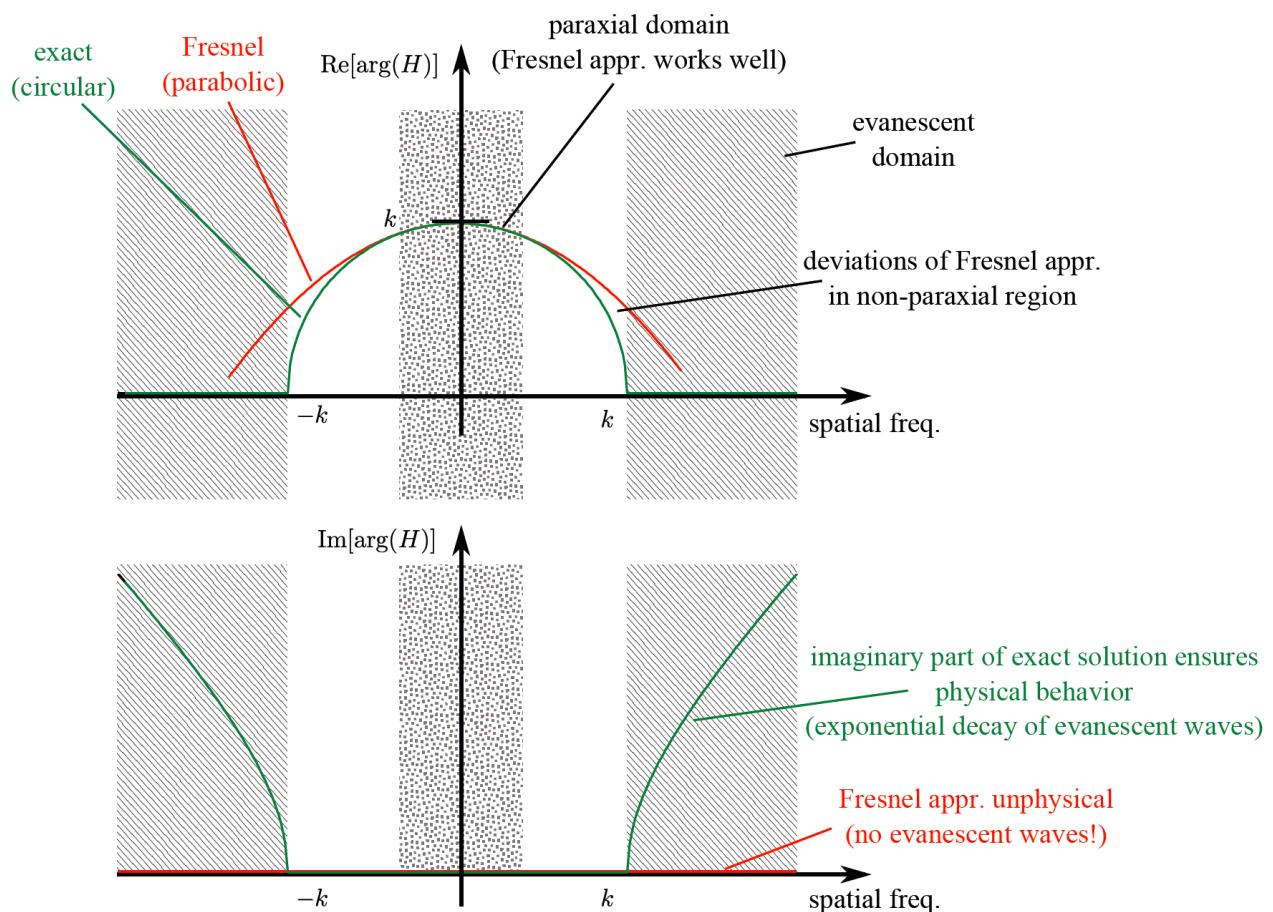
This requirement applies also to the phase of the excitation. Hence it is not sufficient that only the structural features of the intensity have a large scale. The underlying phase of the excitation field must fulfill this condition as well. This includes the condition that the phase of the beam should not have a strong inclination to the principal propagation direction, which in our notation is the positive z-direction. Also, backward propagation is not allowed.

The propagation of the spectrum in Fresnel approximation works in complete analogy to the general case. We simply use the modified transfer function to describe the propagation:

$$U_F(\alpha, \beta; z) = H_F(\alpha, \beta; z) U_0(\alpha, \beta)$$

Summary of Fresnel approximation

For a coarse initial field distribution $u_0(x, y, z)$, the angular spectrum $U_0(\alpha, \beta)$ is nonzero for $\alpha^2 + \beta^2 \ll k^2$ only. In this case, only paraxial plane waves are relevant for transmitting information and the transfer function of homogeneous space can be approximated by $H_F(\alpha, \beta; z)$.



Description in real space

It is also possible to formulate beam propagation in Fresnel (paraxial) approximation in position space:

$$\begin{aligned} u_F(x, y, z) &= \iint_{-\infty}^{\infty} U_F(\alpha, \beta; z) \exp[i(\alpha x + \beta y)] d\alpha d\beta \\ &= \iint_{-\infty}^{\infty} H_F(\alpha, \beta; z) U_0(\alpha, \beta) \exp[i(\alpha x + \beta y)] d\alpha d\beta \\ &= \iint_{-\infty}^{\infty} h_F(x - x', y - y'; z) u_0(x', y') dx' dy' \end{aligned}$$

The spatial response function $h_F(x, y; z)$ follows from the convolution theorem and is the Fourier transform of $H_F(\alpha, \beta; z)$:

$$\begin{aligned} h_F(x, y; z) &= \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} H_F(\alpha, \beta; z) \exp[i(\alpha x + \beta y)] d\alpha d\beta \\ &= \left(\frac{1}{2\pi}\right)^2 \exp(ikz) \iint_{-\infty}^{\infty} \exp\left(-i\frac{\alpha^2 + \beta^2}{2k} z\right) \exp[i(\alpha x + \beta y)] d\alpha d\beta. \end{aligned}$$

This Fourier integral can be solved as a Gaussian integral and we find:

$$h_{\text{F}}(x, y; z) = \exp(\mathbf{i}kz) \left\{ -\frac{\mathbf{i}k}{2\pi z} \exp\left[\mathbf{i}\frac{k}{2z}(x^2 + y^2)\right] \right\} = -\frac{\mathbf{i}k}{2\pi z} \exp\left\{\mathbf{i}kz \left[1 + \frac{(x^2 + y^2)}{2z^2}\right]\right\},$$

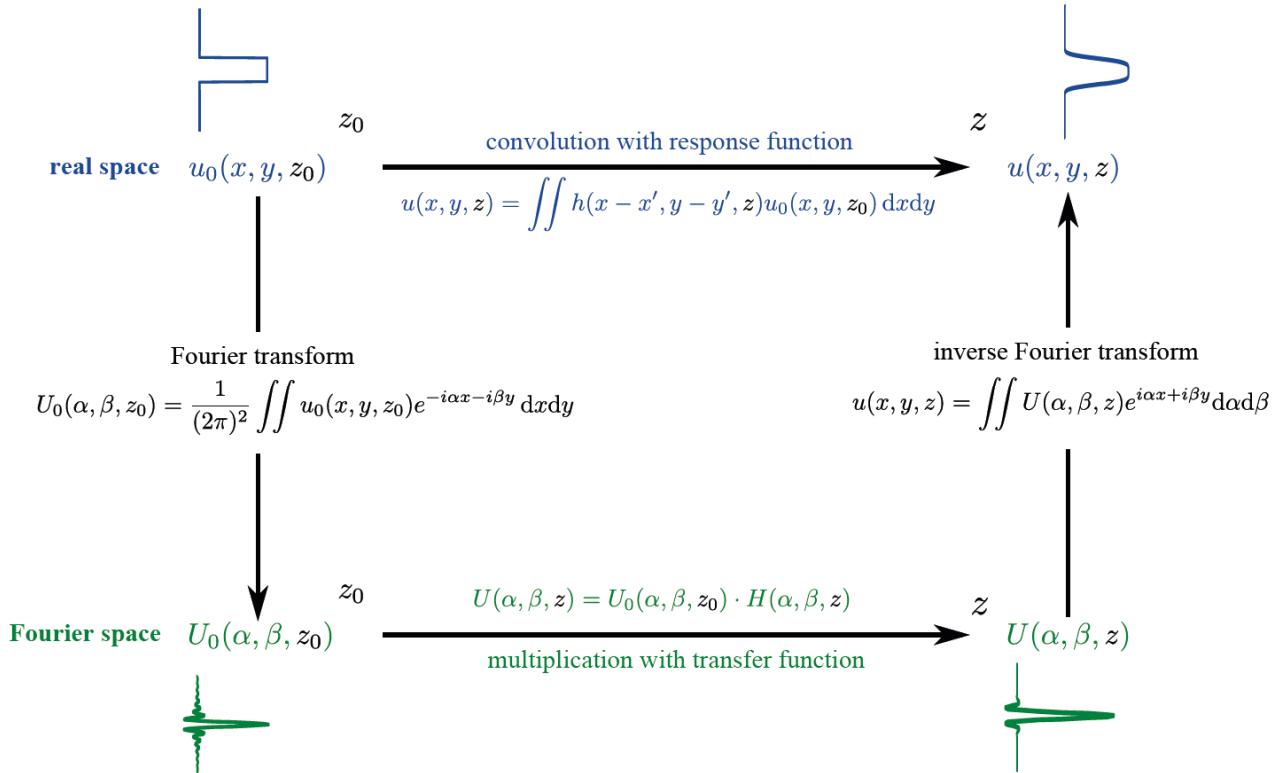
The response function in the paraxial approximation corresponds to a spherical wave. Similar to Huygens principle, where from each point in the object plane a spherical wave is emitted towards the image plane, here paraxial approximations of spherical waves are emitted.

To sum up, in position space paraxial beam propagation is given by the convolution of the initial field with the response function:

$$u_{\text{F}}(x, y, z) = -\frac{\mathbf{i}k}{2\pi z} \exp(\mathbf{i}kz) \iint_{-\infty}^{\infty} u_0(x', y') \exp\left\{\mathbf{i}\frac{k}{2z}[(x-x')^2 + (y-y')^2]\right\} dx' dy'.$$

Of course, the two descriptions in position space and in the spatial Fourier domain are completely equivalent.

The correspondence between real and frequency space



Relation between transfer and response function:

$$h(x, y; z) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} H(\alpha, \beta; z) \exp[\mathbf{i}(\alpha x + \beta y)] d\alpha d\beta$$

Transfer functions for homogeneous space

$$H(\alpha, \beta; z) = \exp[\mathbf{i}\gamma(\alpha, \beta)z] = \exp[i\sqrt{k^2 - \alpha^2 - \beta^2}z] \text{ exact solution}$$

$$H_F(\alpha, \beta; z) = \exp[i\mathbf{k}z] \exp\left[-i\frac{\alpha^2 + \beta^2}{2k}z\right] \text{ Fresnel approximation}$$

$$\text{with } k = k(\omega) = k_0 n(\omega) = \frac{\omega}{c} n(\omega)$$

Remark on the validity of the scalar approximation

In the previous description of the propagation of arbitrary beams, we have used the scalar approximation of the vectorial fields. It is interesting to investigate to what extent this approximation is consistent the conditions which were necessary to derive the Fresnel approximation.

$$\bar{\mathbf{E}}(\mathbf{r}, \omega) = \iint \hat{\mathbf{E}}_0(\alpha, \beta, \omega) e^{i(\alpha x + \beta y + \gamma z)} d\alpha d\beta$$

From the divergence condition, we can derive conditions on the vectorial field components as

$$\operatorname{div} \bar{\mathbf{E}}(\mathbf{r}, \omega) = 0 \rightarrow \alpha \hat{E}_x + \beta \hat{E}_y + \gamma \hat{E}_z = 0.$$

A) One-dimensional beams

- translational invariance in y-direction: $\beta = 0, \alpha \neq 0, \gamma \neq 0$
 - and linear polarization in y-direction: $\hat{E}_y = U, \hat{E}_x = 0, \hat{E}_z = 0$
- scalar approximation is exact since divergence condition is fulfilled exactly

B) Two-dimensional beams

- Finite beam which is localized in the x,y-plane: $\alpha, \beta \neq 0$
 - and linear polarization, w.l.o.g. in y-direction: $\hat{E}_x = 0, \hat{E}_y = U$
- divergence condition: $\beta \hat{E}_y + \gamma \hat{E}_z = 0$

$$\hat{E}_z(\alpha, \beta, \omega) = -\frac{\beta}{\gamma} \hat{E}_y(\alpha, \beta, \omega) = -\underbrace{\frac{\beta}{\sqrt{k^2 - \alpha^2 - \beta^2}}}_{\approx 0} \hat{E}_y(\alpha, \beta, \omega) \approx 0$$

In the paraxial approximation where $\alpha^2 + \beta^2 \ll k^2$, the scalar approximation is thus automatically justified.

3.3.3 Paraxial wave equation

In the paraxial approximation, the propagated spectrum is given by

$$\begin{aligned} U_F(\alpha, \beta; z) &= H_F(\alpha, \beta; z) U_0(\alpha, \beta) \\ &= \exp(i\mathbf{k}z) \exp\left(-i\frac{\alpha^2 + \beta^2}{2k}z\right) U_0(\alpha, \beta) \end{aligned}$$

Let us introduce the slowly varying spectrum $V(\alpha, \beta; z)$ after isolating the fast oscillations:

$$\rightarrow U_F(\alpha, \beta; z) = \exp(\mathbf{i}kz) V(\alpha, \beta; z) \rightarrow V(\alpha, \beta; z) = \exp\left(-\mathbf{i}\frac{\alpha^2 + \beta^2}{2k}z\right) V_0(\alpha, \beta).$$

Differentiation of V with respect to z gives:

$$\mathbf{i} \frac{\partial}{\partial z} V(\alpha, \beta; z) = \frac{1}{2k} (\alpha^2 + \beta^2) V(\alpha, \beta; z)$$

Fourier transformation with respect to α and β back to x and y in position space leads to the so-called paraxial wave equation:

$$\begin{aligned} \mathbf{i} \frac{\partial}{\partial z} \iint_{-\infty}^{\infty} V(\alpha, \beta; z) \exp[\mathbf{i}(\alpha x + \beta y)] d\alpha d\beta \\ = \frac{1}{2k} \iint_{-\infty}^{\infty} (\alpha^2 + \beta^2) V(\alpha, \beta; z) \exp[\mathbf{i}(\alpha x + \beta y)] d\alpha d\beta \\ \rightarrow \mathbf{i} \frac{\partial}{\partial z} v(x, y, z) = -\frac{1}{2k} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \iint_{-\infty}^{\infty} V(\alpha, \beta; z) \exp[\mathbf{i}(\alpha x + \beta y)] d\alpha d\beta \\ \rightarrow \boxed{\mathbf{i} \frac{\partial}{\partial z} v(x, y, z) + \frac{1}{2k} \Delta^{(2)} v(x, y, z) = 0} \text{ paraxial wave equation} \end{aligned}$$

The slowly varying envelope $v(x, y, z)$ (Fourier transform of the slowly varying spectrum) relates to the scalar field by $u_F(x, y, z) = v(x, y, z) \exp(\mathbf{i}kz)$.

Extension of the wave equation to weakly inhomogeneous media

(slowly varying envelope approximation - SVEA)

There is an alternative, more general way to derive the paraxial wave equation, the so-called slowly varying envelope approximation. This approximation even allows us to treat wave propagation in **inhomogeneous media**. We will include inhomogeneous media in this derivation even though the current chapter of this lecture is devoted to homogeneous media.

We start from the scalar Helmholtz equation. However, we should mention that the generalization of our previous discussion to inhomogeneous media requires another approximation. This approximation assumes small spatial fluctuations of $\epsilon(\mathbf{r}, \omega)$ since otherwise the grad-div-term in the general wave equation would not disappear. This is equivalent to having a weak index contrast between different spatial positions.

$$\Delta u(x, y, z) + k^2(\mathbf{r}, \omega) u(x, y, z) = 0 \text{ with } k^2(\mathbf{r}, \omega) = \frac{\omega^2}{c^2} \epsilon(\mathbf{r}, \omega)$$

We use the ansatz of a slowly varying envelope

$$u(x, y, z) = v(x, y, z) \exp(\mathbf{i}\tilde{k}z) \text{ with } \tilde{k} = \langle k \rangle$$

where \tilde{k} is the spatially averaged wavenumber. This is the mean wavenumber in the particular volume of material where wave propagation is considered. Hence, it is the average of a spatially varying material property within the volume of interest.

$$\frac{\partial^2}{\partial z^2}v(x,y,z) + 2\mathbf{i}\tilde{k}\frac{\partial}{\partial z}v(x,y,z) + \Delta^{(2)}v(x,y,z) + \left[k^2(\mathbf{r},\omega) - \tilde{k}^2 \right]v(x,y,z) = 0,$$

With the slowly varying envelope approximation (SVEA)

$$|\tilde{k}v| \gg |\partial v / \partial z|$$

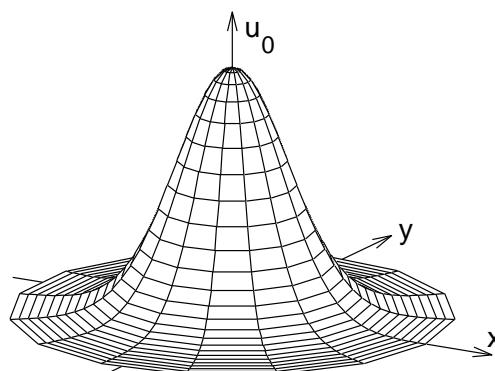
the 2nd derivative with respect to z in the first term can be neglected and we can then simplify the scalar Helmholtz equation as follows:

$$\rightarrow \boxed{\mathbf{i}\frac{\partial}{\partial z}v(x,y,z) + \frac{1}{2\tilde{k}}\Delta^{(2)}v(x,y,z) + \left[\frac{k^2(\mathbf{r},\omega) - \tilde{k}^2}{2\tilde{k}} \right]v(x,y,z) = 0}$$

This is the paraxial wave equation for inhomogeneous media having a weak index contrast.

3.4 Propagation of Gaussian beams

The propagation of Gaussian beams is an important special case. Optical beams with a Gaussian profile are important because many beams in reality have a shape which is at least close to the shape of a Gauss function. In some cases, they even have the exact Gaussian shape, e.g. the transversal fundamental modes of many lasers. The second reason why Gaussian beams are important is the fact that in the paraxial approximation, it is possible to compute the evolution of Gaussian beams analytically.



Fundamental Gaussian beam in focus

The general form of a Gaussian beam at an arbitrary plane ($z=\text{const}$) is elliptic, with a curved phase front:

$$u_0(x,y) = v_0(x,y) = A_0 \exp \left[- \left(\frac{x^2}{w_x^2} + \frac{y^2}{w_y^2} \right) \right] \exp [\mathbf{i}\phi(x,y)].$$

In this plane, the slowly varying envelope is defined to coincide with the field itself. We will restrict ourselves to excitation conditions with rotational symmetry ($w_x^2 = w_y^2 = w_0^2$) and (initially) 'flat' phase $\varphi(x, y) = 0$. We will see later that this corresponds to the focus of a Gaussian beam. The Gaussian beam in such a focal plane (flat phase) is characterized by amplitude A and width w_0 , which are defined by

$$u_0(x^2 + y^2 = w_0^2) = A_0 \exp(-1) = A_0 / e.$$

In practice, the so-called 'full width at half maximum' (FWHM) of the intensity is often used instead of w_0 . The FWHM of the intensity is connected to w_0 of the field by

$$\frac{|u_0(x^2 + y^2)|^2}{A_0^2} = \exp\left(-\frac{w_{\text{FWHM}}^2}{2w_0^2}\right) \doteq \frac{1}{2}$$

Note that the factor of 2 in the denominator of the exponent is derived from a factor of 4 in the denominator since we must divide the FWHM by 2 to compare with w_0 (plus a factor of 2 in the numerator due to the squaring of the field to obtain the intensity).

$$-\frac{w_{\text{FWHM}}^2}{2w_0^2} = -\ln 2 \rightarrow w_{\text{FWHM}}^2 = 2 \ln 2 w_0^2 \approx 1.386 w_0^2$$

3.4.1 Propagation in paraxial approximation

Let us now compute the propagation of a Gaussian beam starting from the focus in the paraxial approximation:

1) Field in focus at $z = 0$:

$$u_0(x, y) = v_0(x, y) = A_0 \exp\left(-\frac{x^2 + y^2}{w_0^2}\right).$$

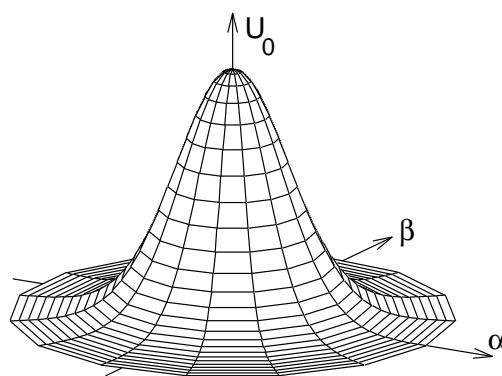
2) Angular spectrum at $z = 0$:

$$\begin{aligned} U_0(\alpha, \beta) = V_0(\alpha, \beta) &= \frac{1}{(2\pi)^2} A_0 \iint_{-\infty}^{\infty} \exp\left(-\frac{x^2 + y^2}{w_0^2}\right) \exp[-i(\alpha x + \beta y)] dx dy \\ &= \frac{A_0}{4\pi} w_0^2 \exp\left(-\frac{\alpha^2 + \beta^2}{4/w_0^2}\right) = \frac{A_0}{4\pi} w_0^2 \exp\left(-\frac{\alpha^2 + \beta^2}{w_s^2}\right), \end{aligned}$$

- We see that the angular spectrum has a Gaussian profile as well and that the widths in position space and Fourier space are linked by

$$w_s \cdot w_0 = 2.$$

- This is to be expected as the Fourier transform of a Gaussian is another Gaussian with widths that are inversely proportional to each other.



Angular spectrum in the focal plane

- We should check whether the paraxial approximation is fulfilled:

We can say that $U_0(\alpha, \beta) \approx 0$ for $(\alpha^2 + \beta^2) \geq 16/w_0^2$, because $\exp(-4) \approx 0.02$.

For the paraxial approximation, we require $k^2 \gg (\alpha^2 + \beta^2)$, where the latter represents spatial frequencies such that $U_0(\alpha, \beta) \neq 0$

$$\rightarrow k^2 \gg 16/w_0^2$$

$$\rightarrow w_0^2 \gg \frac{16}{\left(\frac{2\pi}{\lambda}n\right)^2} = \left(\frac{2\lambda}{\pi n}\right)^2 \approx \left(\frac{\lambda}{n}\right)^2,$$

$$\rightarrow \text{paraxial approximation works for } w_0 \gtrsim 10 \frac{\lambda}{n} = 10\lambda_n$$

- 3) Propagation of the angular spectrum by multiplying with the Fresnel transfer function:

$$\begin{aligned} V(\alpha, \beta; z) &= U_0(\alpha, \beta) \exp\left(-i \frac{\alpha^2 + \beta^2}{2k} z\right) \\ &= \frac{A_0}{4\pi} w_0^2 \exp\left[-w_0^2 \frac{\alpha^2 + \beta^2}{4}\right] \exp\left(-i \frac{\alpha^2 + \beta^2}{2k} z\right). \end{aligned}$$

- 4) Fourier back-transformation to position space, when still using the slowly varying envelope v

$$\begin{aligned}
 v(x, y, z) &= \frac{A_0}{4\pi} w_0^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left[-(\alpha^2 + \beta^2) \left(\frac{w_0^2}{4} + \frac{\mathbf{i}z}{2k} \right) \right] \exp [\mathbf{i}(\alpha x + \beta y)] d\alpha d\beta \\
 &= A_0 \frac{1}{1 + \frac{2\mathbf{i}z}{kw_0^2}} \exp \left[-\frac{x^2 + y^2}{w_0^2 (1 + \frac{2\mathbf{i}z}{kw_0^2})} \right] \\
 &= A_0 \frac{1}{1 + \mathbf{i} \frac{z}{z_0}} \exp \left[-\frac{x^2 + y^2}{w_0^2 (1 + \frac{\mathbf{i}z}{z_0})} \right].
 \end{aligned}$$

where we have defined the Rayleigh length z_0 as a characteristic length scale in the propagation of a Gaussian beam:

$$z_0 = \frac{kw_0^2}{2} = \frac{\pi}{\lambda_n} w_0^2.$$

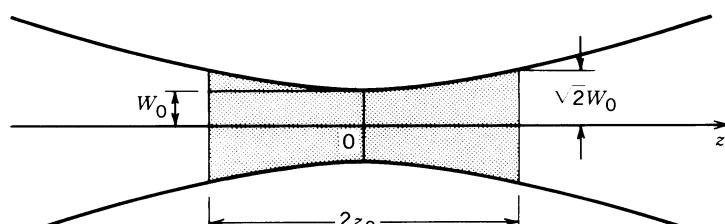
Note that we use the slowly varying envelope v !

Conclusion:

- A Gaussian beam maintains its shape, but its amplitude, width, and phase change upon propagation.
- Two important parameters characterize the beam's evolution: propagation length z and Rayleigh length z_0 .

Some books use the “diffraction length” $L_B = 2z_0$, which is a measure of the depth of focus of the Gaussian beam, where the beam is reasonably collimated E.g.:

$$w_0 \gtrsim 10\lambda_n \rightarrow L_B \gtrsim 600\lambda_n.$$



The depth of focus of a Gaussian beam.

From our computation above we know that the Gaussian beam evolves like:

$$v(x, y, z) = A_0 \frac{1}{1 + \frac{\mathbf{i}z}{z_0}} \exp \left[-\frac{x^2 + y^2}{w_0^2 (1 + \frac{\mathbf{i}z}{z_0})} \right].$$

For practical usage, we can write this expression in terms of the z-dependent amplitude, width and shape of the phase front:

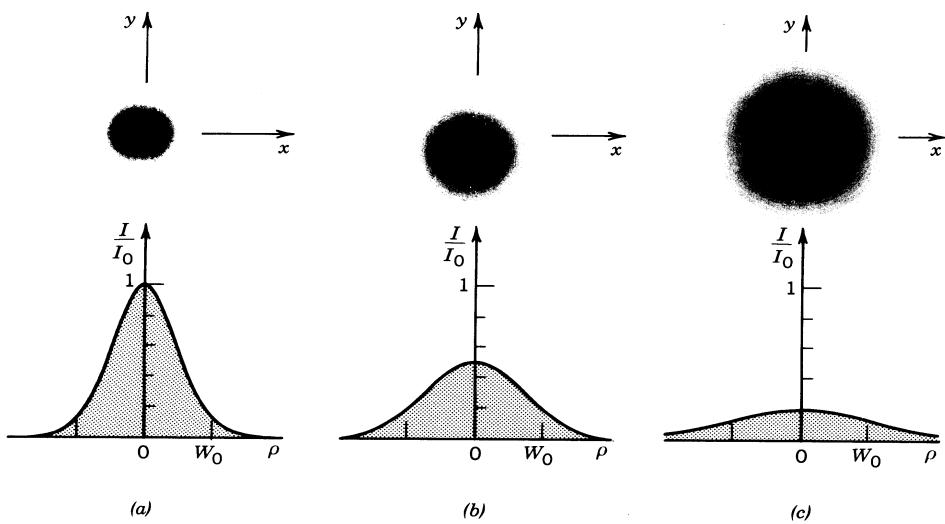
$$v(x, y, z) = A_0 \frac{1 - i \frac{z}{z_0}}{1 + (z/z_0)^2} \exp \left\{ -\frac{x^2 + y^2}{w_0^2 [1 + (z/z_0)^2]} \right\} \exp \left\{ i \frac{(x^2 + y^2)z/z_0}{w_0^2 [1 + (z/z_0)^2]} \right\}$$

$$v(x, y, z) = A_0 \frac{1}{\sqrt{1 + \left(\frac{z}{z_0}\right)^2}} \exp \left\{ -\frac{x^2 + y^2}{w_0^2 \left[1 + \left(\frac{z}{z_0}\right)^2\right]} \right\} \exp \left\{ i \frac{k}{2} \frac{(x^2 + y^2)}{z \left[1 + \left(\frac{z_0}{z}\right)^2\right]} \right\} \exp[i\varphi(z)]$$

Using that $w_0^2 = 2z_0/k$, we have separated the real and imaginary parts of the exponent to describe the evolution of the beam width and curvature of the phase fronts, respectively. Furthermore, we have isolated the amplitude and so-called x, y-independent Gouy phase shift phase $\varphi(z)$ as defined by $\tan \varphi = -z/z_0$.

In conclusion, we see that the propagation of a Gaussian beam is given by a z-dependent amplitude, width, phase curvature and phase shift:

$$v(x, y, z) = A(z) \exp \left[-\frac{x^2 + y^2}{w^2(z)} \right] \exp \left[i \frac{k}{2} \frac{(x^2 + y^2)}{R(z)} \right] \exp[i\varphi(z)]$$



The normalized beam intensity I/I_0 as a function of the radial distance ρ at different axial distances: (a) $z = 0$; (b) $z = z_0$; (c) $z = 2z_0$.

The normalized beam intensity I/I_0 as a function of the radial distance $\rho = (x^2 + y^2)^{1/2}$ at different axial distances: (a) $z = 0$; (b) $z = z_0$; (c) $z = 2z_0$.

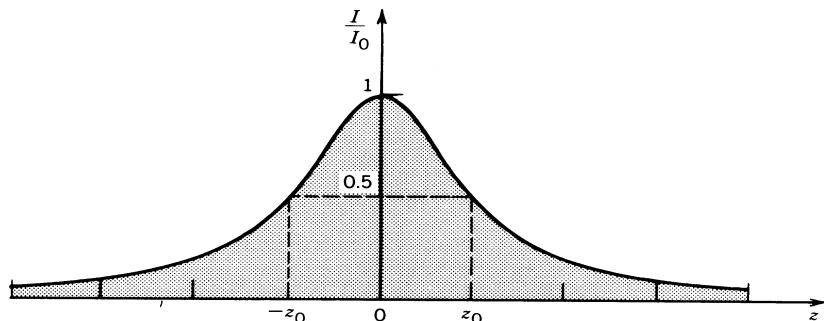
Discussion of beam parameters

Amplitude

The amplitude evolves along z as:

$$A(z) = A_0 \frac{1}{\sqrt{1 + \left(\frac{z}{z_0}\right)^2}} = A_0 \frac{1}{\sqrt{1 + \left(\frac{2z}{L_B}\right)^2}},$$

Hence, we obtain for the intensity profile $I \sim A^2$:



The normalized beam intensity I / I_0 at points on the beam axis ($\rho = 0$) as a function of z .

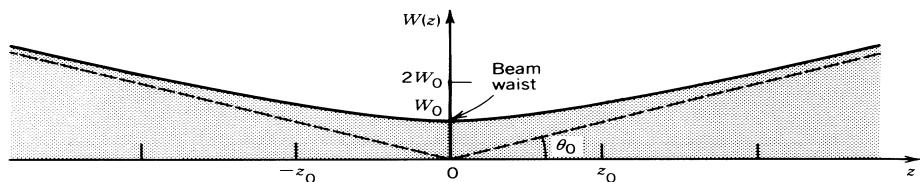
The normalized beam intensity I / I_0 at points on the beam axis ($\rho = (x^2 + y^2)^{1/2} = 0$) as a function of the propagation distance z .

Note that the intensity decreases by a factor of 2 one Rayleigh range z_0 away from the focus.

Width

The beam width evolves along z as:

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_0}\right)^2} = w_0 \sqrt{1 + \left(\frac{2z}{L_B}\right)^2}$$



The beam radius $w(z)$ has its minimum value w_0 at the waist ($z = 0$), reaches $\sqrt{2} w_0$ at $z = \pm z_0$, and increases linearly with z for large z .

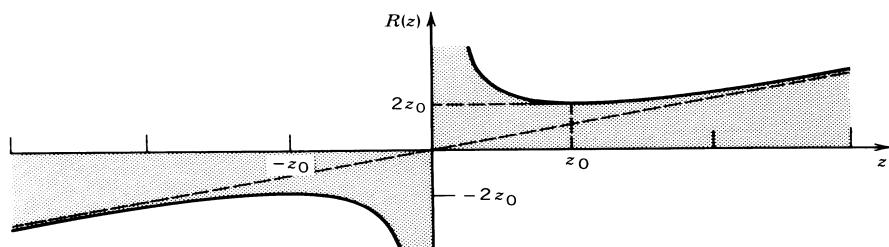
The beam radius $w(z)$ is at its minimum value w_0 at the waist ($z = 0$), reaches $\sqrt{2} w_0$ at $z = \pm z_0$, and increases linearly with z for large z .

Note that the beam width increases from its minimum value W_0 at the waist to $\sqrt{2}W_0$ one Rayleigh range z_0 away.

Phase curvature

The phase fronts of a Gaussian beam can be characterized by their radius of curvature, which is given by

$$R(z) = z \left[1 + \left(\frac{z_0}{z} \right)^2 \right] = z \left[1 + \left(\frac{L_B}{2z} \right)^2 \right]$$



The radius of curvature $R(z)$ of the wavefronts of a Gaussian beam. The dashed line is the radius of curvature of a spherical wave.

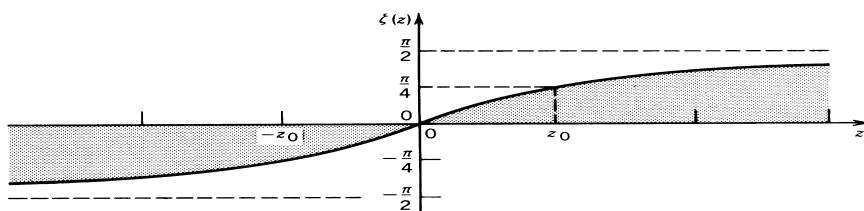
The radius of curvature $R(z)$ of the wavefronts of a Gaussian beam. The dashed line is the radius of curvature of a spherical wave.

The flat phase in the focus ($z=0$) corresponds to an infinite radius of curvature. The strongest curvature (minimum radius) appears at the Rayleigh distance z_0 from the focus. For large z , the radius of curvature increases linearly as with a spherical wave.

Gouy phase

The x, y-independent Gouy phase is given by

$$\tan \varphi = \frac{z}{z_0} = \frac{2z}{L_B}$$

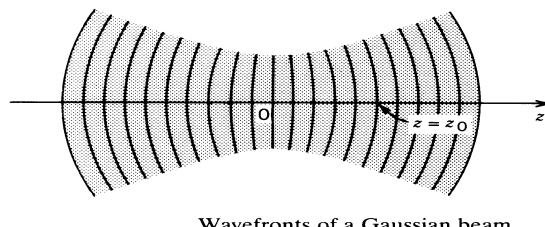


$\xi(z)$ is the phase retardation of the Gaussian beam relative to a uniform plane wave at points on the beam axis.

The so-called Gouy phase $\varphi(z)$ is the retardation of the phase of a Gaussian beam relative to a uniform plane wave at points along the beam axis.

The Gouy phase is not important for many applications because it is ‘flat’ (no transverse dependence). However, in resonators and in the context of nonlinear optics it can play an important role (i.e. harmonic generation in a focused beam geometries). In total, the wavefronts (surfaces of constant phase) of a Gaussian beam are given by

$$\Phi(x, y, z) = \left\{ k \left[z + \frac{x^2 + y^2}{2R(z)} \right] + \varphi(z) \right\} = \text{const.}$$



Wavefronts of a Gaussian beam.

Wavefronts of a Gaussian beam.

3.4.2 Propagation of Gaussian beams with q-parameter formalism

In the previous chapter we derived expressions for Gaussian beam propagation, i.e. we know how the amplitude, width, and phase change with the propagation variable z . Furthermore, we have discovered that the entire beam evolution depends only on the two parameters z and z_0 . Therefore, one could combine these two real-valued parameters into a single complex-valued parameter, which we call the q-parameter:

$$q(z) = z - \mathbf{i}z_0 \quad \text{q-parameter.}$$

This complex beam parameter allows an even simpler computation of the evolution of a Gaussian beam than the expressions of the last section. In fact, if we take the inverse of the q-parameter

$$\frac{1}{q(z)} = \frac{1}{z - \mathbf{i}z_0} = \frac{z}{z^2 + z_0^2} + \mathbf{i} \frac{z_0}{z^2 + z_0^2} = \frac{1}{z \left(1 + \frac{z_0^2}{z^2}\right)} + \mathbf{i} \frac{1}{z_0 \left(1 + \frac{z_0^2}{z^2}\right)},$$

we see that real and imaginary part are directly linked to the radius of curvature of the phase front and the beam width as given above:

$$\frac{1}{q(z)} = \frac{1}{R(z)} + \mathbf{i} \frac{\lambda_n}{\pi w^2(z)}$$

because $z_0 = \frac{kw_0^2}{2} = \frac{\pi}{\lambda_n} w_0^2$

Thus, the q-parameter describes beam propagation for all z !

Example: propagation by $z=d$ in homogeneous space

- 1) determination of initial q-parameter based on initial conditions:

$$\frac{1}{q(0)} = \frac{1}{R(0)} + \mathbf{i} \frac{\lambda_n}{\pi w^2(0)}$$

- 2) propagation (by definition, the real part of the q-parameter increases additively with z while the imaginary part is constant for a given Gaussian beam of fixed z_0):

$$q(d) = q(0) + d$$

- 3) q-parameter at $z=d$ determines new width and radius of curvature

$$\frac{1}{q(d)} = \frac{1}{q(0) + d} \doteq \frac{1}{R(d)} + \mathbf{i} \frac{\lambda_n}{\pi w^2(d)}$$

3.4.3 Gaussian optics

We have seen in the previous chapter that the complex q-parameter formalism enables a simple description of beam propagation. The question is whether it is also possible to treat optical elements (lens, mirror etc.) using this formalism. By investigating beam propagation through such elements, we will go beyond the general scope of this chapter, which was originally devoted only to the investigation of homogeneous space.

Aim: for given R_0, w_0 (i.e. q_0) → pass through n optical elements → calculate R_n, w_n (i.e. q_n)

We will evaluate the q-parameter at certain discrete positions: $q(z_i) \rightarrow q_i$.

Surprising property: We can use the same ABCD-Matrices as for geometrical ray optics!

This is remarkable because here we are doing wave-optics (with Gaussian beams).

A short reminder of geometrical ray optics

A) Propagation through one optical element:

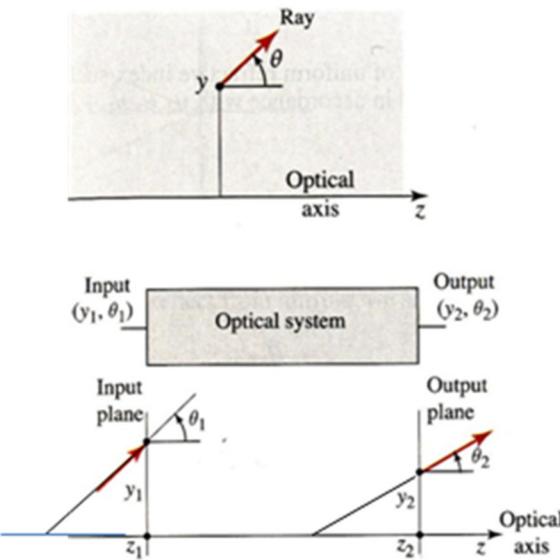
$$\hat{\mathbf{M}} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

B) Propagation through multiple elements:

$$\hat{\mathbf{M}} = \hat{\mathbf{M}}_n \hat{\mathbf{M}}_{n-1} \dots \hat{\mathbf{M}}_1 = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

C) Matrix connects the distances to the optical axis y and inclination angles Θ before and after the element

$$\begin{pmatrix} y_2 \\ \Theta_2 \end{pmatrix} = \hat{\mathbf{M}} \begin{pmatrix} y_1 \\ \Theta_1 \end{pmatrix}.$$



Link to Gaussian beams

Let us consider the distance to the intersection of the ray with the optical axis, as defined in chapter 9.6.1 on "The ray-transfer-matrix":

$$z_1 = y_1 / \tan(\Theta_1)$$

Using the small angle approximation

$$\tan(\Theta_1) \approx \Theta_1$$

we can define this distance as:

$$z_1 = \frac{y_1}{\Theta_1} \rightarrow z_2 = \frac{y_2}{\Theta_2} = \frac{Ay_1 + B\Theta_1}{Cy_1 + D\Theta_1} = \frac{A \frac{y_1}{\Theta_1} + B}{C \frac{y_1}{\Theta_1} + D} = \frac{Az_1 + B}{Cz_1 + D}$$

The distances z_1, z_2 are connected by the elements of the ABCD matrix, but not by the usual matrix multiplication.

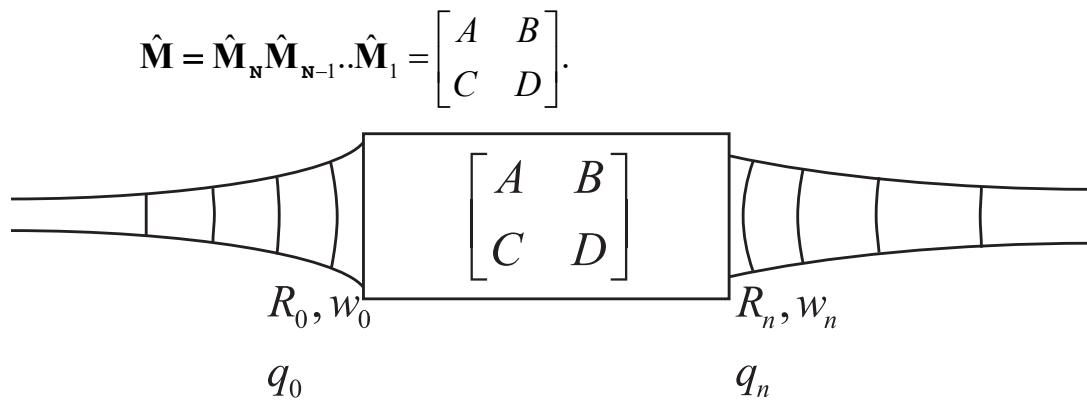
It turns out that this can be applied to Gaussian optics by replacing z with the complex beam parameter q . The propagation of the q -parameter through an optical element is then given by:

$$q_1 = \frac{A_1 q_0 + B_1}{C_1 q_0 + D_1}$$

Accordingly, the propagation through N elements can be described by

$$q_n = \frac{Aq_0 + B}{Cq_0 + D},$$

with the matrix



This works for all ABCD matrices given in chapter 8.6 for ray optics!

Here: we will check two important examples:

i) Propagation in free space over a length $z=d$:

→ propagation (by definition of q -parameter) $q(d)=q(0)+d$

$$\hat{\mathbf{M}} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \rightarrow A_l = 1, B_l = d, C_l = 0, D_l = 1$$

$$q_1 = \frac{A_l q_0 + B_l}{C_l q_0 + D_l} = \frac{q_0 + d}{0 + 1} = q_0 + d$$

ii) Thin lens with focal length f

What is the effect of a thin lens on a Gaussian beam $\exp(-(x^2 + y^2)/w_0^2)$ in the paraxial approximation?

– no change in the width

– but change in phase curvature R_f : $\times \exp\left[i \frac{k(x^2 + y^2)}{2R_f}\right]$

How can we see that?

Trick:

We start at the focus which is produced by the lens. Let w_f represent the width of this focus and $z_0 = z_f = \frac{\pi w_f^2}{\lambda_n}$ the corresponding Rayleigh range. Since the radius of curvature at the focus is infinite, the q -parameter here is

$$\frac{1}{q_0} = i \frac{\lambda_n}{\pi w_f^2}$$

The radius of curvature evolves as:

$$R(z) = z \left[1 + \left(\frac{z_f}{z} \right)^2 \right] \approx z \text{ for } z \gg z_f$$

We can invert the propagation from the focal position to the lens at a distance given by the focal length f and obtain $R_f = -f$.

Now we compare this result to the description of a thin lens using the ABCD matrix from chapter 8.6:

$$f < 0 \quad f > 0 \quad \hat{\mathbf{M}}_{\text{thin lens}} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}$$

$$q_1 = \frac{A_1 q_0 + B_1}{C_1 q_0 + D_1} = \frac{q_0}{-q_0/f + 1}$$

$$\frac{1}{q_1} = \frac{-q_0/f + 1}{q_0} = \frac{1}{-f} + \frac{i\lambda_n}{\pi w_0^2} \quad \text{for } q_0 = -i \frac{\pi w_0^2}{\lambda_n}$$

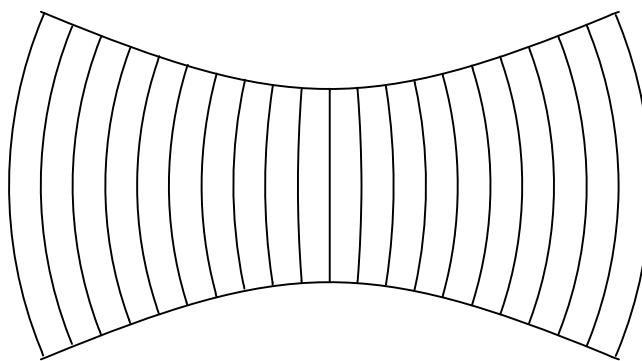
We see that we get the same result as above.

However, we must be careful when using the q-parameter formalism. Gaussian optics only describes the evolution of the beam's width and phase curvature. Changes of amplitude and reflection are not included!

3.4.4 Gaussian modes in a resonator

In this chapter we will use our knowledge of paraxial Gaussian beam propagation to derive stability conditions for resonators. An optical cavity or optical resonator is an arrangement of mirrors that forms a standing wave cavity resonator for light waves. Optical cavities are a major component of lasers, surrounding the gain medium and providing feedback to the laser light (see the He-Ne laser experiment in the experimental optics course at the end of the 1st semester).

3.4.4.1 Transversal fundamental modes (rotational symmetry)



Wave fronts, i.e. planes of constant phase, of a Gaussian beam.

The general idea for obtaining a **stable** light configuration in a resonator is that the **mirrors and wave fronts** (planes of constant phase) **should coincide**. The radiation patterns are then reproduced on every round-trip of the light through the resonator. Those patterns are the so-called *modes* of the resonator.

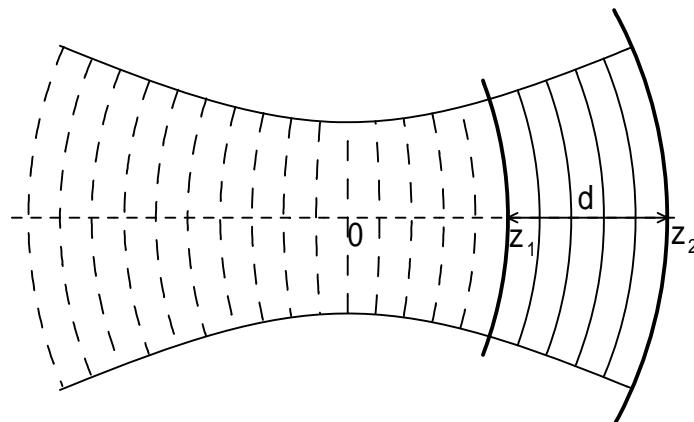
In the paraxial approximation and for Gaussian beams, this condition is easily fulfilled: the radii of mirror and wave front must be identical!

In this lecture we use the following conventions, which is different to that used in the script of the experimental optics course (see remark below):

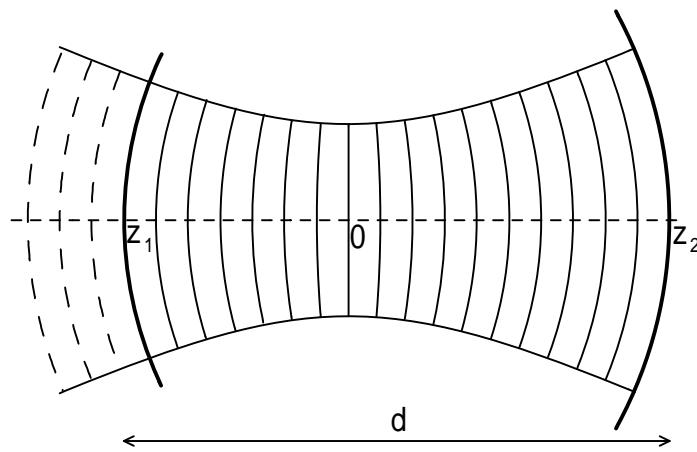
- $z_{1,2}$ are the positions of mirror '1' and '2'; $z = 0$ is the position of the focus!
- d is the distance between the two mirrors $\rightarrow z_2 - z_1 = d$
- since $R(z) = z + \frac{z_0^2}{z}$ \rightarrow radius of wave front < 0 for $z < 0$
- from Chapter 8: beam incident upon concave mirror \rightarrow radius $R_i(i=1,2) < 0$
beam incident upon convex mirror \rightarrow radius $R_i(i=1,2) > 0$

Examples:

A) $R(z_1), R(z_2) > 0; R_1 > 0, R_2 < 0; z_1 > 0, z_2 > 0$



B) $R(z_1) < 0, R(z_2) > 0; R_1, R_2 < 0; z_1 < 0, z_2 > 0$



The question which arises now is whether there exists a Gaussian beam, which fits into any given mirror configuration or if some restrictions apply on R_1 , R_2 , and d .

According to our reasoning above, the conditions for stability are:

$$R_1 = R(z_1), \quad R_2 = -R(z_2)$$

$$\Rightarrow R_1 = z_1 + \frac{z_0^2}{z_1}, \quad -R_2 = z_2 + \frac{z_0^2}{z_2}.$$

In both expressions, we solve for and eliminate the Rayleigh length z_0 :

$$z_1(R_1 - z_1) = -z_2(R_2 + z_2)$$

$$\text{with } z_2 = z_1 + d \text{ we find } z_1 = -\frac{d(R_2 + d)}{R_1 + R_2 + 2d}.$$

Now for given values of R_1, R_2, d , we can compute the location of the focus of the corresponding mode in the resonator. However, we have to make sure that such a mode actually exist. In our calculations above, we have eliminated the Rayleigh length z_0 , which must be a real and positive quantity. Hence, we must check that the so-called **stability condition** $z_0^2 > 0$ is fulfilled! From above,

$$\Rightarrow z_0^2 = R_1 z_1 - z_1^2 = -\frac{d(R_1 + d)(R_2 + d)(R_1 + R_2 + d)}{(R_1 + R_2 + 2d)^2} > 0$$

Since the denominator $(R_1 + R_2 + 2d)^2$ is always positive, we need to fulfill

$$\Rightarrow -d(R_1 + d)(R_2 + d)(R_1 + R_2 + d) > 0$$

If we introduce the so-called resonator parameters

$$g_1 = \left(1 + \frac{d}{R_1}\right), \quad g_2 = \left(1 + \frac{d}{R_2}\right),$$

we can re-express the stability condition as

$$\begin{aligned}
 -d(R_1 + d)(R_2 + d)(R_1 + R_2 + d) &= d g_1 g_2 R_1 R_2 \frac{(1 - g_1 g_2) R_1 R_2}{d} \\
 &= g_1 g_2 (1 - g_1 g_2) (R_1 R_2)^2 > 0.
 \end{aligned}$$

This inequality is fulfilled for

$$(1) \quad g_1 g_2 > 0 \text{ and } 1 - g_1 g_2 > 0$$

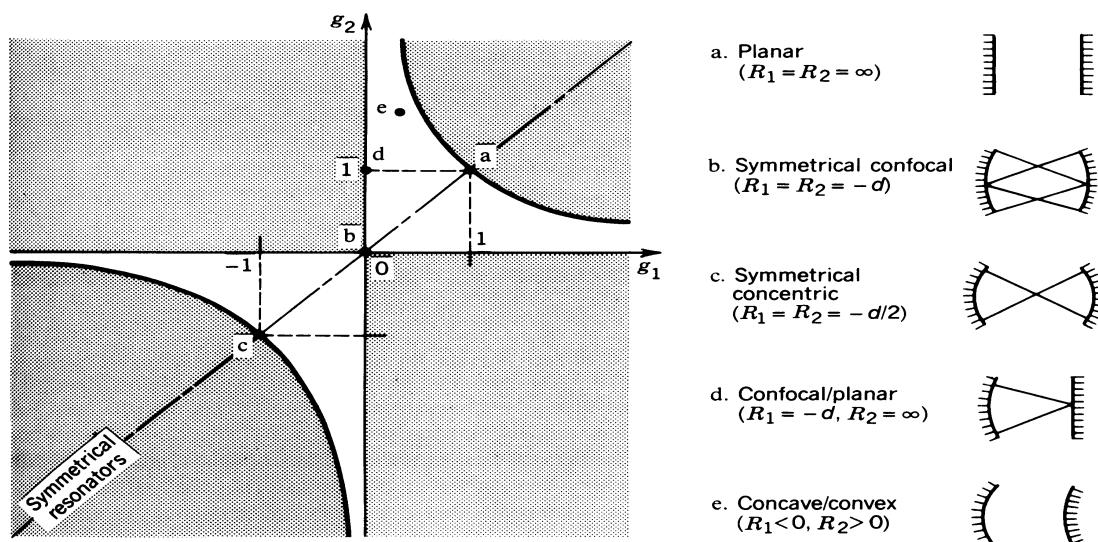
or for

$$(2) \quad g_1 g_2 < 0 \text{ and } 1 - g_1 g_2 < 0.$$

Since the second case cannot be fulfilled for any real-valued g_1 and g_2 , the following condition for the placement of the mirrors and their radii remains in determining a stable cavity

$$0 < g_1 g_2 < 1 \text{ which is equivalent to } 0 < \left(1 + \frac{d}{R_1}\right) \left(1 + \frac{d}{R_2}\right) < 1$$

This final form of the stability condition can be visualized: the region of stability of a resonator lies between the coordinate axes and hyperbolas:

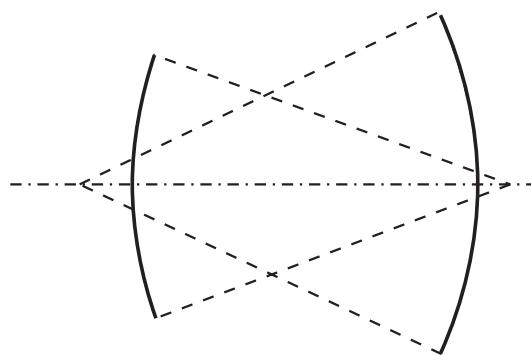


Resonator stability diagram. A spherical-mirror resonator is stable if the parameters $g_1 = 1 + d/R_1$ and $g_2 = 1 + d/R_2$ lie in the unshaded regions bounded by the lines $g_1 = 0$ and $g_2 = 0$, and the hyperbola $g_2 = 1/g_1$. R is negative for a concave mirror and positive for a convex mirror. Various special configurations are indicated by letters. All symmetrical resonators lie along the line $g_1 = g_2$.

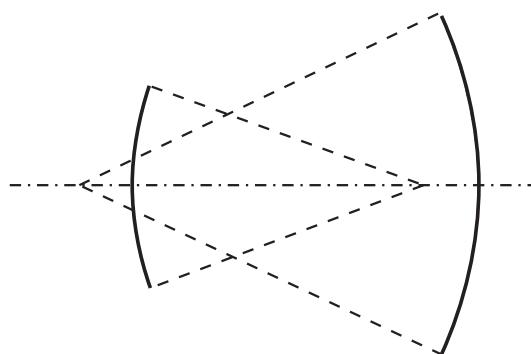
Resonator stability diagram. A spherical-mirror resonator is stable if the parameters $g_1 = 1 + d / R_1$ and $g_2 = 1 + d / R_2$ lie in the unshaded regions bounded by the lines $g_1 = 0$ and $g_2 = 0$, and the hyperbola $g_2 = 1 / g_1$. The convention adopted here is that R is negative for a concave mirror and positive for a convex mirror. Various special configurations are indicated by letters. All symmetrical resonators lie along the line $g_1 = g_2$.

Examples of a stable and an unstable resonator:

A) $R_1, R_2 < 0; |R_1| > d, |R_2| > d; \curvearrowleft 0 \leq g_1 \leq 1, 0 \leq g_2 \leq 1; \curvearrowright 0 \leq g_1 g_2 \leq 1 \curvearrowright \text{stable}$



B) $R_1, R_2 < 0; |R_1| < d, |R_2| > d; \curvearrowleft g_1 \leq 0, 0 \leq g_2 \leq 1; \curvearrowright g_1 g_2 \leq 0 \curvearrowright \text{unstable}$



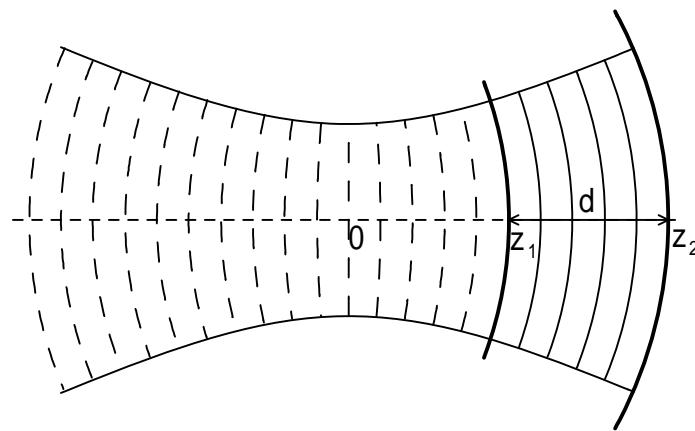
Remark: connection to the script for the HeNe experiment in the experimental optics course (and Wikipedia):

In the experiment's script (he_ne_laser.pdf), a slightly different convention is used:

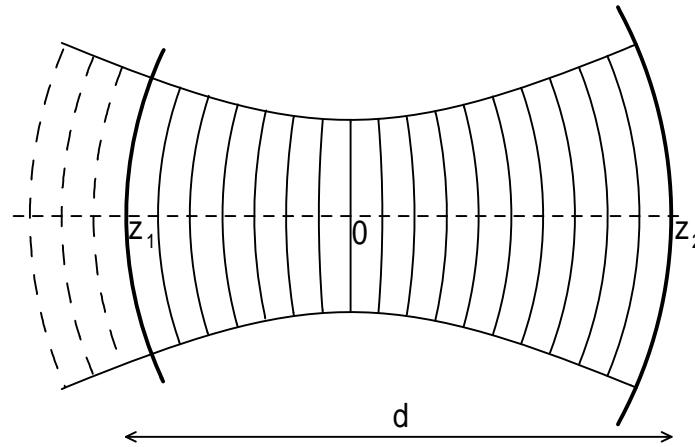
- Direction of z-axis reversed for the two mirrors
- beam incident upon concave mirror → radius $R_i(i=1,2) > 0$.
- beam incident upon convex mirror → radius $R_i(i=1,2) < 0$.
- $z_{1,2}$ is the distance of from mirrors '1' and '2' to the focus!
- d is the distance between the two mirrors → $z_2 + z_1 = d$

Examples:

A) $R(z_1) < 0, R(z_2) > 0 \quad R_1 < 0, R_2 > 0; z_1 < 0, z_2 > 0$



B) $R(z_1) > 0, R(z_2) > 0 ; R_1, R_2 > 0 ; z_1 > 0, z_2 > 0$



Then the conditions for stability are:

$$R_1 = R(z_1), \quad R_2 = R(z_2)$$

With an analogous calculation as above, we find that for the resonator parameters

$$g_1 = \left(1 - \frac{d}{R_1}\right), \quad g_2 = \left(1 - \frac{d}{R_2}\right)$$

the same stability condition

$$g_1 g_2 (1 - g_1 g_2) (R_1 R_2)^2 > 0, \quad 0 < g_1 g_2 < 1.$$

3.4.4.2 Higher order resonator modes

For the derivation of the above stability condition, we only needed the wave fronts. Hence, there may exist other modes with the same wavefronts but different intensity distributions. For the fundamental mode, we have:

$$v_g(x, y, z) = A \frac{w_0}{w(z)} \exp\left[-\frac{x^2 + y^2}{w^2(z)}\right] \exp\left[i \frac{k}{2} \frac{x^2 + y^2}{R(z)}\right] \exp[i\varphi(z)].$$

And for higher order modes this expression generalizes to

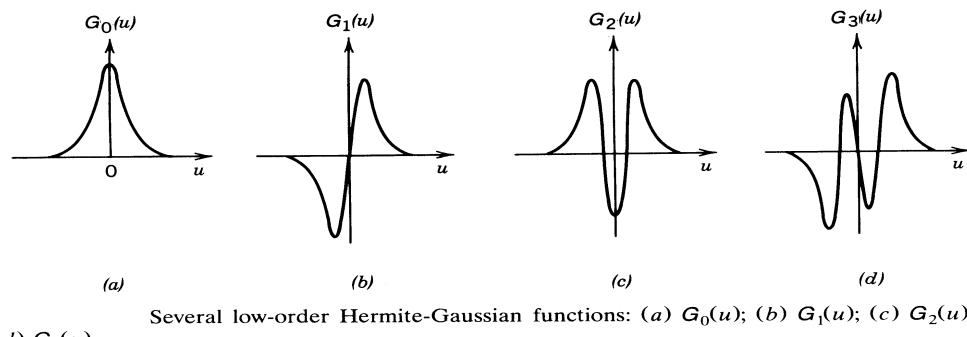
$$u_{l,m}(x,y,z) = A_{l,m} \frac{w_0}{w(z)} G_l \left[\sqrt{2} \frac{x}{w(z)} \right] G_m \left[\sqrt{2} \frac{y}{w(z)} \right] \times \\ \exp \left[i \frac{k}{2} \frac{x^2 + y^2}{R(z)} \right] \exp(i k z) \exp[i(l+m+1)\varphi(z)].$$

While the x, y -dependence of the phase is the same as for the fundamental mode, the amplitude's x, y -dependence is more complex:

$$G_l(\xi) = H_l(\xi) \exp\left(-\frac{\xi^2}{2}\right).$$

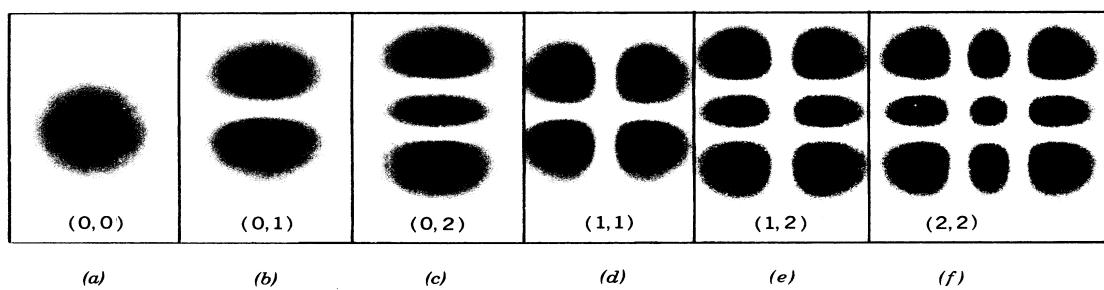
Here, the functions $G_l(\xi)$ are the Hermite-Gaussian functions and $H_l(\xi)$ are the so-called Hermite polynomials with

$$H_0 = 1, H_1 = 2\xi \text{ and } H_{l+1} = 2\xi H_l - 2l H_{l-1}.$$



Several low-order Hermite-Gaussian functions: (a) $G_0(u)$; (b) $G_1(u)$; (c) $G_2(u)$; (d) $G_3(u)$.

Several low-order Hermite-Gaussian functions: (a) $G_0(u)$; (b) $G_1(u)$; (c) $G_2(u)$; and (d) $G_3(u)$.



Intensity distributions of several low-order Hermite-Gaussian beams in the transverse plane. The order (l, m) is indicated in each case.

Intensity distributions of several low-order Hermite-Gaussian beams in the transverse plane. The order (l, m) is indicated in each case.

3.5 Dispersion of pulses in homogeneous isotropic media

3.5.1 Pulses with finite transverse width (pulsed beams)

In the previous chapters, we have treated the propagation of monochromatic beams where the frequency ω was fixed and therefore the wavenumber $k(\omega)$ was constant as well. This is the typical situation when we deal with continuous wave (cw) lasers.

However, for many applications (spectroscopy, nonlinear optics, telecommunication, material processing), we need to consider the propagation of pulsed laser light. In this situation, we have a typical envelope length T_0 of

$$10^{-13} \text{ s} (100 \text{ fs}) \leq T_0 \leq 10^{-10} \text{ s} (100 \text{ ps}).$$

Let us compute the spectrum of a Gaussian pulse:

$$f(t) = \exp(-\mathbf{i}\omega_0 t) \exp\left(-\frac{t^2}{T_0^2}\right)$$

$$F(\omega) \sim \exp\left[-\frac{(\omega - \omega_0)^2}{4/T_0^2}\right] \rightarrow \omega_s^2 = \frac{4}{T_0^2} \rightarrow \omega_s T_0 = 2$$

→ spectral width: $4 \cdot 10^{10} \text{ s}^{-1} \leq \omega_s \leq 4 \cdot 10^{13} \text{ s}^{-1}$

– center frequency of visible light: $\omega_0 = 2\pi\nu \sim 4 \cdot 10^{15} \text{ s}^{-1}$

→ optical cycle: $T_s = 2\pi / \omega_0 \approx 1.6 \cdot 10^{-15} \text{ s} = 1.6 \text{ fs}$

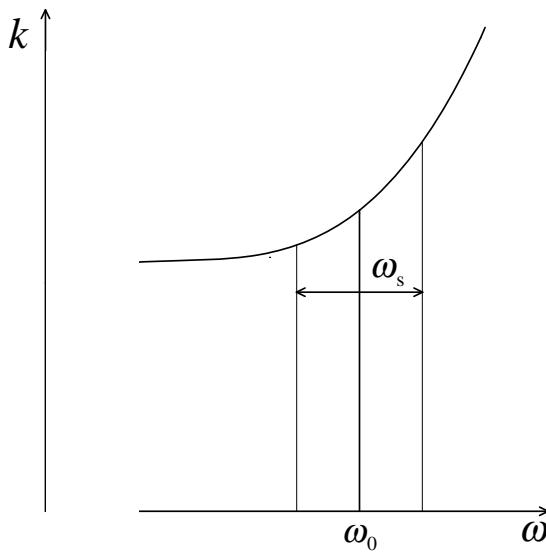
Hence, we have the following order of magnitudes:

$$\boxed{\omega_s \ll \omega_0 \rightarrow \omega - \omega_0 = \bar{\omega} \ll \omega_0}$$

i.e. the spectral bandwidth of the pulse is much smaller than the carrier frequency of visible light.

Thus, out of the very broad range of electromagnetic frequencies, we excite only a very small part. Accordingly, our models for describing the frequency-dependent response of the material through which the light propagates needs to cover only the narrow frequency range of the excited pulse.

In this situation, it can be helpful to replace the complicated frequency dependence (dispersion relation) of the wave vector $\mathbf{k}^2(\omega) = (\omega/c)^2 \epsilon(\omega)$ or the wave number $k(\omega) = (\omega/c)\sqrt{\epsilon(\omega)}$ by a Taylor expansion at the central frequency $\omega = \omega_0$, which can be used to approximate the complicated frequency dependence in a narrow frequency range quite well.



In most cases, a parabolic (or cubic) approximation of the frequency dependence in the dispersion relation will be sufficient:

$$k(\omega) \approx k(\omega_0) + \frac{\partial k}{\partial \omega} \Big|_{\omega_0} (\omega - \omega_0) + \frac{1}{2} \frac{\partial^2 k}{\partial \omega^2} \Big|_{\omega_0} (\omega - \omega_0)^2$$

The three expansion coefficients and their physical significance

The following terminology for the individual expansion coefficients is commonly used in the literature. It associates the three parameters of the Taylor expansion that we have retained with the physics of the dispersion relation.

A) Phase velocity v_{ph}

$$k(\omega_0) = k_0, \rightarrow \frac{1}{v_{\text{ph}}} = \frac{k_0}{\omega_0} = \frac{n(\omega_0)}{c}$$

The phase velocity is the velocity of the phase fronts for light at the central frequency $\omega = \omega_0$.

B) Group velocity v_g

$$\frac{\partial k}{\partial \omega} \Big|_{\omega_0} = \frac{1}{v_g}$$

The group velocity is the velocity of the center of the pulse with central frequency at $\omega = \omega_0$ (see detailed discussion below).

Substituting the dispersion relation

$$k(\omega) = \frac{\omega}{c} n(\omega)$$

into our definition of group velocity,

$$\frac{1}{v_g} = \left. \frac{\partial k}{\partial \omega} \right|_{\omega_0} = \frac{1}{c} \left[n(\omega_0) + \omega_0 \left. \frac{\partial n}{\partial \omega} \right|_{\omega_0} \right]$$

$$v_g = \frac{c}{\left[n(\omega_0) + \omega_0 \left. \frac{\partial n}{\partial \omega} \right|_{\omega_0} \right]} = \frac{c}{n_g(\omega_0)} = v_{\text{PH}} \frac{n(\omega_0)}{n_g(\omega_0)}$$

where $n_g(\omega_0)$ is the so-called group index

$$n_g(\omega_0) = n(\omega_0) + \omega_0 \left. \frac{\partial n}{\partial \omega} \right|_{\omega_0}.$$

For normal dispersion: $\partial n / \partial \omega > 0 \rightarrow n_g > n \rightarrow v_g < v_{\text{PH}}$

For anomalous dispersion: $\partial n / \partial \omega < 0 \rightarrow n_g < n \rightarrow v_g > v_{\text{PH}}$

C) Group velocity dispersion (GVD) or simply dispersion D_ω

$$\left. \frac{\partial^2 k}{\partial \omega^2} \right|_{\omega_0} = D_\omega$$

The GVD changes the pulse shape upon propagation (see detailed discussion below).

$$D = D_\omega = \left. \frac{\partial^2 k}{\partial \omega^2} \right|_{\omega_0} = \frac{\partial}{\partial \omega} \left(\frac{1}{v_g} \right)$$

$$D = \frac{\partial}{\partial \omega} \left(\frac{1}{v_g} \right) = -\frac{1}{v_g^2} \frac{\partial v_g}{\partial \omega}$$

$$\rightarrow D > 0 \curvearrowright \frac{\partial v_g}{\partial \omega} < 0$$

$$\rightarrow D < 0 \curvearrowright \frac{\partial v_g}{\partial \omega} > 0$$

Alternatively, in telecommunication one often uses the derivative with respect to the wavelength

$$D_\lambda = \frac{\partial}{\partial \lambda} \left(\frac{1}{v_g} \right) = -\frac{2\pi}{\lambda^2} c D_\omega$$

which can be related D_ω by the chain rule of differentiation.

The wave equation for pulsed beams

Let us now discuss the propagation of pulsed beams. We start with the **scalar Helmholtz equation**, with the full dispersion relation (no Taylor expansion yet):

$$\Delta \bar{u}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \epsilon(\omega) \bar{u}(\mathbf{r}, \omega) = 0$$

In contrast to monochromatic beam propagation, we now have for each frequency ω one Fourier component of the optical field:

$$\text{dispersion relation: } k^2(\omega) = \frac{\omega^2}{c^2} \epsilon(\omega)$$

Hence, we need to consider the propagation of the Fourier spectrum (Fourier transform in space and time) by a frequency-dependent transfer function:

$$U(\alpha, \beta, \omega; z) = U_0(\alpha, \beta, \omega) \exp[i\gamma(\alpha, \beta, \omega) z]$$

$$\text{with } \gamma(\alpha, \beta, \omega) = \sqrt{k^2(\omega) - \alpha^2 - \beta^2}.$$

The initial spectrum at $z=0$ is $U_0(\alpha, \beta, \omega)$ is obtained by a 3D Fourier transform with respect to space (in transverse directions) and time

$$U_0(\alpha, \beta, \omega) = \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} u_0(x, y, t) \exp[-i(\alpha x + \beta y - \omega t)] dx dy dt.$$

Let us further assume that the Fresnel (paraxial) approximation is justified (narrow spatial frequency spectrum with $k^2(\omega) \gg \alpha^2 + \beta^2$)

$$U(\alpha, \beta, \omega; z) \approx U_0(\alpha, \beta, \omega) \exp[ik(\omega)z] \exp\left[-i\frac{\alpha^2 + \beta^2}{2k(\omega)}z\right].$$

We see that propagation of pulsed beams in the Fresnel approximation in Fourier space is described by the following propagation function (transfer function):

$$H_F(\alpha, \beta, \omega; z) = \exp[ik(\omega)z] \exp\left[-i\frac{\alpha^2 + \beta^2}{2k(\omega)}z\right]$$

For a narrowly excited (temporal) spectrum (long pulses with respect to optical cycle), let us now consider the above Taylor expansion of $k(\omega)$ with respect to ω . If the pulse is not too short, we can replace the wave number $k(\omega)$ by

$$k(\omega) \approx k(\omega_0) + \frac{\partial k}{\partial \omega}\Big|_{\omega_0} (\omega - \omega_0) + \frac{1}{2} \frac{\partial^2 k}{\partial \omega^2}\Big|_{\omega_0} (\omega - \omega_0)^2.$$

Moreover, in the second term $\exp[-i(\alpha^2 + \beta^2)z / \{2k(\omega)\}]$ of the transfer function (which is already small due to paraxiality), we can neglect the frequency dependence of the wave number by $k(\omega) \approx k(\omega_0) = k_0$. This approximation assumes a non-dispersive diffraction term. This is sufficiently accurate to describe the diffraction of

pulsed beams which are not too short. For visible (VIS) or near infrared (NIR) light, this approximation can be applied to describe the diffraction of pulses with a pulse length $T_0 \geq 15\text{fs}$. For shorter pulses the frequency spectrum would become too wide and dispersion and diffraction of the pulses would not be independent of each other (space-time coupling by non-factorizable operators).

By introducing this approximation, we obtain the so-called parabolic approximation of the transfer function:

$$H_{\text{FP}}(\alpha, \beta, \omega; z) \approx \exp[\mathbf{i}k_0 z] \exp\left[\mathbf{i} \frac{1}{v_g} (\omega - \omega_0) z\right] \times \\ \times \exp\left[\mathbf{i} \frac{D}{2} (\omega - \omega_0)^2 z\right] \exp\left[-\mathbf{i} \frac{\alpha^2 + \beta^2}{2k_0} z\right]$$

The resulting propagation integral in the parabolic approximation for the spatiotemporally varying scalar field $u(x, y, z, t)$ is then

$$u(x, y, z, t) = \iiint_{-\infty}^{\infty} U_0(\alpha, \beta, \omega) H_{\text{FP}}(\alpha, \beta, \omega; z) \times \\ \times \exp[\mathbf{i}(\alpha x + \beta y - \omega t)] d\alpha d\beta d\omega.$$

In analogy to the slowly varying spatial envelope already discussed, we can introduce the slowly varying envelope in space and time as

$$u(x, y, z, t) = v(x, y, z, t) \exp[\mathbf{i}(k_0 z - \omega_0 t)].$$

To obtain a transfer function for the slowly varying amplitude, we can remove the factor resulting in rapid phase oscillations along z from the transfer function:

$$H_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) \approx \exp[\mathbf{i}k_0 z] \bar{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z),$$

where we have also introduced the difference $\bar{\omega}$ from the center frequency ω_0 as

$$\bar{\omega} = \omega - \omega_0.$$

The new transfer function $\bar{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z)$ is then

$$\bar{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) = \exp\left[\mathbf{i} z \left(\frac{\bar{\omega}}{v_g} + \frac{1}{2} D \bar{\omega}^2 - \frac{1}{2} \frac{1}{k_0} (\alpha^2 + \beta^2) \right)\right].$$

It appears in the propagation integral for the slowly varying envelope $v(x, y, z, t)$ as

$$v(x, y, z, t) = \iiint_{-\infty}^{\infty} V_0(\alpha, \beta, \bar{\omega}) \bar{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) \exp[\mathbf{i}(\alpha x + \beta y - \bar{\omega} t)] d\alpha d\beta d\bar{\omega}$$

where the initial spectrum is defined by the Fourier transform of the slowly varying spatiotemporal envelope $v(x, y, z, t)$ with respect to the frequency difference $\bar{\omega} = \omega - \omega_0$:

$$u_0(x, y, t) = v_0(x, y, t) \exp(-i\omega_0 t)$$

$$\rightarrow V_0(\alpha, \beta, \bar{\omega}) = \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} v_0(x, y, t) \exp[-i(\alpha x + \beta y - \bar{\omega}t)] dx dy dt$$

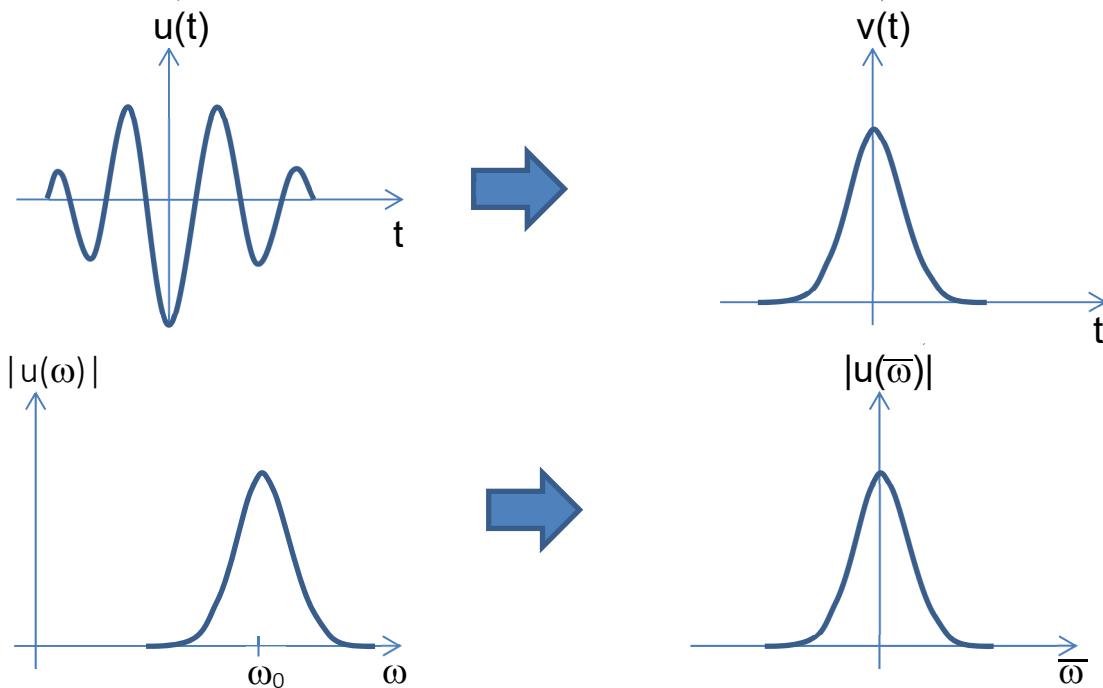
Note that this coincides with the Fourier transform of the full field $u(x, y, z, t)$ in the initial excitation plane $z = 0$, evaluated at ω . Hence, the propagation integral may be equivalently be expressed as:

$$v(x, y, z, t) = \iiint_{-\infty}^{\infty} U_0(\alpha, \beta, \bar{\omega} + \omega_0) \bar{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) \times \exp[i(\alpha x + \beta y - \bar{\omega}t)] d\alpha d\beta d\bar{\omega}$$

Remark: For the full rapidly varying field, the propagation integral is obtained by multiplying with the rapidly oscillating carrier wave

$$u(x, y, z, t) = \exp[i(k_0 z - \omega_0 t)] \iiint_{-\infty}^{\infty} U_0(\alpha, \beta, \omega) \bar{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) \times \exp[i(\alpha x + \beta y - \bar{\omega}t)] d\alpha d\beta d\bar{\omega}.$$

Illustration of the slowly varying envelope in the spectral domain



Co-moving reference frame

The next step is to introduce a co-moving reference frame by factoring out of the transfer function a term which simply describes the trivial propagation of the center of the pulse at the group velocity

$$\bar{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) = \exp\left[i \frac{\bar{\omega}}{v_g} z\right] \tilde{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z).$$

The new transfer function for the slowly varying envelope in the co-moving reference frame is then

$$\tilde{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) = \exp\left[i \frac{z}{2} \left(D\bar{\omega}^2 - \frac{\alpha^2 + \beta^2}{k_0}\right)\right].$$

Based on this transfer function, the evolution of the slowly varying envelope would be described as

$$v(x, y, z, t) = \iiint_{-\infty}^{\infty} V_0(\alpha, \beta, \bar{\omega}) \tilde{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z) \exp\left[i(\alpha x + \beta y - \bar{\omega}(t - z/v_g))\right] d\alpha d\beta d\bar{\omega}.$$

In this expression, the comoving reference frame of the pulse appears in the coupling of spatial and temporal coordinates in the term $t - z/v_g$. This can be conveniently expressed by introducing the so-called comoving reference time τ as

$$\tau = t - \frac{z}{v_g}.$$

This frame is called comoving since the frame moves along with the pulse such that it appears to be at rest and we can study any dispersion effects without the trivial motion of the entire pulse at the group velocity.

The linear ω -dependence in Fourier space had resulted in a shift in the time domain. Due to the shift theorem for Fourier transforms, multiplication with $\exp[i(\bar{\omega}/v_g)z]$ in the frequency domain corresponds to a shift of the coordinate in the time domain by $t - (z/v_g)$.

In contrast, $\tilde{H}_{\text{FP}}(\alpha, \beta, \bar{\omega}; z)$ is now purely quadratic in $\bar{\omega}$. Thus, the slowly varying envelope in the comoving frame evolves as:

$$\begin{aligned} \tilde{v}(x, y, z, \tau) &= \iiint_{-\infty}^{\infty} V_0(\alpha, \beta, \bar{\omega}) \exp\left[i \frac{z}{2} \left(D\bar{\omega}^2 - \frac{\alpha^2 + \beta^2}{k_0}\right)\right] \times \\ &\quad \times \exp\left\{i[\alpha x + \beta y - \bar{\omega} \tau]\right\} d\alpha d\beta d\bar{\omega}. \end{aligned}$$

The optical field u in the comoving frame becomes:

$$\tilde{u}(x, y, z, \tau) = \tilde{v}(x, y, z, \tau) \exp\left[\mathbf{i}(k_0 z - \omega_0 t)\right] = \tilde{v}(x, y, z, \tau) \exp\left[\mathbf{i}\left(k_0 z - \frac{\omega_0}{v_g} z - \omega_0 \tau\right)\right].$$

Propagation equation in real space

Finally, let us derive the propagation equation for the slowly varying envelope in the comoving frame. We start from the transfer function

$$\tilde{V}(\alpha, \beta, \bar{\omega}; z) = V_0(\alpha, \beta, \bar{\omega}) \exp\left[\mathbf{i} \frac{z}{2} \left(D\bar{\omega}^2 - \frac{\alpha^2 + \beta^2}{k_0}\right)\right].$$

We then take the spatial derivative of the transfer function along the propagation direction z

$$\mathbf{i} \frac{\partial \tilde{V}(\alpha, \beta, \bar{\omega}; z)}{\partial z} = -\frac{1}{2} \left(D\bar{\omega}^2 - \frac{\alpha^2 + \beta^2}{k_0}\right) \tilde{V}(\alpha, \beta, \bar{\omega}; z).$$

This is an ordinary differential equation describing the spatial evolution along z of the spatiotemporal frequency spectrum for the slowly varying envelope in the comoving reference frame.

As before in the case of monochromatic beams, we use the Fourier back-transformation to obtain the differential equation in the time-position domain

$$\mathbf{i} \frac{\partial \tilde{v}(x, y, z, \tau)}{\partial z} - \frac{D}{2} \frac{\partial^2}{\partial \tau^2} \tilde{v}(x, y, z, \tau) + \frac{1}{2k_0} \Delta^{(2)} \tilde{v}(x, y, z, \tau) = 0.$$

This is the scalar paraxial equation for propagation of so-called pulsed beams.

Comment: Extension to inhomogeneous media

By using the slowly varying envelope approximation, it is possible to generalize the scalar paraxial equation to **inhomogeneous** media, when a weak index contrast is assumed.

$$\mathbf{i} \frac{\partial}{\partial z} \tilde{v}(x, y, z, \tau) - \frac{D}{2} \frac{\partial^2}{\partial \tau^2} \tilde{v}(x, y, z, \tau) + \frac{1}{2\bar{k}_0} \Delta^{(2)} \tilde{v}(x, y, z, \tau) + \left[\frac{k_0^2(\mathbf{r}) - \bar{k}_0^2}{2\bar{k}_0} \right] \tilde{v}(x, y, z, \tau) = 0$$

where $\bar{k}_0 \approx \langle k_0(\mathbf{r}) \rangle$ is the averaged wave number

For $D = 0$, the equation would simply reduce to the case of diffraction which was derived earlier for beam propagation.

3.5.2 Pulses with infinite transverse extension (pulse propagation)

Diffraction plays a negligible role for sufficiently small propagation lengths. For broad beams, the spatial Rayleigh length $z_{0,\text{spatial}}$, over which diffraction becomes considerable, can be rather large compared to propagation distances of interest. We can then assume $\alpha = \beta \approx 0$, which corresponds to the assumption that we only have a

single plane wave propagating in z-direction. Later in this lecture series, we will see that this case is also valid for mode propagation in waveguides, e.g. optical telecommunication fibers.

Description in frequency domain

- 1) initial condition (expressed in terms of slowly varying temporal envelope):

$$u_0(t) = v_0(t) \exp(-\mathbf{i}\omega_0 t)$$

- 2) initial spectrum: $V_0(\bar{\omega}) = U_0(\omega)$

- 3) propagation of the spectrum by multiplication with transfer function in parabolic and non-diffractive approximation:

$$V(\bar{\omega}; z) = V_0(\bar{\omega}) \exp\left[\mathbf{i} z \frac{D}{2} \bar{\omega}^2\right]$$

- 4) back-transformation to τ leads to the following evolution of the slowly varying envelope in the comoving frame:

$$\tilde{v}(z, \tau) = \int_{-\infty}^{\infty} V_0(\bar{\omega}) \exp\left[\mathbf{i} z \frac{D}{2} \bar{\omega}^2\right] \exp[-\mathbf{i}\bar{\omega}\tau] d\bar{\omega}$$

Description in time domain

- A) In time domain, it is possible to describe pulse propagation by means of a response function:

$$\tilde{H}_p(\bar{\omega}; z) = \exp\left[\mathbf{i} z \frac{D}{2} \bar{\omega}^2\right] \rightarrow \text{FT}^{-1} \rightarrow \tilde{h}_p(\tau; z) = \sqrt{\frac{2}{-\mathbf{i}\pi Dz}} \exp\left[-\mathbf{i} \frac{\tau^2}{2Dz}\right]$$

and the evolution is then described by the convolution integral

$$\tilde{v}(z, \tau) = \int_{-\infty}^{\infty} \tilde{h}_p(\tau - \tau'; z) \tilde{v}_0(\tau') d\tau'.$$

- B) Neglecting diffraction in the evolution equation derived previously, the propagation of a slowly varying pulse envelope in the comoving frame is described by

$$\mathbf{i} \frac{\partial \tilde{v}(z, \tau)}{\partial z} - \frac{D}{2} \frac{\partial^2}{\partial \tau^2} \tilde{v}(z, \tau) = 0.$$

3.5.3 Analogy of diffraction and dispersion

BEAM DIFFRACTION

approximation: Fresnel/paraxial \leftrightarrow

restriction: monochromacy \leftrightarrow

$$\mathbf{i} \frac{\partial}{\partial z} v(x, y, z) + \frac{1}{2k} \Delta^{(2)} v(x, y, z) = 0 \quad \leftrightarrow$$

PULSE DISPERSION

Taylor expansion of $k(\omega)$ up to quadratic term (+ co-moving frame)

plane wave

$$\mathbf{i} \frac{\partial}{\partial z} \tilde{v}(z, \tau) - \frac{D}{2} \frac{\partial^2}{\partial \tau^2} \tilde{v}(z, \tau) = 0$$

$$\begin{aligned}
 (x, y) &\leftrightarrow \tau \\
 \nabla &\leftrightarrow \frac{\partial}{\partial \tau} \\
 \frac{1}{k_0} \text{ (always positive)} &\leftrightarrow -D \text{ (\(D \leq 0\) can vary sign)} \\
 (\alpha, \beta) &\leftrightarrow \bar{\omega} \\
 \tilde{H}_F(\alpha, \beta; z) = \exp\left(-i \frac{\alpha^2 + \beta^2}{2k} z\right) &\leftrightarrow \tilde{H}_P(\bar{\omega}; z) = \exp\left[i z \frac{D}{2} \bar{\omega}^2\right]
 \end{aligned}$$

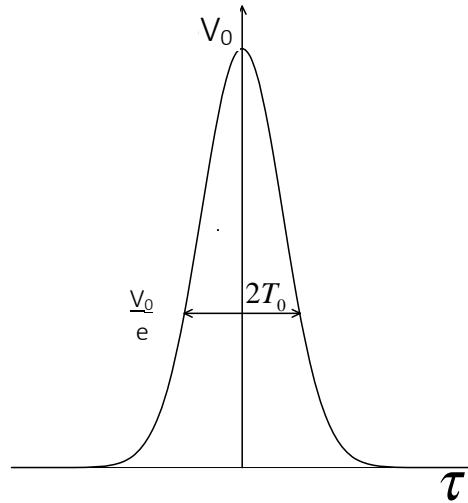
In the following we will study two typical examples of pulse propagation.

3.5.4 Propagation of a Gaussian pulse without chirp

use analogies to spatial diffraction

1) Initial pulse shape

pulse without chirp → corresponds to Gaussian pulse at the waist (focus) with flat phase



$$u_0(t) = A_0 \exp\left(-\frac{t^2}{T_0^2}\right) \exp(-i\omega_0 t) \rightarrow v_0(\tau) = A_0 \exp\left(-\frac{\tau^2}{T_0^2}\right)$$

2) Initial pulse spectrum (by Fourier transform)

$$V_0(\bar{\omega}) = A_0 \frac{T_0}{2\sqrt{\pi}} \exp\left(-\frac{\bar{\omega}^2 T_0^2}{4}\right)$$

$$\rightarrow \text{spectral width: } \omega_s^2 = 4/T_0^2$$

Use results from propagation of Gaussian beams:

Rayleigh range $z_0 = \left(\frac{k}{2} w_0^2\right)$ was used to describe Gaussian beam

Now the Gaussian pulse is described by an analogous dispersion parameter

$$z_0 = -\frac{1}{2} \frac{T_0^2}{D} \gtrless 0$$

Hence, anomalous GVD is equivalent to 'normal' diffraction.

Dispersion length: $L_D = 2|z_0|$

3) Propagated pulse spectrum (by product with transfer function)

$$V(\bar{\omega}; z) = A_0 \frac{T_0}{2\sqrt{\pi}} \exp\left(-\frac{\bar{\omega}^2 T_0^2}{4}\right) \exp\left(iz \frac{D}{2} \bar{\omega}^2\right)$$

4) Propagated pulse (by inverse Fourier transform)

$$\tilde{v}(z, \tau) = A_0 \sqrt{\frac{T_0}{T(z)}} \exp\left[-\frac{\tau^2}{T(z)^2}\right] \exp\left[-\frac{i}{2D} \frac{\tau^2}{R(z)}\right] \exp[i\varphi(z)]$$

with the following parameters

Amplitude

$$A(z) = A_0 \sqrt{\frac{1}{1 + \left(\frac{z}{z_0}\right)^2}}$$

Pulse length

$$T(z) = T_0 \sqrt{1 + \left(\frac{z}{z_0}\right)^2}$$

with $A^2(z)T(z) = \text{const.}$

Note that this must be the case due to energy conservation, since the amplitude squared gives the instantaneous power of the pulse, which when multiplied with the pulse duration gives the total energy of the pulse.

Phase profile

'Phase curvature' is not appropriate for describing pulses since it cannot be measured directly → introduction of new parameter: **chirp**

We can look at the following analogy:

Recall that the phase fronts $\Phi(x, y, z)$ of a Gaussian beam are given by:

$$\Phi(x, y, z) = \left\{ k \left[z + \frac{x^2 + y^2}{2R(z)} \right] + \varphi(z) \right\} = \text{const.}$$

These phase fronts can thus be described by differential equation:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \Phi(x, y, z) = \frac{k}{R(z)}$$

Likewise, we can investigate the temporal dynamics of the phase. For monochromatic fields, this is given by:

$$\Phi(\tau) = -\omega\tau \rightarrow -\frac{\partial\Phi(\tau)}{\partial\tau} = \omega$$

For an arbitrary time dependence of the phase, the same equation applies, but the instantaneous frequency may vary along the pulse

$$-\frac{\partial\Phi(\tau)}{\partial\tau} = \omega(\tau) \text{ and } -\frac{\partial^2\Phi(\tau)}{\partial\tau^2} = \frac{\partial\omega(\tau)}{\partial\tau} \neq 0 \rightarrow \text{chirp}$$

The chirp of a pulse describes the variation of the temporal frequency of the electric field in the pulse.

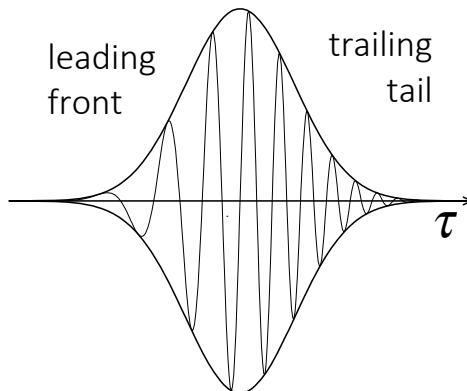
- parabolic approximation → quadratic temporal dependence of phase
- 'chirp' constant (second temporal derivative of phase)
- dimensionless chirp parameter (often just chirp)

$$C = -\frac{T_0^2}{2} \frac{\partial^2\Phi(\tau)}{\partial\tau^2}$$

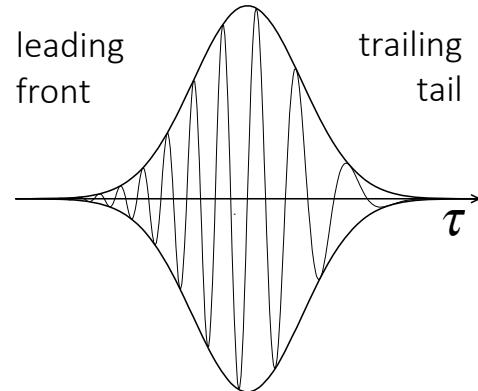
Integration over τ leads to:

$$-\frac{\partial\Phi(\tau)}{\partial\tau} = \omega(\tau) = \omega_0 + 2C \frac{\tau}{T_0^2}, \quad -\Phi(\tau) = \omega_0\tau + C \frac{\tau^2}{T_0^2}$$

$C > 0 \rightarrow$ up-chirp



$C < 0 \rightarrow$ down-chirp



Note that the trailing tail of the pulse occurs for larger values of τ .

Phase of Gaussian pulse :

in analogy to the above consideration,

$$\Phi(\tau) = -\omega_0 \left(\tau + \frac{z}{v_g} \right) - \frac{\tau^2}{2DR(z)} \equiv -\omega_0 \left(\tau + \frac{z}{v_g} \right) - C(z) \frac{\tau^2}{T_0^2}$$

where the first term is derived from the carrier wave. We have now allowed the chirp $C(z)$ to vary with z :

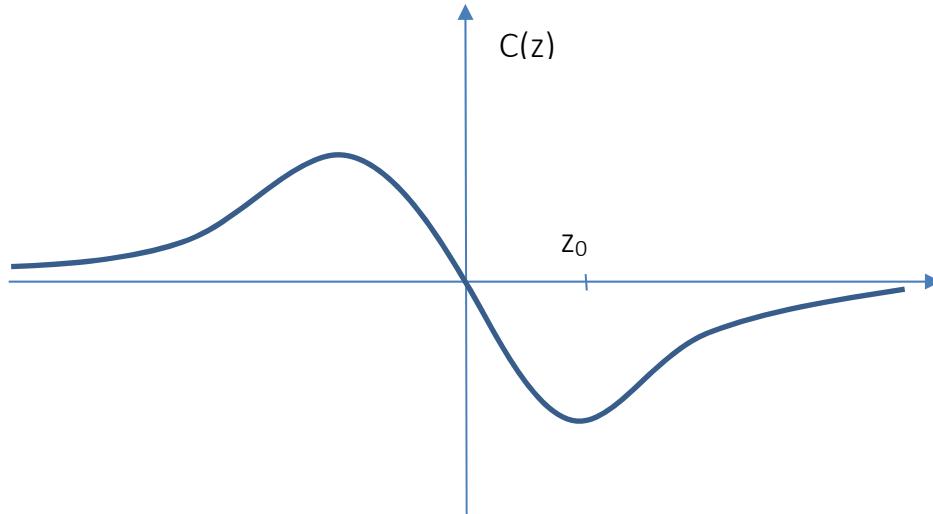
$$\rightarrow C(z) = \frac{T_0^2}{2DR(z)} = -\frac{z_0}{R(z)}.$$

By analogy with the case of beam propagation, we see that the chirp of the pulse must evolve as

$$R(z) = \frac{z^2 + z_0^2}{z} \rightarrow C(z) = -\frac{z_0 z}{z^2 + z_0^2} = -\frac{z}{z_0 \left(1 + \frac{z^2}{z_0^2}\right)}$$

$$\rightarrow C(0) = 0, \quad C(|z_0|) = -\frac{1}{2} \operatorname{sgn}(z_0), \quad C(z \rightarrow \infty) = -\frac{z_0}{z} \text{ with } z_0 = -\frac{T_0^2}{2D}$$

Attention: Chirp depends on sign of z_0 and hence on D .

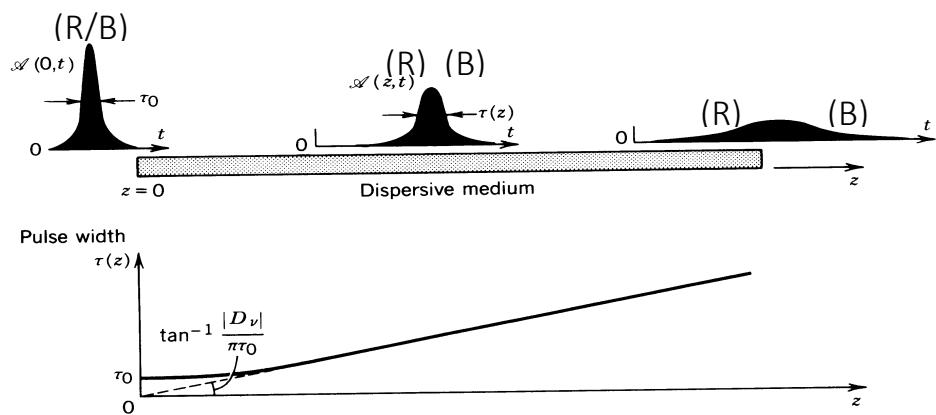


Evolution of the chirp parameter $C(z)$

Complete field:

$$u(z, \tau) = A_0 \sqrt{\frac{T_0}{T(z)}} \exp\left[-\frac{\tau^2}{T(z)^2}\right] \exp\left[-iC(z) \frac{\tau^2}{T_0^2}\right] \exp[i\varphi(z)] \exp[i(k_0 z - \omega_0 t)]$$

Dynamics of a pulse is equivalent to that of a beam.



Gaussian pulse spreading as a function of distance. For large distances, the width increases at the rate $|D|/\pi\tau_0$, which is inversely proportional to the initial width τ_0 .

Gaussian pulse spreading as a function of distance z . For large distances, the width increases linearly at a rate $2|D|/T_0$, which is inversely proportional to the initial width T_0 (for $D > 0$).

important parameter: dispersion parameter $z_0 = -\frac{T_0^2}{2D}$

Different regimes of pulse propagation:

- 1) $z \ll |z_0|$: no effect
- 2) $z \approx |z_0|$: complicated dispersion behavior which needs to be described by the above formula
- 3) $z \gg |z_0|$: asymptotic linear dependence

$$T(z) \approx T_0 \frac{z}{|z_0|} \rightarrow T(z) \approx \frac{2|D|}{T_0} z$$

$$T(z)/z = T_0/|z_0| = \frac{2|D|}{T_0} = \text{const.}$$

Remark: $D \leq 0$ is only important if initial pulse is chirped, since otherwise the same quadratic dependence is observed, independently of the sign of D .

3.5.5 Propagation of a chirped Gaussian pulse

Important because of:

- short pulse lasers → chirped pulses
- chirp is introduced on purpose, for subsequent pulse compression
- analogy to curved phase → focusing
- chirped pulse amplification (CPA) → Petawatt lasers (Nobel Prize 2018)

1) Initial pulse shape

$$v_0(\tau) = A_0 \exp\left[-\frac{\tau^2(1+iC_0)}{T_0^2}\right] \quad C_0 - \text{initial chirp}$$

2) Initial pulse spectrum

$$V_0(\bar{\omega}) = A_0 \exp\left[-\frac{\bar{\omega}^2 T_0^2 (1-iC_0)}{4(1+C_0^2)}\right]$$

$$\rightarrow \text{spectral width: } \bar{\omega}_s^2 = \frac{4(1+C_0^2)}{T_0^2}$$

- Spectral width of chirped pulse is larger than that of unchirped pulse of the same length.

$$(\bar{\omega}_s^2 = 4/T_0^2) \text{ only for transform limited pulses}$$

3) Propagation of pulse

Aim: calculation of the dependence of pulse width and chirp on z for given initial conditions.

Instead of using the product of the initial spectrum with the transfer function and the subsequent inverse Fourier transform to calculate the pulse propagation, we will use the q-parameter formalism, which we used for Gaussian beams before.

Gaussian beam → q -parameter → similar to Gaussian pulse

Analogously: $q(z) = q(0) + z$.

Recall that for beams beams:

$$\frac{1}{q(z)} = \frac{1}{R(z)} + i \frac{2}{kw^2(z)}.$$

$$k \rightarrow -\frac{1}{D}, w^2(z) \rightarrow T^2(z), \frac{1}{R(z)} \rightarrow \frac{2DC(z)}{T_0^2}$$

$$\rightarrow \frac{1}{q(z)} = \frac{2DC(z)}{T_0^2} - i \frac{2D}{T^2(z)}$$

$$\boxed{\frac{1}{q(z)} = \frac{2D}{T_0^2} \left[C(z) - i \frac{T_0^2}{T^2(z)} \right]} \quad (*)$$

Important: T_0 here is defined as the pulse width at $z=0$, which does not necessarily coincide with the 'focus' or waist.

Initial q-parameter at $z=0$:

$$\rightarrow \frac{1}{q(0)} = \frac{2D}{T_0^2} [C_0 - i] \text{ with } C_0 = C(0)$$

Propagation through homogeneous space over distance z

a) $q(z) = q(0) + z$ with $\frac{1}{q(0)} = \frac{2D}{T_0^2} [C_0 - i]$

b) invert to calculate $\frac{1}{q(z)}$

c) determine $T(z)$, $C(z)$ from $\frac{1}{q(z)} = \frac{2D}{T_0^2} \left[C(z) - i \frac{T_0^2}{T^2(z)} \right]$.

Generally: 2 equations (real and imaginary part) with the 5 parameters $C_0, T_0, z, C(z), T(z) \rightarrow 3$ values must be known

Let us consider the individual steps of the calculation for a propagation distance $z = d$

1) Determination of q parameter at input

$$q(0) = \frac{T_0^2}{2D} \frac{(C_0 + i)}{(1 + C_0^2)}$$

2) Evolution of q parameter

$$q(d) = q(0) + d = \frac{T_0^2}{2D} \frac{(C_0 + i)}{(1 + C_0^2)} + d = \frac{2Dd(1 + C_0^2) + C_0 T_0^2 + i T_0^2}{2D(1 + C_0^2)} \quad (**)$$

3) Inversion of general equation (*) for $q(d)$

$$\frac{1}{q(d)} = \frac{2D}{T_0^2} \left[C(d) - i \frac{T_0^2}{T^2(d)} \right]$$

$$q(d) = \frac{T_0^2 T^2(d) [C(d) T^2(d) + i T_0^2]}{2D [C^2(d) T^4(d) + T_0^4]}$$

4) Set equations (*) & (**) equal

$$\frac{2Dd(1 + C_0^2) + C_0 T_0^2 + i T_0^2}{2D(1 + C_0^2)} = \frac{T_0^2 T^2(d) [C(d) T^2(d) + i T_0^2]}{2D [C^2(d) T^4(d) + T_0^4]}$$

split into real and imaginary part:

a) real part $\frac{2Dd(1 + C_0^2) + C_0 T_0^2}{(1 + C_0^2)} = \frac{C(d) T_0^2 T^4(d)}{[C^2(d) T^4(d) + T_0^4]} \quad (****)$

b) imaginary part $\frac{1}{(1 + C_0^2)} = \frac{T_0^2 T^2(d)}{[C^2(d) T^4(d) + T_0^4]} \quad (*****)$

If 3 parameters are known (e.g. $C_0, T_0, C(d)$), we can determine the other 2 unknown parameters ($d, T(d)$).

Important question: Where is the pulse compressed to its smallest length?

given: C_0, T_0 & in the focus: $C(d)=0$

unknown: $d, T(d)$

a) From $C(d) = 0 \rightarrow$ real part (****) must be zero

$$\rightarrow [2Dd(1+C_0^2) + C_0T_0^2] = 0$$

$$\rightarrow d = -\frac{1}{2} \frac{T_0^2 C_0}{D(1+C_0^2)} = -\frac{1}{2} \text{sgn}(D) \frac{C_0}{(1+C_0^2)} L_D$$

b) From $C(d)=0 \rightarrow$ imaginary part (*****) becomes

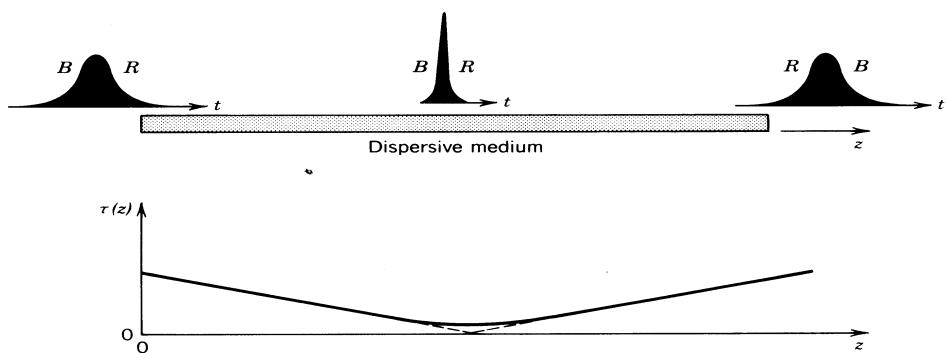
$$\rightarrow T^2(d) = \frac{T_0^2}{(1+C_0^2)}$$

Resulting properties

- 1) A pulse can be compressed when the product of initial chirp and dispersion is negative $\rightarrow C_0 D < 0$. This follows from (a), since only then is the distance d positive.
- 2) The possible compression increases with initial chirp C_0 .

Physical interpretation

If e.g. $C_0 < 0$ and $D > 0 \rightarrow \partial v_g / \partial \omega < 0 \rightarrow$ 'red' is faster than 'blue'



Compression of a chirped pulse in a medium with normal dispersion. The low frequency (marked R) occurs after the high frequency (marked B) in the initial pulse, but it catches up since it travels faster. Upon further propagation, the pulse spreads again as the R component arrives earlier than the B component.

Compression of a chirped pulse in a medium with normal dispersion. The low frequency (marked R for red) occurs after the high frequency (marked B for blue) in the initial pulse (down-chirp), but it catches up since it travels faster. Upon further propagation, the pulse spreads again as the R component arrives earlier than the B component (up-chirp).

- 1) Initially, the 'red tail' of the pulse catches up with the 'blue front' until the chirp $C(z)=0$ at the waist, i.e. the pulse is compressed. At this propagation distance the pulse has no remaining chirp.
- 2) Subsequently, $C(z)>0$ and red is in front. The 'red front' is faster than the 'blue tail', i.e. the pulse gets wider.

$$C(z) = -\frac{z}{z_0 \left(1 + \frac{z^2}{z_0^2}\right)} \quad z_0 = -\frac{T_0^2}{2D}$$

4. Diffraction theory

4.1 Interaction with plane masks

In this chapter, we will use our knowledge of beam propagation to analyze diffraction effects. In particular, we will treat the interaction of light with thin planar masks /apertures. We would like to understand how a given transversally localized field distribution propagates in a half-space.

There are different approximations commonly used to describe light propagation behind an amplitude mask:

- A) If we use geometrical optics, we obtain a simple shadow.
- B) We can use scalar diffraction theory with the light-aperture interaction approximated by a **complex transmission function**

$$t(x, y) \text{ with } t(x, y) = 0 \text{ for } |x|, |y| > a \text{ (aperture)}$$

Here we consider the description based on scalar diffraction theory. We can then split our diffraction problem into three sequential processes:

- i) propagation from light source to aperture
→ not important, generally plane wave (no diffraction)
- ii) multiply field distribution of illuminating wave by transmission function

$$u_+(x, y, z_A) = t(x, y) u_-(x, y, z_A)$$

- iii) propagation of the modified field distribution behind the aperture through homogeneous space

$$u(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(\alpha, \beta; z - z_A) U_+(\alpha, \beta; z_A) \exp[i(\alpha x + \beta y)] d\alpha d\beta$$

or

$$u(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x - x', y - y', z - z_A) u_+(x', y', z_A) dx' dy'$$

$$\text{with } h = \frac{1}{(2\pi)^2} \text{FT}^{-1}[H]$$

In the following we will use the notation $z_B = z - z_A$. According to our choice of the propagation function H , resp. h , we can compute this propagation either exactly or in the paraxial approximation (Fresnel). In the following, we will see that a further approximation is possible for very large z_B , the so-called Fraunhofer approximation.

4.2 Propagation using different approximations

4.2.1 General case - small aperture

We know from before that for arbitrary fields (with arbitrarily wide angular spectra), we must use the general propagation function

$$H(\alpha, \beta; z_B) = \exp(i\gamma(\alpha, \beta)z_B) \text{ with } \gamma^2 = k^2(\omega) - \alpha^2 - \beta^2.$$

We then have no constraints with respect to the spatial frequencies α, β . We get **homogeneous and evanescent** waves and can treat arbitrary small structures in the aperture by:

$$u(x, y, z) = \iint_{-\infty}^{\infty} U_+(\alpha, \beta) H(\alpha, \beta; z_B) \exp[i(\alpha x + \beta y)] d\alpha d\beta$$

where $U_+(\alpha, \beta) = \text{FT}[u_+(x, y)]$

Derivation of the response function of homogeneous space

We start from the Weyl expansion of an outgoing spherical wave as a superposition of plane waves:

$$\frac{1}{r} \exp(ikr) = \frac{i}{2\pi} \iint_{-\infty}^{\infty} \frac{1}{\gamma} \exp[i(\alpha x + \beta y + \gamma z)] d\alpha d\beta$$

Now we can compute the general response function h , which we did not do in the previous chapter as we only computed h_F (in the Fresnel approximation). By taking the derivative of the above equation, we get

$$\frac{\partial}{\partial z} \left[\frac{1}{r} \exp(ikr) \right] = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[i(\alpha x + \beta y + \gamma z)] d\alpha d\beta = -\frac{1}{2\pi} \text{FT}^{-1}[H] = -2\pi h$$

and therefore

$$h(x, y, z) = -\frac{1}{2\pi} \frac{\partial}{\partial z} \left[\frac{1}{r} \exp(ikr) \right] \text{ with } r = \sqrt{x^2 + y^2 + z^2}.$$

Remark: The derivative operator in the response function signifies that by merely defining the initial field, the problem of wave propagation without the Fresnel approximation is not well-defined. For such a problem, which is described by an elliptic partial differential equation, one needs to define the field at the entire boundary of the half-space or alternatively its derivative at the excitation plane $z = 0$. From the point of view of physics, this corresponds to defining the direction of field propagation and not just the value of the field in a single plane.

The resulting expression in position space for the propagation of monochromatic beams is also called 'Rayleigh-formula':

$$u(x, y, z_A + z_B) = \iint_{-\infty}^{\infty} h(x - x', y - y', z_B) u_+(x', y', z_A) dx' dy'$$

4.2.2 Fresnel approximation (paraxial approximation)

From the previous chapter, we know that we can apply the Fresnel approximation if $\alpha^2 + \beta^2 \ll k^2$. Thus, this approximation is valid only for a limited angular spectrum, which corresponds to a large size of the structures inside the aperture. Then

$$H_F(\alpha, \beta; z_B) = \exp(i k z_B) \exp\left[-i \frac{z_B}{2k} (\alpha^2 + \beta^2)\right]$$

$$h_F(x, y, z_B) = -\frac{i}{\lambda z_B} \exp(i k z_B) \exp\left[i \frac{k}{2z_B} (x^2 + y^2)\right]$$

4.2.3 Paraxial Fraunhofer approximation (far field approximation)

A further simplification of the beam propagation is possible for many diffraction problems. Let us assume a narrow angular spectrum

$$\alpha^2 + \beta^2 \ll k^2$$

and the additional condition for the so-called Fresnel number N_F

$$N_F \lesssim 0.1 \text{ with } N_F = \frac{a}{\lambda} \frac{a}{z_B}$$

where a is the (largest) size of the aperture (likthe "beam width"). Obviously, a larger aperture requires a larger distance z_B to fulfill the condition $N_F \lesssim 0.1$. Hence, the approximation which we derive in the following, is only valid in the so-called '**far field**', which means far away from the aperture.

To understand the implication of this new condition on the Fresnel number, we consider beam propagation in the paraxial approximation:

$$u_F(x, y, z_B) = \iint_{-\infty}^{\infty} h_F(x - x', y - y'; z_B) u_+(x', y') dx' dy'$$

$$= -\frac{i}{\lambda z_B} \exp(i k z_B) \iint_{-\infty}^{\infty} u_+(x', y') \exp\left\{i \frac{k}{2z_B} [(x - x')^2 + (y - y')^2]\right\} dx' dy'$$

Instead of working in the Fourier domain as we have previously done, in this situation it is easier to treat the beam propagation in position space by solving the above convolution integral, because

$$u_+(x, y) = t(x, y) u_-(x, y), \text{ and } t(x, y) = 0 \text{ for } |x|, |y| > a \text{ (aperture)}$$

$$\rightarrow u_+(x, y) = 0 \text{ for } |x|, |y| > a$$

This means that in the above convolution integral, we only need to integrate over the aperture and not over an infinite plane:

$$u_F(x, y, z_B) = -\frac{i}{\lambda z_B} \exp(i k z_B) \iint_{-\infty}^{\infty} u_+(x', y') \exp\left\{i \frac{k}{2z_B} [(x - x')^2 + (y - y')^2]\right\} dx' dy'$$

Now let us take a closer look at the exponential function in the integral:

$$\begin{aligned}
 u_F(x, y, z_B) &= -\frac{i}{\lambda z_B} \exp(ikz_B) \\
 &\quad \iint_{-\infty}^{\infty} u_+(x', y') \exp \left\{ i \frac{k}{2z_B} [x^2 - 2xx' + x'^2 + y^2 - 2yy' + y'^2] \right\} dx' dy' \\
 &= -\frac{i}{\lambda z_B} \exp(ikz_B) \\
 &\quad \iint_{-\infty}^{\infty} u_+(x', y') \exp \left\{ i \frac{k}{2z_B} [x^2 + y^2] \right\} \exp \left\{ -i \left[\frac{kx}{z_B} x' + \frac{ky}{z_B} y' \right] \right\} \exp \left\{ i \frac{k}{2z_B} [x'^2 + y'^2] \right\} dx' dy'.
 \end{aligned}$$

So far, we have only rearranged the factors. The trick is that because of the limits of integration, we have $x', y' < a$ and therefore

$$\rightarrow \frac{k}{2z_B} [x'^2 + y'^2] < \frac{ka^2}{z_B} = 2\pi N_F$$

Hence,

$$\text{for } N_F \lesssim 0.1 \rightarrow \exp \left\{ i \frac{k}{2z_B} [x'^2 + y'^2] \right\} \approx 1$$

This means that we can neglect the phase term quadratic in x', y' and obtain for the field far from the object:

$$\begin{aligned}
 u_{FR}(x, y, z_B) &= -\frac{i}{\lambda z_B} \exp(i k z_B) \exp \left[i \frac{k}{2z_B} (x^2 + y^2) \right] \\
 &\quad \times \iint_{-\infty}^{\infty} u_+(x', y') \exp \left\{ -i \left(\frac{kx}{z_B} x' + \frac{ky}{z_B} y' \right) \right\} dx' dy' \\
 &= -i \frac{(2\pi)^2}{\lambda z_B} \exp(i k z_B) U_+ \left(k \frac{x}{z_B}, k \frac{y}{z_B}; z_A \right) \exp \left[i \frac{k}{2z_B} (x^2 + y^2) \right]
 \end{aligned}$$

This is the so-called far-field in the paraxial Fraunhofer approximation or sometimes just the far-field approximation or Fraunhofer approximation. We see that the amplitude in the far-field is given by the Fourier transform of field behind the aperture. Likewise, the intensity distribution in the far field (position space) is given by the Fourier transform of the field distribution at the aperture

$$I_{FR}(x, y, z_B) \sim \frac{1}{(\lambda z_B)^2} \left| U_+ \left(k \frac{x}{z_B}, k \frac{y}{z_B}; z_A \right) \right|^2$$

Note that the intensity decreases with the square of propagation distance z_B . This is simply the inverse square law for intensities, which results from energy conservation.

Interpretation

For any plane $z = z_B$ in the far field, only one spatial frequency ($\alpha = kx/z_B; \beta = ky/z_B$) with spectral amplitude $U_+(kx/z_B, ky/z_B)$ contributes to the field distribution at each point x, y . This is in contrast to the previously considered cases, where all spatial frequencies contributed to the field at a given point.

In summary, we have shown that in the (paraxial) Fraunhofer approximation, the propagated field, or the diffraction pattern, is very straightforward to calculate. We simply need to compute the Fourier transform of the field at the aperture. In order to apply this approximation we have to check that:

- A) $\alpha^2 + \beta^2 \ll k^2 \rightarrow$ smallest features $\Delta x, \Delta y \gg \lambda$
 \rightarrow narrow angular spectrum (paraxiality)
- B) $N_F = \frac{a^2}{\lambda z_B} \ll 1 \rightarrow$ largest feature a determines $z_B \gg \frac{a^2}{\lambda}$
 \rightarrow minimum propagation distance to the far field

Example: $\Delta x, \Delta y = 10\lambda, a = 100\lambda, \lambda = 1\mu\text{m} \rightarrow z_B \gg 10^4\lambda \approx 1\text{ cm}$

4.2.4 Non-paraxial Fraunhofer approximation

The concept that the angular components of the input spectrum separate in the far field due to diffraction also applies beyond the paraxial approximation.

If we have arbitrary spatial frequencies in our spectrum, only those satisfying $\alpha^2 + \beta^2 \leq k^2$ contribute to the far field distribution.

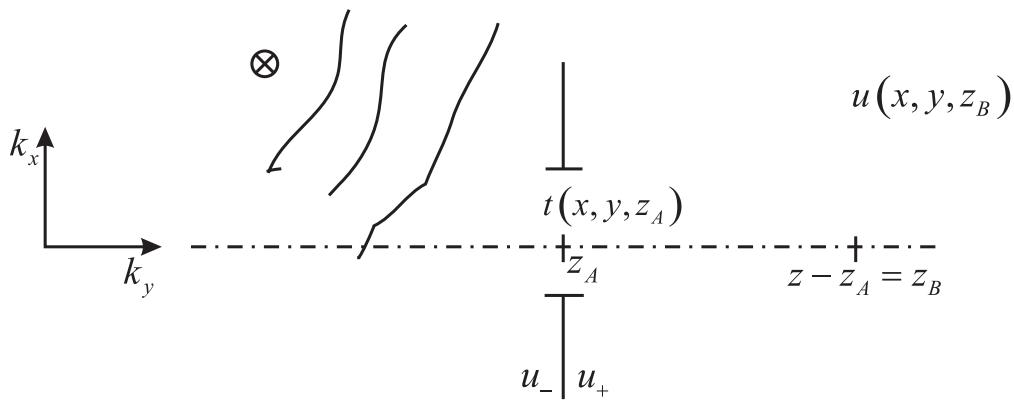
Evanescence waves decay for $kz_B \gg 1 \rightarrow z_B \gg \lambda$.

$$N_F \lesssim 0.1 \quad \text{with } N_F = \frac{a}{\lambda} \frac{a}{z_B} \quad \left(= \frac{1}{\pi} \frac{z_0}{z_B} \right)_{\text{for Gaussian beams}}$$

$$u_{\text{FR non-paraxial}}(x, y, z_B) = -i \frac{(2\pi)^2}{\lambda \sqrt{x^2 + y^2 + z_B^2}} \frac{z_B}{\sqrt{x^2 + y^2 + z_B^2}} \\ \times U_+ \left(\frac{kx}{\sqrt{x^2 + y^2 + z_B^2}}, \frac{ky}{\sqrt{x^2 + y^2 + z_B^2}}; z_A \right) \exp \left(ik \sqrt{x^2 + y^2 + z_B^2} \right)$$

4.3 Paraxial Fraunhofer diffraction at plane masks

Let us now combine our results and investigate examples of diffraction patterns induced by plane masks within the (paraxial) Fraunhofer approximation.



Let us consider an incident plane wave with a wavevector which is inclined with respect to the optical axis z

$$u_-(x, y, z_A) = A \exp\left[\mathbf{i}(k_x x + k_y y + k_z z_A)\right]$$

The field behind the mask is given by:

$$u_+(x, y, z_A) = u_-(x, y, z_A) t(x, y) = A \exp\left[\mathbf{i}(k_x x + k_y y + k_z z_A)\right] t(x, y) (*)$$

From the previous chapter, we know that the diffraction pattern in the far field in the paraxial Fraunhofer approximation is given by:

$$I(x, y, z_B) \sim |u(x, y, z_B)|^2 \sim \frac{1}{(\lambda z_B)^2} \left| U_+\left(k \frac{x}{z_B}, k \frac{y}{z_B}\right) \right|^2$$

Hence, the diffraction pattern is proportional to the spectrum of the field behind the mask at

$$\alpha = k \frac{x}{z_B}, \beta = k \frac{y}{z_B}$$

This spectrum is calculated by the Fourier transform of the field (*) as:

$$\begin{aligned} & U_+\left(k \frac{x}{z_B}, k \frac{y}{z_B}\right) \\ &= \frac{A}{(2\pi)^2} \exp(\mathbf{i} k_z z_A) \int_{-\infty}^{\infty} t(x', y') \exp\left[-\mathbf{i}\left(k \frac{x}{z_B} - k_x\right)x' - \mathbf{i}\left(k \frac{y}{z_B} - k_y\right)y'\right] dx' dy' \\ &= A \exp(\mathbf{i} k_z z_A) T\left(k \frac{x}{z_B} - k_x, k \frac{y}{z_B} - k_y\right) \end{aligned}$$

Hence, the intensity distribution of the diffraction pattern is given by:

$$I(x, y, z_B) \sim \frac{1}{(\lambda z_B)^2} \left| T\left(k \frac{x}{z_B} - k_x, k \frac{y}{z_B} - k_y\right) \right|^2$$

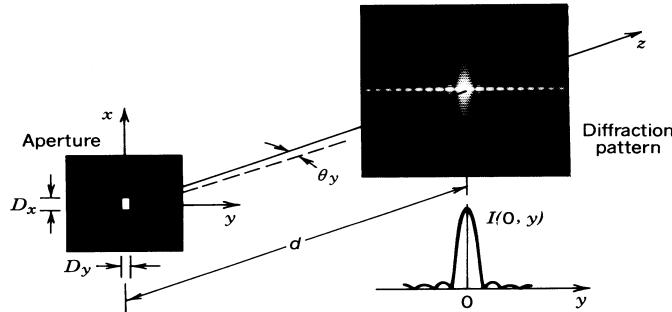
This is related to the modulus squared of the Fourier transform of the aperture function. In the paraxial approximation, an inclination of the illuminating plane wave simply shifts the pattern transversely, as would be expected.

Examples

A) Rectangular aperture illuminated by normal plane wave

$$t(x, y) = \begin{cases} 1 & \text{for } |x| \leq a, |y| \leq b \\ 0 & \text{elsewhere} \end{cases}$$

$$I(x, y, z_B) \sim \operatorname{sinc}^2\left(ka \frac{x}{z_B}\right) \operatorname{sinc}^2\left(kb \frac{y}{z_B}\right)$$



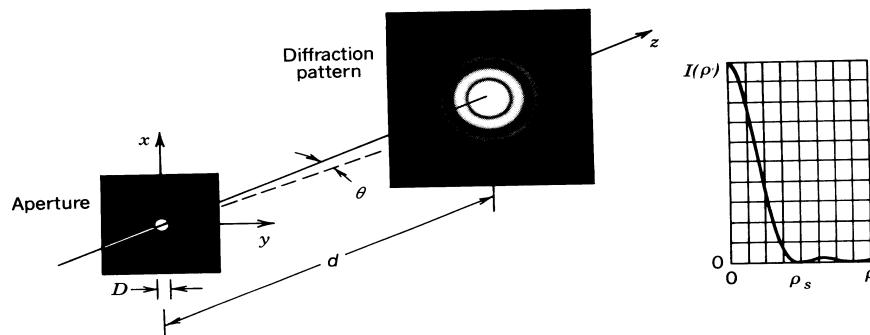
Fraunhofer diffraction from a rectangular aperture. The central lobe of the pattern has half-angular widths $\theta_x = \lambda/D_x$ and $\theta_y = \lambda/D_y$.

Fraunhofer diffraction pattern from a rectangular aperture. The central lobe of the pattern has half-angular widths $\theta_x = \lambda/D_x$ and $\theta_y = \lambda/D_y$.

B) Circular aperture (pinhole) illuminated by normal plane wave

$$t(x, y) = \begin{cases} 1 & \text{for } x^2 + y^2 \leq a^2 \\ 0 & \text{elsewhere} \end{cases}$$

$$I(x, y, z_B) \sim \left[\frac{J_1\left(\frac{ka}{z_B} \sqrt{x^2 + y^2}\right)}{\frac{ka}{z_B} \sqrt{x^2 + y^2}} \right]^2 \rightarrow \text{Bessel function}$$



The Fraunhofer diffraction pattern from a circular aperture produces the Airy pattern with the radius of the central disk subtending an angle $\theta = 1.22\lambda/D$.

The Fraunhofer diffraction pattern from a circular aperture produces the Airy pattern with the radius of the central disk subtending an angle $\theta \approx 1.22\lambda/D$.

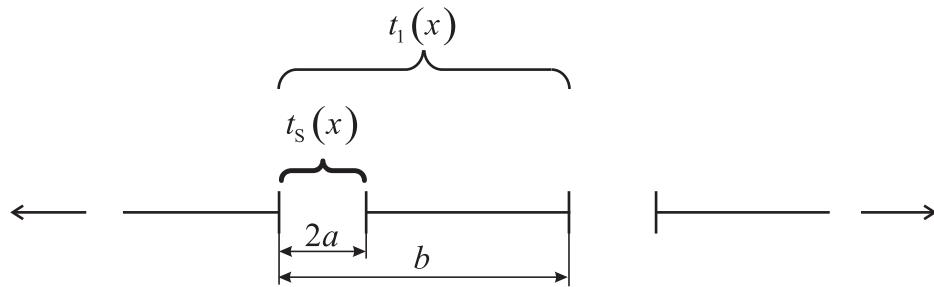
The first zero of the Bessel function (defines size of Airy disk):

$$\frac{ka}{z_B} \rho = 1.22\pi \rightarrow \frac{\rho}{z_B} = \frac{0.61\lambda}{a} \text{ with } \rho^2 = x^2 + y^2$$

So-called angle of aperture: $\Theta = \frac{2\rho}{z_B} = \frac{1.22\lambda}{a}$ (in small angle approximation)

C) One-dimensional periodic structure (grating) illuminated by normal plane wave

For periodic arrangements of slits, we can gain a deeper insight into the structure of the diffraction pattern. Let us assume a periodic arrangement of apertures with a period b and a size of each slit $2a$:



Then, we can express the mask function $t(x)$ as:

$$t(x) = \sum_{n=0}^{N-1} t_l(x - nb) \text{ with } t_l(x) = \begin{cases} t_s(x) & \text{for } |x| \leq a \\ 0 & \text{elsewhere} \end{cases}$$

Here $t_s(x)$ is the transmission function of a single slit and N is the total number of slits.

The Fourier transform of the mask is then given by

$$T\left(k \frac{x}{z_B}\right) \sim \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} t_l(x' - nb) \exp\left(-ik \frac{x}{z_B} x'\right) dx'$$

With the variable substitution $x' - nb = X'$, we can simplify further:

$$T\left(k \frac{x}{z_B}\right) \sim \sum_{n=0}^{N-1} \int_{-\alpha}^{\alpha} t_s(X') \exp\left(-ik \frac{x}{z_B} X'\right) \exp\left(-ik \frac{x}{z_B} nb\right) dX'$$

$$T\left(k \frac{x}{z_B}\right) \sim T_s\left(k \frac{x}{z_B}\right) \sum_{n=0}^{N-1} \exp\left(-ik \frac{x}{z_B} nb\right)$$

We see that the Fourier transform T_s of the single slit t_s appears. The second factor has its origin in the periodic arrangement. We can identify this expression as a geometrical series and perform the summation by using the following formula to define the so-called grating function:

$$\left| \sum_{n=0}^{N-1} \exp(-i\delta n) \right| = \left| \frac{\sin(N \frac{\delta}{2})}{\sin(\frac{\delta}{2})} \right|$$

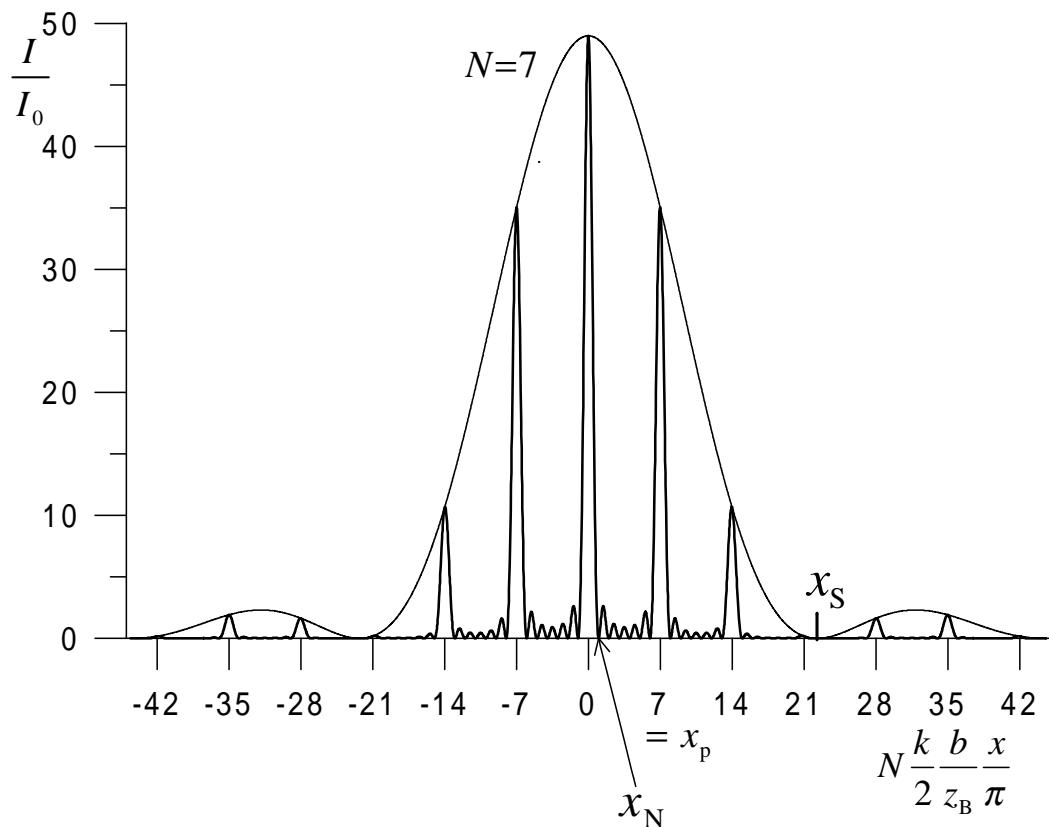
Thus, we finally write:

$$T\left(k \frac{x}{z_B}\right) \sim T_s\left(k \frac{x}{z_B}\right) \frac{\sin\left(N \frac{k}{2} \frac{x}{z_B} b\right)}{\sin\left(\frac{k}{2} \frac{x}{z_B} b\right)}$$

For the particular case of a simple grating of slit apertures with $t_s(x)=1$ we have

$$T_1\left(k \frac{x}{z_B}\right) = \text{sinc}\left(k \frac{x}{z_B} a\right)$$

and therefore $I \sim \text{sinc}^2\left(k \frac{x}{z_B} a\right) \frac{\sin^2\left(N \frac{k}{2} \frac{x}{z_B} b\right)}{\sin^2\left(\frac{k}{2} \frac{x}{z_B} b\right)}$



We find three important parameters for the diffraction pattern of such a grating:

- Global width of the diffraction pattern → first zero of slit function T_s

$$k \frac{x_s}{z_B} a = \pi \rightarrow x_s = \frac{\lambda z_B}{2a}$$

The width x_s of the entire far-field diffraction pattern (largest length scale in the pattern) is determined by the size a of the individual slit (smallest length scale in the mask).

- Positions of local maxima of the diffraction pattern
→ maxima of grating function, where the denominator vanishes

$$\max \left(\frac{\sin^2 \left(N \frac{k}{2} \frac{x}{z_B} b \right)}{\sin^2 \left(\frac{k}{2} \frac{x}{z_B} b \right)} \right) \rightarrow \frac{k}{2} \frac{x_p}{z_B} b = n\pi \rightarrow x_p = n \frac{\lambda z_B}{b}$$

These are the so-called diffraction orders (second largest length scale in the pattern), which are determined exclusively by the grating period (second smallest length scale in the mask).

- Width of local maxima → zeroes of grating function

$$N \frac{k}{2} \frac{x_N}{z_B} b = \pi \rightarrow x_N = \frac{\lambda z_B}{Nb}$$

The width of a maximum in the far-field diffraction pattern x_N (smallest length scale in the pattern) is determined by $N * b$, which is the total size of the mask (largest length scale of the mask).

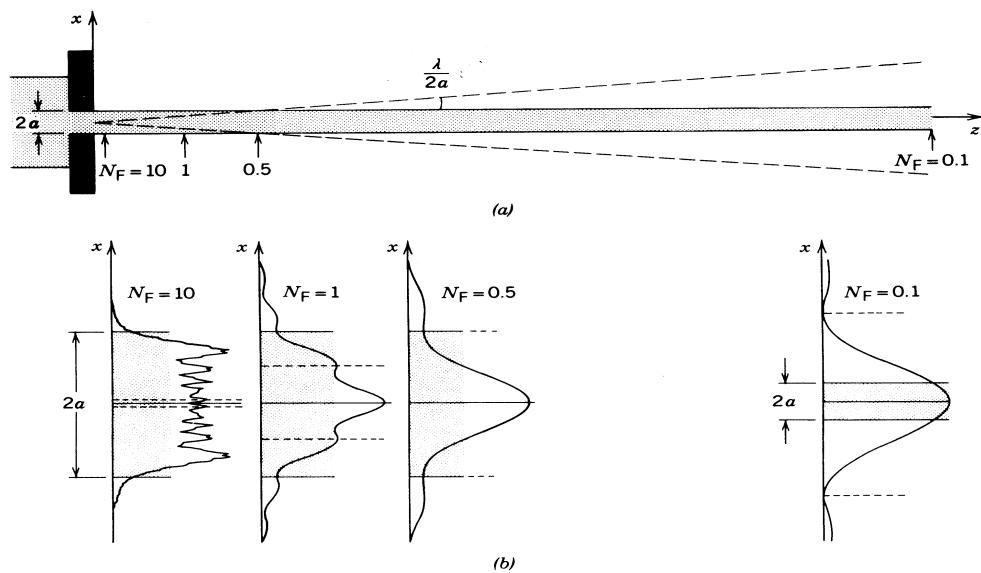
These observations are consistent with the general property of the Fourier-transform: small scales in position space give rise to a broad angular spectrum and vice versa.

4.4 Remarks on Fresnel diffraction

Fresnel number $N_F = \frac{a}{\lambda} \frac{a}{z_B}$

diffraction length from Gaussian beams $z_0 = \frac{\pi a^2}{\lambda}$

- $N_F \gtrsim 10$ (a large, λz_B small, $z_B < 1/30 z_0$) → shadow
- $10 \gtrsim N_F \gtrsim 0.1$ ($1/30 z_0 < z_B < 3z_0$) → Fresnel diffraction
- $N_F \lesssim 0.1$ ($z_B > 3z_0$) → Fraunhofer
→ FT of aperture



Fresnel diffraction from a slit of width $D = 2a$. (a) Shaded area is the geometrical shadow of the aperture. The dashed line is the width of the Fraunhofer diffracted beam. (b) Diffraction pattern at four axial positions marked by the arrows in (a) and corresponding to the Fresnel numbers $N_F = 10, 1, 0.5$, and 0.1 . The shaded area represents the geometrical shadow of the slit. The dashed lines at $|x| = (\lambda/D)d$ represent the width of the Fraunhofer pattern in the far field. Where the dashed lines coincide with the edges of the geometrical shadow, the Fresnel number $N_F = a^2/\lambda d = 0.5$.

Fresnel diffraction from a slit of width $D = 2a$. (a) Shaded area is the geometrical shadow of the aperture. The dashed line is the width of the Fraunhofer diffraction beam. (b) Diffraction pattern at four axial positions marked by the arrows in (a) and corresponding to the Fresnel numbers $N_F = 10, 1, 0.5$ and 0.1 . The shaded area represents the geometrical shadow of the slit. The dashed lines at $|x| = (\lambda / D)d$ represent the width of the Fraunhofer pattern in the far field. Where the dashed lines coincide with the edges of the geometrical shadow, the Fresnel number $N_F = a^2 / \lambda d = 0.5$.

5. Fourier optics- optical filtering

From previous chapters, we know how to propagate the optical field through homogeneous space, and we also know the transfer function of a thin lens. Thus, we have all the tools at hand to describe optical imaging. While detailed designs of high-resolution optical systems require a consideration of non-paraxial effects, the paraxial approximation is usually sufficient to obtain a fundamental understanding of optical systems. Hence, we use the paraxial approximation here.

Many imaging systems exploit the appearance of the Fourier transform of the original object in the so-called Fourier plane of the system in order to manipulate the angular spectrum of the object in this plane. Accordingly, this field of optical science is called Fourier optics. In the following, we will see that with the right setup of our imaging system, we can generate the Fourier transform of the object at a much shorter distance than by far field diffraction in the Fraunhofer approximation.

The general idea of Fourier optics is the following:

- 1) An imaging system generates the Fourier transform of the object in the Fourier plane.
- 2) A spatial filter (e.g. an aperture) in the Fourier plane manipulates the field.
- 3) Another imaging system performs the Fourier back-transform and hence results in a manipulated image.

Mathematical concept:

- propagation in free space → calculated in Fourier domain
- interaction with lens or filter → calculated in position space

5.1 Imaging of arbitrary optical fields with a thin lens

5.1.1 Transfer function of a thin lens

A thin lens changes only the phase of the optical field, since no diffraction occurs due to its infinitesimal thickness. By definition, it transforms a spherical wave into a plane wave (in real space). If we write down this definition in paraxial approximation, we obtain

$$\underbrace{-\frac{i}{\lambda f} \exp(ikf) \exp\left[i \frac{k}{2f}(x^2 + y^2)\right]}_{\text{spherical wave in paraxial approximation}} t_L(x, y) = \underbrace{-\frac{i}{\lambda f} \exp(ikf)}_{\text{plane wave}}$$

And therefore the response function for a thin lens is given by (see chapter 9.6):

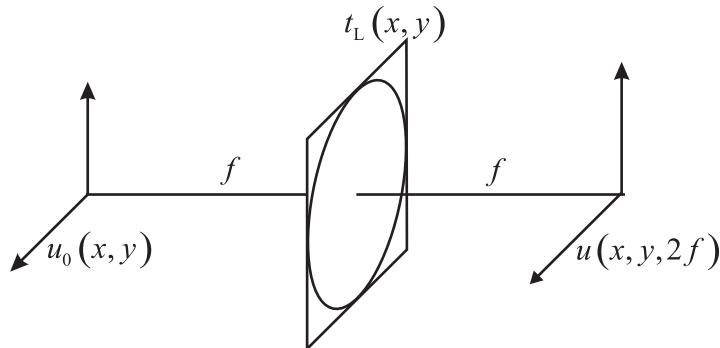
$$t_L(x, y) = \exp\left[-i \frac{k}{2f}(x^2 + y^2)\right] \quad (\text{by product})$$

By taking the Fourier transform of the response function, we find the transfer function in the Fourier domain:

$$T_L(\alpha, \beta) = -i \frac{\lambda f}{(2\pi)^2} \exp \left[i \frac{f}{2k} (\alpha^2 + \beta^2) \right] \text{ (by convolution)}$$

5.1.2 Optical imaging using the 2f-setup

Let us now consider optical imaging. We place our object in the first focus of a thin lens, with a field distribution $u_0(x, y)$, and follow the usual recipe for light propagation.



A) Spectrum in object plane

$$U_0(\alpha, \beta) = \text{FT}[u_0(x, y)]$$

B) Propagation from object to lens in the paraxial approximation
(lens positioned at distance f)

$$U_-(\alpha, \beta; f) = H_F(\alpha, \beta; f) U_0(\alpha, \beta)$$

$$U_-(\alpha, \beta; f) = \exp(i k f) \exp \left[-\frac{i}{2k} (\alpha^2 + \beta^2) f \right] U_0(\alpha, \beta)$$

C) Interaction with lens

(multiplication in position space or convolution in Fourier domain, as one plane wave transforms into multiple due to the action of the lens)

$$u_+(x, y, f) = t_L(x, y) u_-(x, y, f)$$

$$U_+(\alpha, \beta; f) = T_L(\alpha, \beta) * U_-(\alpha, \beta; f)$$

$$\begin{aligned} &= -i \frac{\lambda f}{(2\pi)^2} \exp(i k f) \int \int_{-\infty}^{\infty} \exp \left\{ i \frac{f}{2k} [(\alpha - \alpha')^2 + (\beta - \beta')^2] \right\} \\ &\quad \cdot \exp \left[-\frac{i}{2k} (\alpha'^2 + \beta'^2) f \right] U_0(\alpha', \beta') d\alpha' d\beta' \end{aligned}$$

RED: transfer function of lens; BLUE: transfer function of free space

D) Propagation from lens to image plane

$$U(\alpha, \beta; 2f) = H_F(\alpha, \beta; f) U_+(\alpha, \beta; f)$$

$$U(\alpha, \beta; 2f) = -\frac{\lambda f}{(2\pi)^2} \exp(2ikf) \iint_{-\infty}^{\infty} \exp\left\{ i \frac{f}{2k} [(\alpha - \alpha')^2 + (\beta - \beta')^2] \right\} \cdot \exp\left[-\frac{i}{2k} (\alpha'^2 + \beta'^2) f \right] \exp\left[-\frac{i}{2k} (\alpha^2 + \beta^2) f \right] U_0(\alpha', \beta') d\alpha' d\beta'$$

Quadratic terms from the free space transfer function $\left[-\frac{i}{2k} (\alpha'^2 + \beta'^2) f \right]$ and $\left[-\frac{i}{2k} (\alpha^2 + \beta^2) f \right]$ cancel with quadratic terms from the lens transfer function $i \frac{f}{2k} [(\alpha - \alpha')^2 + (\beta - \beta')^2]$ to leave only the mixed terms.

$$U(\alpha, \beta; 2f) = -\frac{\lambda f}{(2\pi)^2} \exp(2ikf) \iint_{-\infty}^{\infty} U_0(\alpha', \beta') \exp\left[-i \frac{f}{k} (\alpha \alpha' + \beta \beta') \right] d\alpha' d\beta'$$

$$= -\frac{\lambda f}{(2\pi)^2} \exp(2ikf) u_0\left(-\frac{f}{k}\alpha, -\frac{f}{k}\beta\right)$$

We see that the spectrum in the image plane is given by the optical field in the object plane.

E) Fourier back transform in the image plane

$$u(x, y, 2f) = \text{FT}^{-1}[U(\alpha, \beta; 2f)]$$

$$= -\frac{\lambda f}{(2\pi)^2} \exp(2ikf) \iint_{-\infty}^{\infty} u_0\left(-\frac{f}{k}\alpha, -\frac{f}{k}\beta\right) \exp[i(\alpha x + \beta y)] d\alpha d\beta$$

Note that as expected, the arguments of the object plane field in the above integral has dimensions of length. Thus, we perform the coordinate transformation

$$x' = -\frac{f}{k}\alpha, \quad y' = -\frac{f}{k}\beta \quad \rightarrow \quad d\alpha = -\frac{2\pi}{\lambda f} dx', \quad d\beta = -\frac{2\pi}{\lambda f} dy'$$

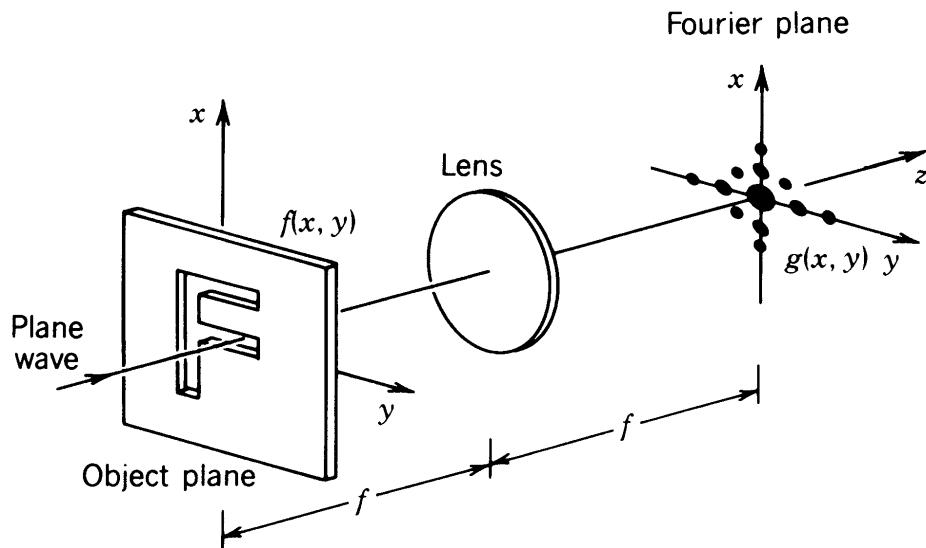
to simplify the expression for the object plane field in the above Fourier back transform:

$$\rightarrow u(x, y, 2f) = -i \frac{1}{\lambda f} \exp(2ikf) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_0(x', y') \exp\left[-i \frac{k}{f}(xx' + yy')\right] dx' dy'$$

$$u(x, y, 2f) = -i \frac{(2\pi)^2}{\lambda f} \exp(2ikf) U_0\left(\frac{k}{f}x, \frac{k}{f}y\right)$$

The image in the second focal plane is thus just the Fourier transform of the optical field in the object plane. This is similar to the far field in Fraunhofer approximation, but but with $z_B \leftrightarrow f$, i.e. the propagation distance is replaced by the focal length of the lens. This finding allows us to optically perform a Fourier transform over shorter distances than relying on beam propagation and with adjustable size in the Fourier plane (by varying the focal length).

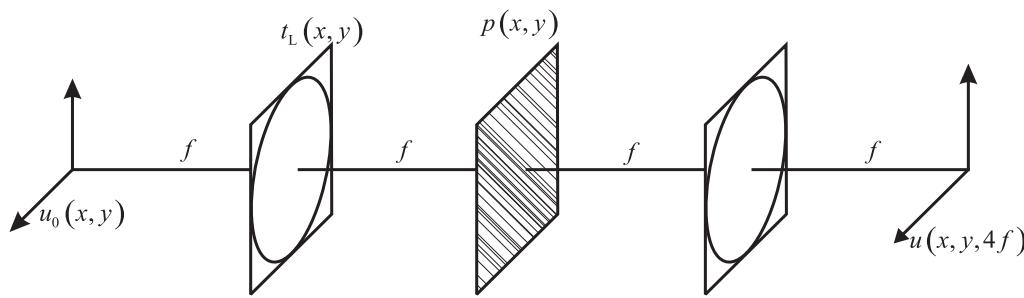
Furthermore, it is possible to directly manipulate the spectrum in the Fourier plane.



5.2 Optical filtering and image processing

5.2.1 4f-setup

For image manipulation (spectral filtering), it would be advantageous if we could also perform a Fourier back transform by means of optical imaging. This leads to the so-called 4f-setup:



The filtering (manipulation) occurs in the focal plane (Fourier plane after $2f$) by applying a transmission mask $p(x,y)$. In order to retrieve the filtered image, i.e. transform it back to real space, we use a second lens:

We know from the previous section that the image in the Fourier plane is the FT of the optical field in the object plane.

$$u(x,y,2f) = AU_0\left(\frac{k}{f}x, \frac{k}{f}y\right)$$

We have to compute the image of the second lens after the manipulation of the spectrum in the Fourier plane.

Our goal is to derive the transfer function $H_A(\alpha,\beta;4f)$ of the complete imaging system, as defined by:

$$u(-x,-y,4f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_A(\alpha,\beta;4f) U_0(\alpha,\beta) \exp[-i(\alpha x + \beta y)] d\alpha d\beta$$

Note: We will see in the following calculation that the second lens performs a Fourier transform $\sim \exp[-i(\alpha x + \beta y)]$. In order to obtain the required back transform, we must transform to the mirrored coordinates $x \rightarrow -x$, $y \rightarrow -y$.

Furthermore, we will see that the transmission mask $p(x,y)$ contains all constraints of the system (e.g. a limited aperture) and optical filtering (which we can design).

A) Field behind the transmission mask

$$u_+(x,y,2f) = u(x,y,2f)p(x,y) \sim AU_0\left(\frac{k}{f}x, \frac{k}{f}y\right)p(x,y)$$

B) Second lens \rightarrow Fourier back transform of field distribution

$$u(x,y,4f) = -i \frac{(2\pi)^2}{\lambda f} \exp(2ikf) U_+\left(\frac{k}{f}x, \frac{k}{f}y; 2f\right)$$

This can be related to the initial spectrum in the object plane U_0 :

$$\begin{aligned} u(x, y, 4f) &\sim \iint_{-\infty}^{\infty} u_+(x', y', 2f) \exp\left[-i \frac{k}{f}(xx' + yy')\right] dx' dy' \\ &\sim \iint_{-\infty}^{\infty} U_0\left(\frac{k}{f}x', \frac{k}{f}y'\right) p(x', y') \exp\left[-i \frac{k}{f}(xx' + yy')\right] dx' dy' \end{aligned}$$

Here we are not concerned about the specific amplitudes and simply write the proportionality symbol \sim :

To obtain the anticipated form of a Fourier back transform, we need to perform a coordinate transformation:

$$\alpha = \frac{k}{f}x', \beta = \frac{k}{f}y'$$

This simplifies the arguments of U_0 :

$$u(x, y, 4f) \sim \iint_{-\infty}^{\infty} U_0(\alpha, \beta) p\left(\frac{f}{k}\alpha, \frac{f}{k}\beta\right) \exp\left[-i(\alpha x + \beta y)\right] d\alpha d\beta$$

By transforming to mirrored coordinates $x \rightarrow -x$, $y \rightarrow -y$, we get obtain a Fourier back transform

$$u(-x, -y, 4f) \sim \iint_{-\infty}^{\infty} p\left(\frac{f}{k}\alpha, \frac{f}{k}\beta\right) U_0(\alpha, \beta) \exp\left[i(\alpha x + \beta y)\right] d\alpha d\beta$$

Hence, by comparing with the above definition of the transfer function, we can identify

$$H_A(\alpha, \beta; 4f) \sim p\left(\frac{f}{k}\alpha, \frac{f}{k}\beta\right)$$

Note that the image is inverted compared to the original object.

Summary

- Fourier amplitudes in the object plane are multiplied by the transmission mask
- transmission mask \rightarrow transfer function
- coordinates of image \rightarrow mirrored coordinates of object

In position space, we can formulate the propagation through a 4f-system by convolution with the response function $h_A(x, y)$

$$u(-x, -y, 4f) = \iint_{-\infty}^{\infty} h_A(x - x', y - y') u_0(x', y') dx' dy'$$

As usual, the response function is given by the Fourier transform of the transfer function:

$$h_A(x, y) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} H_A(\alpha, \beta; 4f) \exp\{i[\alpha x + \beta y]\} d\alpha d\beta$$

From above, we have the transfer function $H_A(\alpha, \beta; 4f) \sim p\left(\frac{f}{k}\alpha, \frac{f}{k}\beta\right)$

$$h_A(x, y) \sim \iint_{-\infty}^{\infty} p\left(\frac{f}{k}\alpha, \frac{f}{k}\beta\right) \exp\left\{i[\alpha x + \beta y]\right\} d\alpha d\beta$$

We introduce the coordinate transform $\left(\bar{x} = \frac{f}{k}\alpha, \bar{y} = \frac{f}{k}\beta\right)$

$$h_A(x, y) \sim \iint_{-\infty}^{\infty} p(\bar{x}, \bar{y}) \exp\left\{i\frac{k}{f}[-\bar{xx} + \bar{yy}]\right\} d\bar{x} d\bar{y} \sim P\left[-\frac{k}{f}x, -\frac{k}{f}y\right]$$

The response function is proportional to the Fourier transform of the transmission mask.

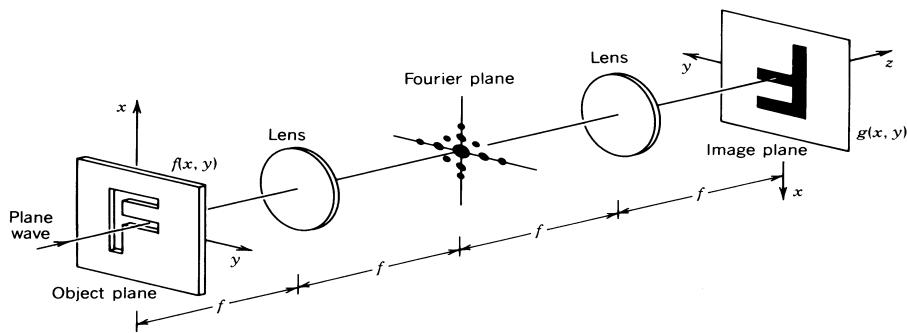
$$u(-x, -y, 4f) \sim \iint_{-\infty}^{\infty} P\left[\frac{k}{f}(x' - x), \frac{k}{f}(y' - y)\right] u_0(x', y') dx' dy'$$

As usual, the field after the 4f-system is obtained by convolving the object plane field with the response function. Note again that the image is inverted with respect to the object.

5.2.2 Examples of aperture functions

Example 1: The ideal image (infinite aperture)

Note that we use the paraxial approximation → requires limited angular spectrum



The 4-f system performs a Fourier transform followed by an inverse Fourier transform, so that the image is a perfect replica of the object.

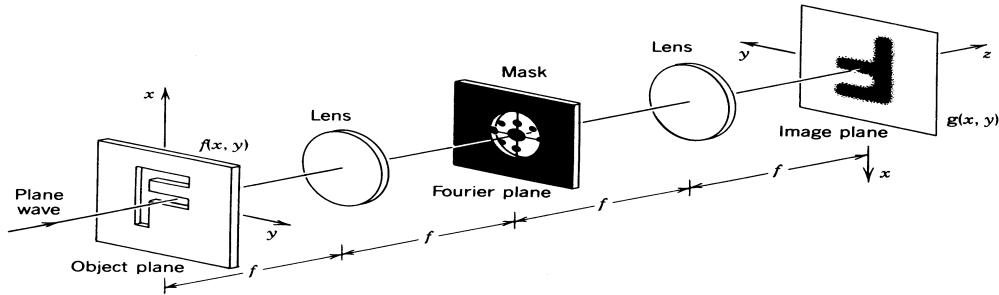
The 4-f system performs a Fourier transform, followed by an inverse Fourier transform such that the image is a perfect replica of the object (perfect only within the Fresnel approximation).

For an infinite, perfectly transmitting aperture, the transmission mask is uniform

$$p = 1 \rightarrow P \sim \delta(x)\delta(y) \rightarrow u(-x, -y, 4f) \sim u_0(x, y) \rightarrow \text{mirrored original}$$

Example 2: Finite circular aperture

$$p(x, y) = \begin{cases} 1 & \text{for } x^2 + y^2 \leq (D/2)^2 \\ 0 & \text{elsewhere} \end{cases}$$



Spatial filtering. The transparencies in the object and Fourier planes have complex amplitude transmittances $f(x, y)$ and $p(x, y)$. A plane wave traveling in the z direction is modulated by the object transparency, Fourier transformed by the first lens, multiplied by the transmittance of the mask in the Fourier plane and inverse Fourier transformed by the second lens. As a result, the complex amplitude in the image plane $g(x, y)$ is a filtered version of $f(x, y)$. The system has a transfer function $\mathcal{H}(\nu_x, \nu_y) = p(\lambda f \nu_x, \lambda f \nu_y)$.

Spatial filtering. The transparencies in the object and Fourier planes have complex amplitude transmittances $f(x, y)$ and $p(x, y)$. A plane wave traveling in the z direction is modulated by the object transparency, Fourier transformed by the first lens, multiplied by the transmittance of the mask in the Fourier plane and inverse Fourier transformed by the second lens. As a result, the complex amplitude in the image plane $g(x, y)$ is a filtered version of the original field $f(x, y)$ in the object plane. The system has a transfer function $H(\nu_x, \nu_y) = p(\lambda f \nu_x, \lambda f \nu_y)$.

Transmission function (obtained by the appropriate substitution into the transmission mask $p(x, y)$) :

$$H_A(\alpha, \beta; 4f) \sim \begin{cases} 1 & \text{for } \left(\frac{f}{k}\alpha\right)^2 + \left(\frac{f}{k}\beta\right)^2 \leq (D/2)^2 \\ 0 & \text{elsewhere} \end{cases}$$

- finite aperture truncates large angular frequencies beyond a certain cutoff which obviously increases with the diameter of the aperture and decreases with focal length (since the lens would focus less tightly on to the filter)
- acts as a low-pass filter
- determines optical resolution

5.2.3 Optical resolution

A finite aperture acts as a low pass filter for angular frequencies:

$$\left(\frac{f}{k}\alpha\right)^2 + \left(\frac{f}{k}\beta\right)^2 \leq (D/2)^2 \rightarrow (\alpha)^2 + (\beta)^2 \leq \left(\frac{k}{f}D/2\right)^2$$

With $\rho^2 = \alpha^2 + \beta^2$ we can define an upper limit for the angular frequencies ρ_{\max} which are transmitted (bandwidth of the system)

$$\rho_{\max}^2 = \frac{k^2}{f^2} \left(\frac{D}{2} \right)^2 \rightarrow \boxed{\rho_{\max} = \frac{2\pi n D}{\lambda f} / 2}$$

Translated to position space, the smallest transmitted structural information is given by:

$$\Delta r_{\min} \approx \frac{2\pi}{\rho_{\max}} = \frac{2\lambda f}{nD}$$

A more precise definition of the optical resolution can be derived the following way:

$$\begin{aligned} u(-x, -y, 4f) &\sim \int \int_{-\infty}^{\infty} P \left[\frac{k}{f} (x' - x), \frac{k}{f} (y' - y) \right] u_0(x', y') dx' dy' \\ &\sim \int \int_{-\infty}^{\infty} \frac{J_1 \left[\frac{kD}{2f} \sqrt{(x' - x)^2 + (y' - y)^2} \right]}{\frac{kD}{2f} \sqrt{(x' - x)^2 + (y' - y)^2}} u_0(x', y') dx' dy', \end{aligned}$$

where the response function takes the form of a Bessel function for a circular aperture. In other words, each point of the object x_0, y_0 gives rise to an Airy disk (pixel) in the image:

$$\sim \left[\frac{J_1 \left(\frac{kD}{2f} \sqrt{(x_0 - x)^2 + (y_0 - y)^2} \right)}{\frac{kD}{2f} \sqrt{(x_0 - x)^2 + (y_0 - y)^2}} \right]^2$$

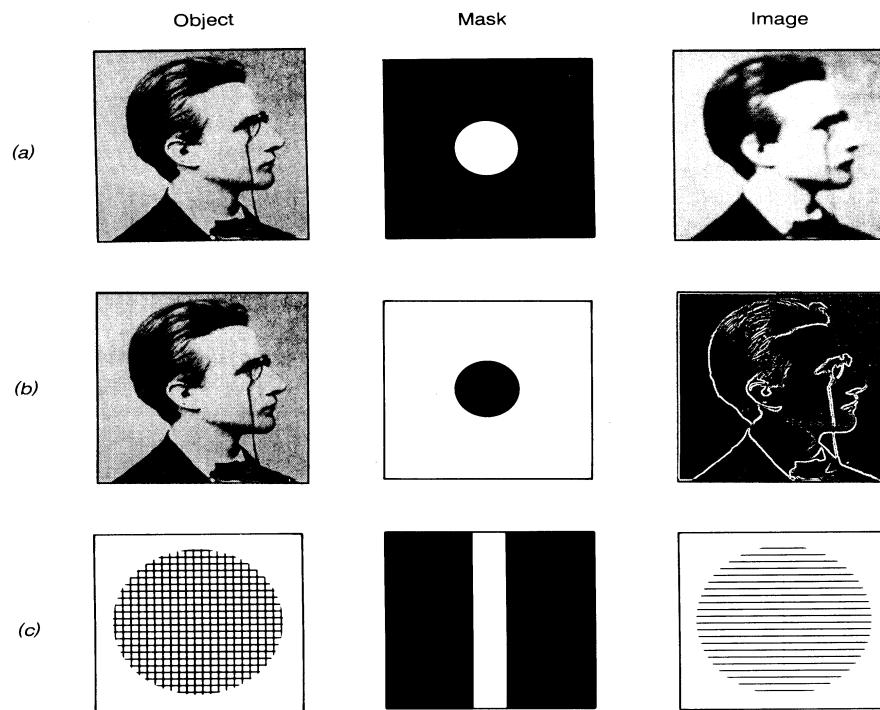
We can thus define the limit of optical resolution by the following requirement:

Two objects in the object plane can be independently resolved in the image plane as long as the intensity maximum of one object is not closer to the other object than its first intensity minimum:

$$\frac{kD}{2f} \Delta r_{\min} = 1.22\pi$$

Hence we find: $\boxed{\Delta r_{\min} = \frac{1.22\lambda f}{nD}}$

Further examples for 4f filtering



Examples of object, mask, and filtered image for three spatial filters: (a) low-pass filter; (b) high-pass filter; (c) vertical-pass filter. Black means the transmittance is zero and white means the transmittance is unity.

Examples of object, mask, and filtered image for three spatial filters: (a) low-pass filter; (b) high-pass filter; (c) vertical pass filter. Black represents a zero transmittance while white represents a unity transmittance.

6. The polarization of electromagnetic waves

We are interested in the temporal evolution of the electric field vector $\mathbf{E}_r(\mathbf{r},t)$. In the previous chapters, we mostly used a scalar description, assuming linearly polarized light. However, in general one has to consider the vectorial nature, i.e. the polarization state, of the electric field vector.

We know that the normal modes of homogeneous isotropic dielectric media are plane waves $\mathbf{E}(\mathbf{r},t) = \mathbf{E} \exp\{\mathbf{i}[\mathbf{k}(\omega)\mathbf{r} - \omega t]\}$.

If we assume propagation in z direction (\mathbf{k} -vector points in z -direction), $\text{div}\mathbf{E}(\mathbf{r},t) = 0$ implies that we can have two nonzero transverse field components, which are the x and y components E_x, E_y .

The orientation and shape of the area which the (real) electric field vector traces out is in general an ellipse. There are two special cases:

- The ellipse can degrade to a line: linear polarization.
- The ellipse can become a circle: circular polarization.

6.1 Polarization of normal modes in isotropic media

$$\mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ k \end{pmatrix} \quad \rightarrow \text{propagation in } z \text{ direction}$$

The evolution of the real electric field vector is given by

$$\mathbf{E}_r(\mathbf{r},t) = \Re\{\mathbf{E} \exp[\mathbf{i}(kz - \omega t)]\}$$

Because the field is transversal, we have two free complex field components

$$\mathbf{E} = \begin{pmatrix} E_x \exp(\mathbf{i}\varphi_x) \\ E_y \exp(\mathbf{i}\varphi_y) \\ 0 \end{pmatrix} \quad \text{with } E_{x,y} \text{ and } \varphi_{x,y} \text{ being real}$$

Thus the **real** electric field vector is given by

$$\mathbf{E}_r(\mathbf{r},t) = \begin{pmatrix} E_x \cos(\omega t - kz - \varphi_x) \\ E_y \cos(\omega t - kz - \varphi_y) \\ 0 \end{pmatrix}$$

Usually, only the relative phase $\delta = \varphi_y - \varphi_x$ is of interest.

Conclusion

In general, normal modes in isotropic, dispersive media are elliptically polarized. The field amplitudes E_x, E_y and the phase difference $\delta = (\varphi_y - \varphi_x)$ are free parameters.

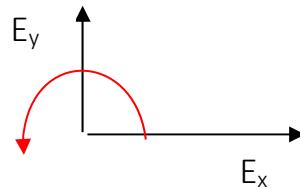
6.2 Polarization states

Let us have a look at the different possible parameter values:

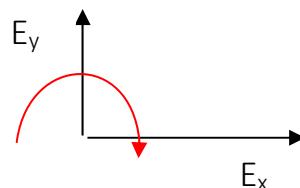
A) linear polarization $\rightarrow \delta = n\pi$ or $E_x = 0$ or $E_y = 0$

B) circular polarization $\rightarrow E_x = E_y = E, \delta = \pm\pi/2$

$\delta = +\pi/2 \rightarrow$ counterclockwise rotation



$\delta = -\pi/2 \rightarrow$ clockwise rotation



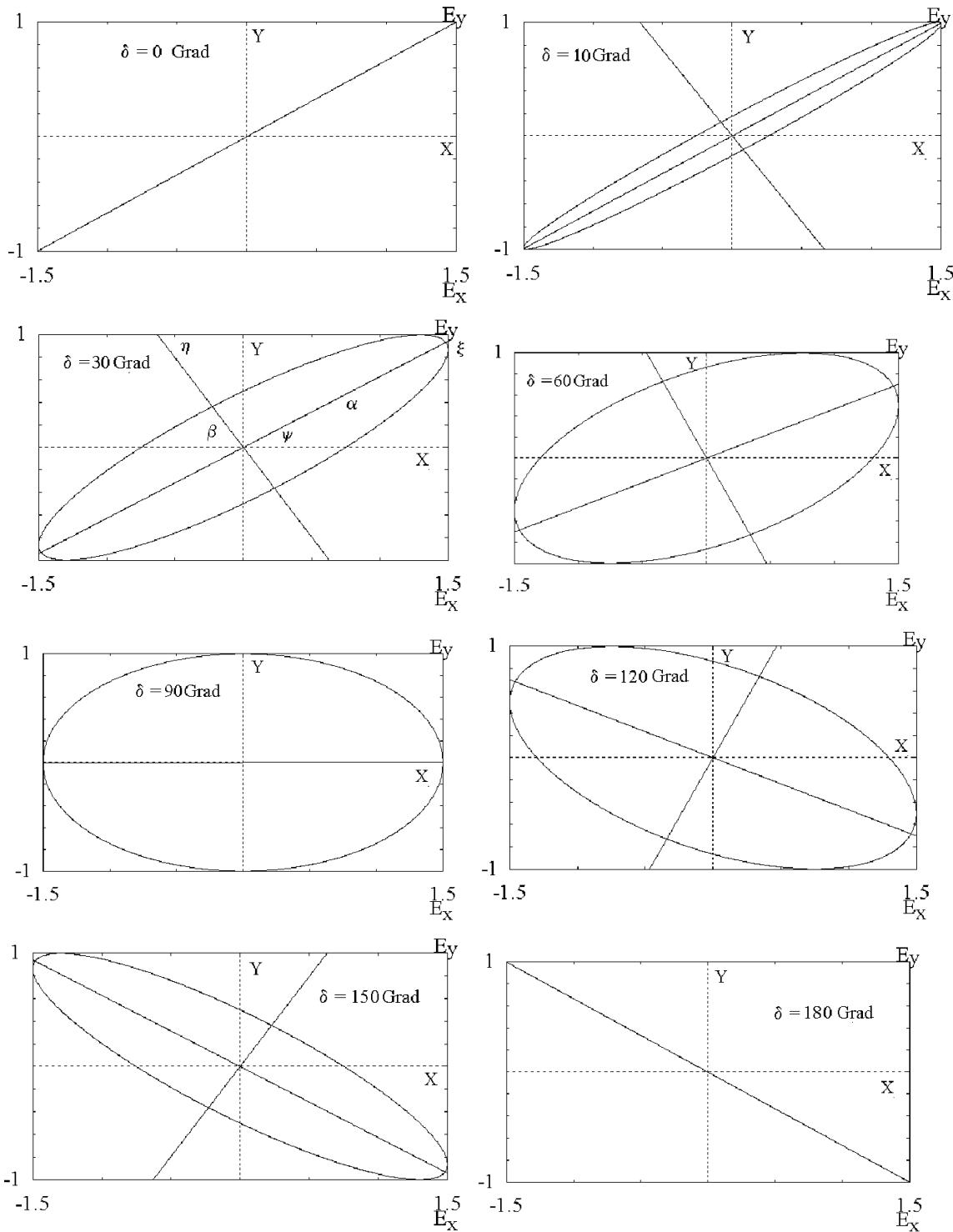
Note that these diagrams are in the perspective of an observer looking contrary to the propagation direction.

C) elliptic polarization $\rightarrow E_x \neq E_y \neq 0, \delta \neq n\pi$

$0 < \delta < \pi \rightarrow$ counterclockwise

$\pi < \delta < 2\pi \rightarrow$ clockwise

Examples



Remark on the correspondence between linear and circular polarization basis

A linearly polarized wave can be written as a superposition of two counter-rotating circularly polarized waves. Example: Let's observe the temporal evolution at a fixed position $kz = 0$ with $\delta = \pm\pi/2$.

$$E \begin{pmatrix} \cos(\omega t) \\ \sin(\omega t) \\ 0 \end{pmatrix} + E \begin{pmatrix} \cos(\omega t) \\ -\sin(\omega t) \\ 0 \end{pmatrix} = 2E \begin{pmatrix} \cos(\omega t) \\ 0 \\ 0 \end{pmatrix}$$

7. Principles of optics in crystals

In this chapter, we will treat light propagation in anisotropic media, which is a much more complicated case than for the isotropic media from before. As in the isotropic case, we will seek the normal modes and assume homogeneous anisotropic media in order to keep things simple.

7.1 Susceptibility and dielectric tensor

before: **isotropy** (optical properties independent of direction)

now: **anisotropy** (optical properties depend on direction)

The common reason for anisotropy in many optical media (in particular crystals) is that the polarization \mathbf{P} depends on the direction of the electric field vector. The underlying reason is that in crystals the atoms have a periodic distribution with different symmetries in different directions.

Prominent examples of anisotropic materials and their applications are:

- Lithium niobate → electro-optical modulators,
- Quartz → polarizer,
- liquid crystals → displays.

In order to keep things as simple as possible, we make the following **assumptions**:

- monochromaticity → single frequency ω ,
- plane wave → single spatial frequency,
- no absorption → real valued $\epsilon = \epsilon'$

From previous chapters we know that in isotropic media the normal modes are elliptically polarized, monochromatic plane waves. The question is how the normal modes and their dispersion relation in an anisotropic medium look like.

Before (isotropic)

$$\bar{\mathbf{P}}(\mathbf{r}, \omega) = \epsilon_0 \chi(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

$$\bar{\mathbf{D}}(\mathbf{r}, \omega) = \epsilon_0 \epsilon(\omega) \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

In the following, we will write $\bar{\mathbf{E}} \rightarrow \mathbf{E}$, because we assume monochromatic light and the frequency ω is just a parameter.

Now (anisotropic)

$$P_i(\mathbf{r}, \omega) = \epsilon_0 \sum_{j=1}^3 \underbrace{\chi_{ij}(\omega)}_{\text{tensor components}} E_j(\mathbf{r}, \omega)$$

The linear susceptibility tensor has $3 \times 3 = 9$ tensor components. Direct consequences of this relation between polarization \mathbf{P} and electric field \mathbf{E} are:

- $\mathbf{P} \nparallel \mathbf{E}$: the polarization is **not** necessarily **parallel** to the electric field

- The tensor elements χ_{ij} depend on the structure of the crystal. However, we do not need to know the microscopic structure because of the different length scales involved (optics: $5 \cdot 10^{-7} \text{ m}$; crystal: $5 \cdot 10^{-10} \text{ m}$), but the field is influenced by the symmetries of the crystal (see next section).
- In complete analogy we find for the D field:

$$D_i(\mathbf{r}, \omega) = \epsilon_0 \sum_{j=1}^3 \epsilon_{ij}(\omega) E_j(\mathbf{r}, \omega)$$

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \hat{\boldsymbol{\epsilon}}(\omega) \mathbf{E}(\mathbf{r}, \omega)$$

As for the polarization we find:

- $\mathbf{D} \nparallel \mathbf{E}$

We introduce the following notation:

- $\hat{\boldsymbol{\chi}} = (\chi_{ij})$ → susceptibility tensor
- $\hat{\boldsymbol{\epsilon}} = (\epsilon_{ij})$ → dielectric tensor
- $\hat{\boldsymbol{\sigma}} = (\hat{\boldsymbol{\epsilon}})^{-1} = (\sigma_{ij})$ → inverse dielectric tensor $\sum_{j=1}^3 \sigma_{ij}(\omega) D_j(\mathbf{r}, \omega) = \epsilon_0 E_i(\mathbf{r}, \omega)$

The following properties of the dielectric and inverse dielectric tensor are important:

- $\sigma_{ij}, \epsilon_{ij}$ are real in the transparent region (omit ω) since we have no losses (see our assumptions above)
- The tensors are symmetric (Hermitian), only 6 components are independent $\epsilon_{ij} = \epsilon_{ji}, \sigma_{ij} = \sigma_{ji}$.
- It is known (see any book on linear algebra) that such symmetric tensors/matrices can be diagonalized by rotation (orthogonal transformation) to principal axes.
- If we write this for σ_{ij} , it means that we are looking for directions where $\mathbf{D} \parallel \mathbf{E}$, i.e., our principal axes:

$$\epsilon_0 E_i = \sum_{j=1}^3 \sigma_{ij} D_j \doteq \lambda D_i$$

This is a so-called eigenvalue problem, with eigenvalues λ . If we want to solve for the eigenvalues we obtain

$$\det[\sigma_{ij} - \lambda I_{ij}] = 0, \quad \text{with } I_{ij} = \delta_{ij}$$

This leads to a third order equation in λ , hence we expect three solutions (roots) $\lambda^{(\alpha)}$. The corresponding eigenvectors can be computed from

$$\sum_{j=1}^3 \sigma_{ij} D_j^{(\alpha)} = \lambda^{(\alpha)} D_i^{(\alpha)}.$$

The eigenvectors are orthogonal:

$$\sum_{i=1}^3 D_i^{(\beta)} D_i^{(\alpha)} = 0 \text{ for } \lambda^{(\alpha)} \neq \lambda^{(\beta)}$$

Physically, the directions of the principal axes (defined by the eigenvectors) correspond to the symmetry axes of the crystal.

The diagonalized dielectric and inverse dielectric tensors are linked simply by reciprocating the diagonal elements:

$$\varepsilon_{ij} = \varepsilon_i \delta_{ij}, \quad \sigma_{ij} = \sigma_i \delta_{ij} = \frac{1}{\varepsilon_i} \delta_{ij}$$

$$\left(\varepsilon_{ij} \right) = \begin{bmatrix} \varepsilon_1(\omega) & 0 & 0 \\ 0 & \varepsilon_2(\omega) & 0 \\ 0 & 0 & \varepsilon_3(\omega) \end{bmatrix}$$

The above reasoning shows that in general anisotropic media are characterized by **three** independent dielectric functions (in the **principal coordinate system**).

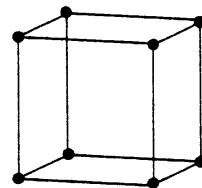
It is easier to do all calculations in the principal coordinate system (coordinate system of the crystal) and **back-transform** the final results to the **laboratory system**.

7.2 Optical classification of crystals

Let us now give a brief overview of crystal classes and their optical properties:

A) isotropic

- three crystallographically equivalent orthogonal axes
- **cubic** crystals (diamond, Si....)

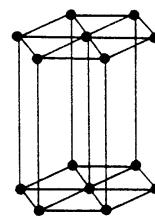
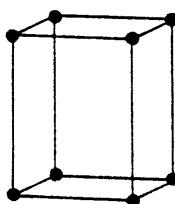
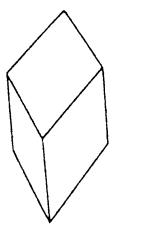


$$\varepsilon_1(\omega) = \varepsilon_2(\omega) = \varepsilon_3(\omega) \rightarrow D_i = \varepsilon_0 \varepsilon(\omega) E_i$$

With no anisotropy, cubic crystals behave like gases, amorphous solids and liquids.

B) uniaxial

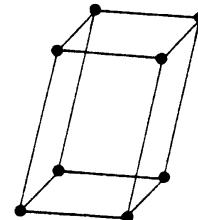
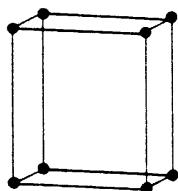
- two crystallographically equivalent directions
- **trigonal** (quartz, lithium niobate), **tetragonal**, **hexagonal**



$$\epsilon_1(\omega) = \epsilon_2(\omega) \neq \epsilon_3(\omega)$$

C) biaxial

- no crystallographically equivalent directions
- orthorhombic, monoclinic, triclinic



$$\epsilon_1(\omega) \neq \epsilon_2(\omega) \neq \epsilon_3(\omega)$$

7.3 Index ellipsoid

The index ellipsoid offers a simple geometrical interpretation of the **inverse** dielectric tensor $\hat{\sigma} = [\hat{\epsilon}]^{-1}$. The defining equation for the index ellipsoid is

$$\sum_{i,j=1}^3 \sigma_{ij} x_i x_j = 1$$

which describes a surface in three dimensional space.

Remark on the physics of the index ellipsoid:

The index ellipsoid defines a surface of constant electric energy density in three-dimensional field space:

$$\sum_{i,j=1}^3 \sigma_{ij} D_i D_j = \epsilon_0 \sum_{i=1}^3 E_i D_i = 2w_{\text{el}}$$

In the principal coordinate system, the defining equation of the index ellipsoid simplifies to

$$\sigma_1 x_1^2 + \sigma_2 x_2^2 + \sigma_3 x_3^2 = 1$$

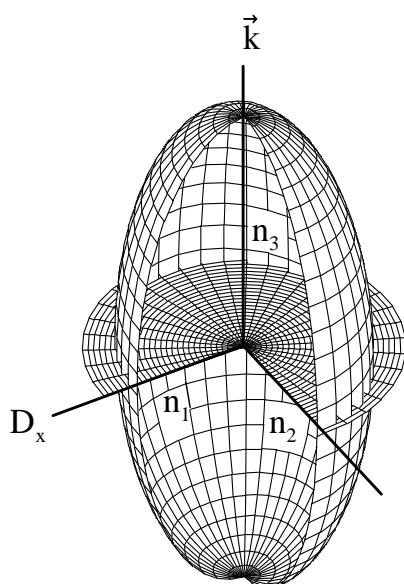
and can be expressed by the three diagonal dielectric tensor components as

$$\frac{x_1^2}{\epsilon_1} + \frac{x_2^2}{\epsilon_2} + \frac{x_3^2}{\epsilon_3} = 1$$

This equation can be interpreted as the defining equation of an ellipsoid having semi-principal axes of length $\sqrt{\epsilon_i}$. From our discussion of the normal modes in isotropic media, we know that $\sqrt{\epsilon}$ corresponded to the refractive index of the normal modes. We will show in the following discussion that for anisotropic media, there are also special cases where the $\sqrt{\epsilon_i}$ determine the phase velocity of normal modes. Hence the elements of the dielectric tensor which define the semi-principal axes of the epsilon ellipsoid can be related to refractive indexes

$$n_i = \sqrt{\epsilon_i}$$

This is the reason why the ellipsoid, which represents graphically the epsilon tensor, is called index ellipsoid.



Graphical representation of the epsilon tensor of an anisotropic crystal by the so-called index ellipsoid.

- The index ellipsoid is degenerate for special cases:
 - isotropic crystal: sphere
 - uniaxial crystal: rotational symmetric with respect to z-axis and $n_1 = n_2$

7.4 Normal modes in anisotropic media

Let us now look for the normal modes in crystals. A normal mode is:

- a solution to the wave equation, which shows only phase dynamics during propagation while amplitude and polarization remain constant
→ simplest solution $\sim \exp\{i[\mathbf{k}(\omega)\mathbf{r} - \omega t]\}$
- a solution where the spatial and temporal evolution of the phase are connected by a dispersion relation $\omega = \omega(\mathbf{k})$ or $\mathbf{k} = \mathbf{k}(\omega)$

Before – isotropic media

In isotropic media the normal modes are monochromatic plane waves

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E} \exp\left\{i[\mathbf{k}(\omega)\mathbf{r} - \omega t]\right\}$$

with the dispersion relation

$$\mathbf{k}^2(\omega) = k^2(\omega) = \frac{\omega^2}{c^2} \epsilon(\omega)$$

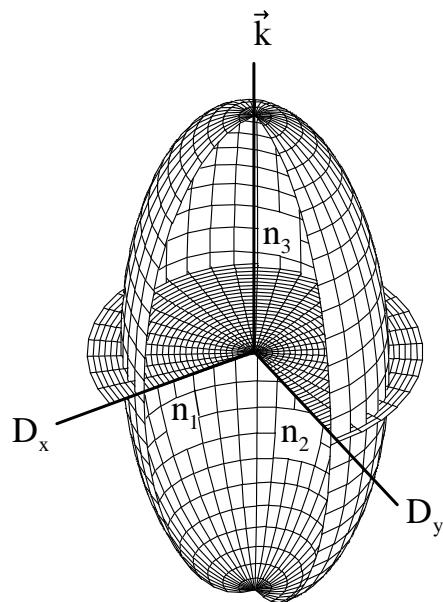
where $\epsilon(\omega) > 0$ and is real. In addition, the transversality condition $\mathbf{k} \cdot \mathbf{E} = \mathbf{k} \cdot \mathbf{D} = 0$ applies. The normal modes are elliptically polarized and the polarization is conserved during propagation.

Now – anisotropic media

What are the normal modes in anisotropic media?

7.4.1 Normal modes propagating in principal directions

Let us first calculate the normal modes for propagation in the direction of the principal axes of the index ellipsoid, which is the simplest case.



We assume without loss of generality that the orthogonal principal axes are in x, y, z direction and the light propagates in z direction ($\mathbf{k} \rightarrow k_z$). Then, the fields are arbitrary in the x, y -plane

$$D_x, D_y \neq 0$$

and due to the diagonal form of the dielectric tensor

$$D_i = \epsilon_0 \epsilon_i E_i$$

In general, we have $\mathbf{E} \neq \mathbf{D}$. However, from the above dielectric tensor equation in principal coordinates, a vanishing component of \mathbf{D} along the z -direction implies a vanishing z -component of \mathbf{E} . Thus, in this particular case of propagation along one of

the principal axes, $\mathbf{k} \cdot \mathbf{D} = 0 \Rightarrow \mathbf{k} \cdot \mathbf{E} = 0$, even if \mathbf{D} and \mathbf{E} are not parallel to each other. The two polarization directions are then decoupled:

$$D_x, \varepsilon_x \sim D_x \exp[\mathbf{i}(k_1 z - \omega t)] \text{ with } k_1^2 = \frac{\omega^2}{c^2} \varepsilon_x(\omega)$$

$$D_y, \varepsilon_y \sim D_y \exp[\mathbf{i}(k_2 z - \omega t)] \text{ with } k_2^2 = \frac{\omega^2}{c^2} \varepsilon_y(\omega)$$

We see that in contrast to isotropic media, normal modes can't be elliptically polarized, since the different polarizations along the principal axes x and y undergo phase evolution at different rates. Thus, polarization direction would change during propagation. But for linear polarization in the direction of a **principal axis** (x or y), only the phase changes during propagation. Thus, we have found our normal modes:

$$\mathbf{D}^{(a)} = \left\{ D_x \exp[\mathbf{i}(\mathbf{k}_a \mathbf{r} - \omega t)] \right\} \mathbf{e}_x \rightarrow \mathbf{k}_a^2 = \frac{\omega^2}{c^2} \varepsilon_x \rightarrow \text{normal mode a}$$

$$\mathbf{D}^{(b)} = \left\{ D_y \exp[\mathbf{i}(\mathbf{k}_b \mathbf{r} - \omega t)] \right\} \mathbf{e}_y \rightarrow \mathbf{k}_b^2 = \frac{\omega^2}{c^2} \varepsilon_y \rightarrow \text{normal mode b}$$

→ For light propagation in a principal direction, we find two perpendicular linearly polarized normal modes with $\mathbf{E} \parallel \mathbf{D}$.

Remark on the indices in the index ellipsoid

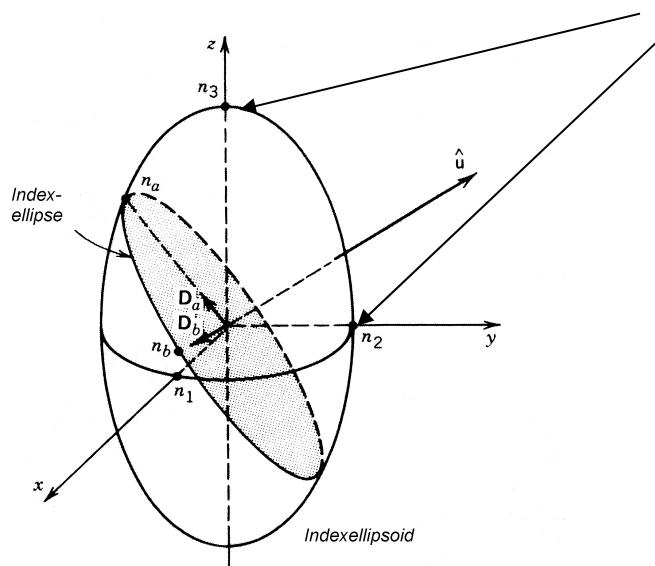
The indices $n_i = \sqrt{\varepsilon_i}$ in the index ellipsoid are connected to the indices n_a and n_b of the normal modes propagating along the principal axis. However please be careful about the direction correspondence. For example, the two normal modes propagating along the z direction have phase velocities determined by the indices $n_a = n_x$ and $n_b = n_y$ associated with the directions of their electric fields rather than the direction of their propagation.

7.4.2 Normal modes for arbitrary propagation direction

7.4.2.1 Geometrical construction

Before we mathematically derive the normal modes and their dispersion relations, let us preview the results visualized in the index ellipsoid. It is actually possible to construct the normal modes geometrically. We start from the normal modes which we have determined for propagation along the principal directions of the crystal and try to generalize to arbitrary propagation directions:

- For a specific crystal and a given frequency ω , we take the ε_i in the principal axis system and construct the index ellipsoid.
- We then fix the propagation direction of the normal mode of interest by the unit vector $\mathbf{k} / k = \mathbf{u}$.
- We draw a plane through the origin of index ellipsoid which is perpendicular to \mathbf{u} .



- The resulting intersection between the constructed plane and index ellipsoid is an ellipse, the so-called **index ellipse**.
- The semi-major and semi-minor axes of this index ellipse correspond to the refractive indices n_a, n_b of the normal modes for the propagation direction $\mathbf{u} = \mathbf{k} / k$

$$\rightarrow k_a = \frac{\omega}{c} n_a \text{ and } k_b = \frac{\omega}{c} n_b$$

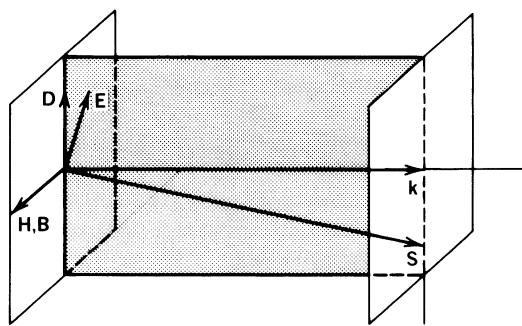
- The directions of the major and minor axes of the index ellipse are the polarization directions of the normal modes $\mathbf{D}^{(a)}$ and $\mathbf{D}^{(b)}$.
- The electric field vectors of the normal modes $\mathbf{E}^{(a)}$ and $\mathbf{E}^{(b)}$ follow from

$$E_i^{(a)} = \frac{D_i^{(a)}}{\epsilon_0 \epsilon_i}, \quad E_i^{(b)} = \frac{D_i^{(b)}}{\epsilon_0 \epsilon_i}$$

- Thus, $\mathbf{D}^{(a,b)}$ and $\mathbf{E}^{(a,b)}$, and $\mathbf{E}^{(a,b)}$ are not perpendicular to \mathbf{k} , unlike the case of isotropic media.
- This has a direct consequence on the Poynting vector:

$$\langle \mathbf{S} \rangle = \frac{1}{2} \Re(\mathbf{E} \times \mathbf{H}^*)$$

hence \mathbf{k} is not parallel to $\langle \mathbf{S} \rangle$ because $\langle \mathbf{S} \rangle \perp \mathbf{E}$; i.e. the direction of phase evolution and energy transport need not coincide for general propagation in anisotropic media.



- If the index ellipse is a circle, the direction of this particular \mathbf{k} -vector defines the **optic axis** of the crystal where both normal modes have the same refractive index. Thus, any polarization (\mathbf{D} orthogonal to optic axis) is preserved by propagation along this optic axis.

7.4.2.2 Mathematical derivation of dispersion relation

Let us now derive mathematically the dispersion relation for normal modes of the form

$$\mathbf{E}(\mathbf{r},t) = \mathbf{E} \exp\left\{i[\mathbf{k}(\omega)\mathbf{r} - \omega t]\right\}$$

$$\mathbf{D}(\mathbf{r},t) = \mathbf{D} \exp\left\{i[\mathbf{k}(\omega)\mathbf{r} - \omega t]\right\}$$

with \mathbf{E} and \mathbf{D} being connected by $D_i = \epsilon_0 \epsilon_i E_i$ along principal axes.

In the **isotropic case**, we found the dispersion relation

$$\mathbf{k}^2(\omega) = k^2(\omega) = \frac{\omega^2}{c^2} \epsilon(\omega)$$

where the magnitude of the \mathbf{k} -vector is **independent of its direction**. The fields of the normal modes are elliptically polarized.

In the **anisotropic** case, the normal modes are again monochromatic plane waves, but the wavenumbers depend on the directions \mathbf{u} of propagation, where $\mathbf{u} = \mathbf{k} / k$. Hence

$$k = k(\omega, \mathbf{u})$$

and the polarizations of the normal modes are **not elliptic**.

In the following, we start again from Maxwell's equations and plug in the plane wave ansatz. We will use the following notation for the directional dependence of \mathbf{k} :

$$\mathbf{k} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = k \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \text{ with } u_1^2 + u_2^2 + u_3^2 = 1$$

Our aim is to derive $\omega = \omega(k_1, k_2, k_3)$ or $\omega = \omega(k, \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ or $k = k(\omega, u_1, u_2, u_3)$.

We start from Maxwell's equations for the plane wave Ansatz:

$$\mathbf{k} \cdot \mathbf{D} = 0$$

$$\mathbf{k} \times \mathbf{E} = \omega \mu_0 \mathbf{H}$$

$$\mathbf{k} \cdot \mathbf{H} = 0$$

$$\mathbf{k} \times \mathbf{H} = -\omega \mathbf{D}$$

Now we follow the usual derivation of the wave equation:

$$-\left[\mathbf{k} \times (\mathbf{k} \times \mathbf{E})\right] = \frac{\omega^2}{c^2} \frac{1}{\epsilon_0} \mathbf{D} \rightarrow -\mathbf{k}(\mathbf{k} \cdot \mathbf{E}) + \mathbf{k}^2 \mathbf{E} = \frac{\omega^2}{c^2} \frac{1}{\epsilon_0} \mathbf{D}$$

- Here $\mathbf{k} \cdot \mathbf{E}$ does not vanish as it would have in the isotropic case since the transversality condition from the divergence equation ($\text{div}\mathbf{D}(\mathbf{r}, t) = 0 \rightarrow \mathbf{D} \perp \mathbf{k}$) applies only to the \mathbf{D} field.
- In the principal coordinate system where $D_i = \epsilon_0 \epsilon_i E_i$, we find

$$-k_i \sum_j k_j E_j + k^2 E_i = \frac{\omega^2}{c^2} \epsilon_i E_i$$

$$\left(\frac{\omega^2}{c^2} \epsilon_i - k^2 \right) E_i = -k_i \sum_j k_j E_j$$

Remark: for isotropic media the r.h.s. of this equation would vanish ($\mathbf{k} \cdot \mathbf{E} = 0$).

Now, we have the following problem to solve:

$$\begin{bmatrix} \frac{\omega^2}{c^2} \epsilon_1 - k_1^2 - k_2^2 & k_1 k_2 & k_1 k_3 \\ k_2 k_1 & \frac{\omega^2}{c^2} \epsilon_2 - k_1^2 - k_3^2 & k_2 k_3 \\ k_3 k_1 & k_3 k_2 & \frac{\omega^2}{c^2} \epsilon_3 - k_1^2 - k_2^2 \end{bmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The general way to solve this problem is by setting the determinant $\det[\dots] = 0$, which gives the dispersion relation $\omega = \omega(k)$ for given k_i / k . But this results in a complicated mathematical problem from which it would be difficult to derive intuition on the properties of the normal modes.

However, there is an easy way to show some general properties of the dispersion relation. We start from the following trick:

$$\left(\frac{\omega^2}{c^2} \epsilon_i - k^2 \right) E_i = -k_i \sum_j k_j E_j \rightarrow E_i = -\frac{k_i}{\left(\frac{\omega^2}{c^2} \epsilon_i - k^2 \right)} \sum_j k_j E_j$$

Now we multiply this equation by k_i , perform a summation over the index ' i ' and relabel $i \leftrightarrow j$ on the l.h.s.:

$$\sum_j k_j E_j = -\sum_i \frac{k_i^2}{\left(\frac{\omega^2}{c^2} \epsilon_i - k^2 \right)} \sum_j k_j E_j.$$

Because $\text{div}\mathbf{E} = \sum_j k_j E_j \neq 0$ we can safely divide away this factor and obtain the (implicit) dispersion relation:

$$\sum_i \frac{k_i^2}{\left(k^2 - \frac{\omega^2}{c^2} \epsilon_i \right)} = 1$$

With $\begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = k(\omega) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \frac{\omega}{c} n(\omega) \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ we can write

$$\sum_i \frac{k^2 u_i^2}{\left(k^2 - \frac{\omega^2}{c^2} \epsilon_i \right)} = 1 \rightarrow \sum_i \frac{u_i^2}{\left(1 - \frac{\epsilon_i}{n^2} \right)} = 1$$

$$\rightarrow \boxed{\sum_i \frac{u_i^2}{\left[n^2 - \epsilon_i \right]} = \frac{1}{n^2}} \text{ final form of dispersion relation}$$

Discussion of results

For given $\epsilon_i(\omega)$ and direction (u_1, u_2) we can compute the refractive index $n(\omega, u_1, u_2)$ experienced by the normal mode. Because $u_1^2 + u_2^2 + u_3^2 = 1$, it is sufficient to fix two components (u_1, u_2) of \mathbf{u} to determine the direction.

A more explicit form of the dispersion relation can be obtained by multiplying by the lowest common denominator:

$$u_1^2 (n^2 - \epsilon_2) (n^2 - \epsilon_3) n^2 + u_2^2 (n^2 - \epsilon_1) (n^2 - \epsilon_3) n^2 + u_3^2 (n^2 - \epsilon_1) (n^2 - \epsilon_2) n^2 = \\ (n^2 - \epsilon_1) (n^2 - \epsilon_2) (n^2 - \epsilon_3)$$

The resulting equation is quadratic in $(n^2)^2$ since the n^6 -terms cancel. As expected, we obtain two (positive) solutions n_a, n_b and therefore $k_a = n_a (\omega / c)$ and $k_b = n_b (\omega / c)$ for the two orthogonally polarized normal modes $\mathbf{D}^{(a)}$ and $\mathbf{D}^{(b)}$.

In particular, for propagation in the direction of a principal axis ($u_3 = 1$ and $u_1 = u_2 = 0$, see 6.4.1) we find:

$$\rightarrow (n^2 - \epsilon_1) (n^2 - \epsilon_2) n^2 = (n^2 - \epsilon_1) (n^2 - \epsilon_2) (n^2 - \epsilon_3) \\ \curvearrowright (n^2 - \epsilon_1) (n^2 - \epsilon_2) \epsilon_3 = 0 \\ \curvearrowright \boxed{n_a^2 = \epsilon_1, \quad n_b^2 = \epsilon_2}$$

Finally, we can derive some properties of the fields of the normal modes, i.e. the eigenfunctions.

We start from the above eigenvalue equation which still involves the eigenfunctions

$$\left(\frac{\omega^2}{c^2} \epsilon_i - k^2 \right) E_i = -k_i \sum_j k_j E_j.$$

For cases where the first factor of the l.h.s. is not equal to zero (propagation directions not parallel to the principal axes) we can divide by this term

$$E_i = -\frac{k_i}{\left(\frac{\omega^2}{c^2}\epsilon_i - k^2\right)} \sum_j k_j E_j.$$

The sum does not depend on the dummy index i . Hence the last term of the equation must be constant

$$\rightarrow \sum_j k_j E_j = \text{const.}$$

Knowing that the last term is a constant, we can derive the ratio of the individual field components from the first term of the equation

$$\curvearrowright E_1 : E_2 : E_3 = \frac{k_1}{\frac{\omega^2}{c^2}\epsilon_1 - k^2} : \frac{k_2}{\frac{\omega^2}{c^2}\epsilon_2 - k^2} : \frac{k_3}{\frac{\omega^2}{c^2}\epsilon_3 - k^2}$$

and with $D_i = \epsilon_0 \epsilon_i E_i$

$$\curvearrowright D_1 : D_2 : D_3 = \frac{\epsilon_1 k_1}{\frac{\omega^2}{c^2}\epsilon_1 - k^2} : \frac{\epsilon_2 k_2}{\frac{\omega^2}{c^2}\epsilon_2 - k^2} : \frac{\epsilon_3 k_3}{\frac{\omega^2}{c^2}\epsilon_3 - k^2}$$

Please be aware that this relation can only be applied for propagation directions not parallel to the principal axes.

How are the normal modes polarized?

- The ratio between the field components is real
 \rightarrow phase difference 0 \rightarrow linear polarization

How do we see the orthogonality $\mathbf{D}^{(a)} \cdot \mathbf{D}^{(b)} = 0$? (be careful: $\mathbf{E}^{(a)} \cdot \mathbf{E}^{(b)} \neq 0$) Using the above equation for the components of \mathbf{D} :

$$\begin{aligned} \mathbf{D}^{(a)} \cdot \mathbf{D}^{(b)} &\sim \sum_i \frac{k_a k_b \epsilon_i^2 u_i^2}{\left(k_a^2 - \frac{\omega^2}{c^2} \epsilon_i\right) \left(k_b^2 - \frac{\omega^2}{c^2} \epsilon_i\right)} \\ &= \frac{c^2}{\omega^2} \frac{k_a k_b}{(k_b^2 - k_a^2)} \left[k_a^2 \sum_i \frac{\epsilon_i u_i^2}{\left(k_a^2 - \frac{\omega^2}{c^2} \epsilon_i\right)} - k_b^2 \sum_i \frac{\epsilon_i u_i^2}{\left(k_b^2 - \frac{\omega^2}{c^2} \epsilon_i\right)} \right] \end{aligned}$$

Since the two red terms vanish due to the dispersion relation, it follows that $\mathbf{D}^{(a)} \cdot \mathbf{D}^{(b)} = 0$. The vanishing of the red terms can be seen when rewriting the dispersion relation:

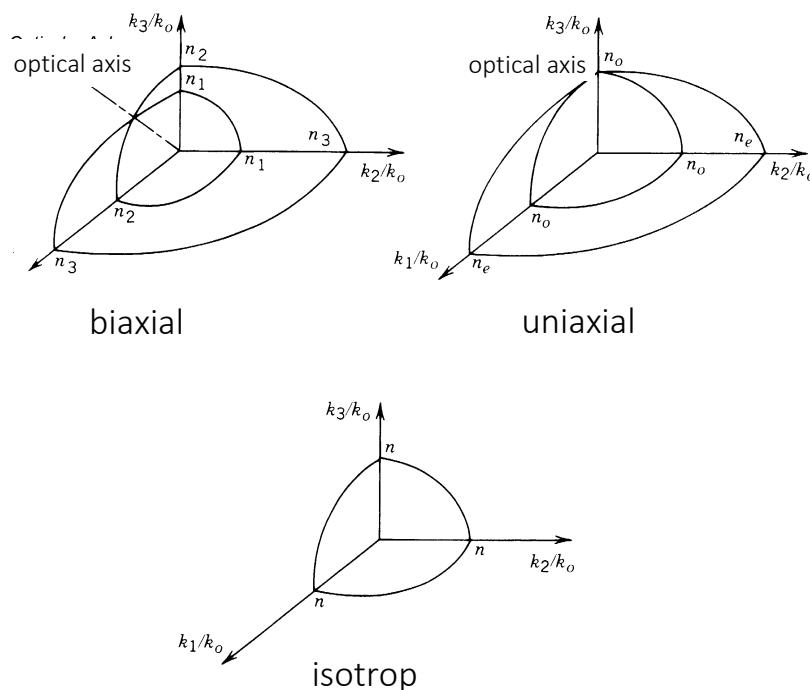
$$1 = \sum_i \frac{k_{a,b}^2 u_i^2}{\left(k_{a,b}^2 - \frac{\omega^2}{c^2} \epsilon_i\right)} = \sum_i \frac{\left(k_{a,b}^2 - \frac{\omega^2}{c^2} \epsilon_i + \frac{\omega^2}{c^2} \epsilon_i\right) u_i^2}{\left(k_{a,b}^2 - \frac{\omega^2}{c^2} \epsilon_i\right)} = 1 + \frac{\omega^2}{c^2} \sum_i \frac{\epsilon_i u_i^2}{\left(k_{a,b}^2 - \frac{\omega^2}{c^2} \epsilon_i\right)}$$

using the definition of \mathbf{u} as a unit vector.

7.4.3 Normal surfaces of normal modes

In addition to the index ellipsoid, which is a graphical representation of the material properties of crystals from which the properties of the normal modes can be interpreted as shown above, we can derive a direct graphical representation of the dispersion relation of normal modes in crystals. This graphical representation of the dispersion relation is called **normal surfaces**:

If we plot the refractive indices (wave number or norm of the k-vector divided by k_0) of the normal modes in the k_i -space (normal surfaces), we obtain a centrosymmetric, two-layered surface.



Normal surfaces as the graphical representation of the dispersion relation of normal modes in crystals.

isotropic: sphere

uniaxial: 2 points with $n_a = n_b$ at the poles → connecting line defines the optical axis
(for $\epsilon_1 = \epsilon_2 = \epsilon_{or}$, $\epsilon_3 = \epsilon_e$ the z-axis is the optical axis)

biaxial: 4 points with $n_a = n_b$ → connecting lines define two optical axes

How to read the figure:

- fix propagation direction (u_1, u_2) → intersection with surfaces
- distances from origin to intersections with surfaces correspond to refractive indices of normal modes
- definition of optical axis → $n_a = n_b$

Summary: there are two geometrical constructions:

- index ellipsoid (visualization of dielectric tensor)

- fix propagation direction \curvearrowright index ellipse \curvearrowright semi-major/minor axes give n_a, n_b (refractive indices of the normal modes)
 - optical axis \rightarrow index ellipse is a circle
 - for uniaxial crystals the optical axis coincides with one principal axis
- B) normal surfaces (visualization of dispersion relation)
- fix propagation direction \curvearrowright intersection with surfaces \curvearrowright distances from origin give n_a, n_b
 - optical axis connects points with $n_a = n_b$

Conclusion

In anisotropic media and for a given propagation direction we find two normal modes, which are linearly polarized monochromatic plane waves with two different phase velocities $c/n_a, c/n_b$ and two orthogonal polarization directions $\mathbf{D}^{(a)}, \mathbf{D}^{(b)}$.

7.4.4 Special case: uniaxial crystals

Let us now investigate the special and simpler case of uniaxial crystals. In biaxial crystals, we do not find any other effects, but the description is more complicated.

The main advantage of uniaxial crystals is that we have rotational symmetry in one plane. Therefore, all three-dimensional graphs (index-ellipsoid, normal surfaces) can be reduced to two dimensions and we can sketch them more easily. As we have seen before, uniaxial crystals have trigonal, tetragonal, or hexagonal symmetry.

The following classification for uniaxial crystals is commonly used

$\epsilon_{or} > \epsilon_e \rightarrow$ negative uniaxial

$\epsilon_{or} < \epsilon_e \rightarrow$ positive uniaxial

Let us assume (without loss of generality) that the index ellipsoid is rotationally symmetric around the z-axis, and we have

$$\epsilon_1 = \epsilon_2 = \epsilon_{or}, \quad \epsilon_3 = \epsilon_e,$$

which are the ordinary and extraordinary dielectric constants.

Then, we expect two normal modes:

A) ordinary wave

- n_a independent of propagation direction
- The ordinary wave $\mathbf{D}^{(or)}$ is polarized perpendicular to the z-axis and the k-vector and it does not interact with ϵ_e .

B) extraordinary wave

- n_b depends on propagation direction
- The extraordinary wave $\mathbf{D}^{(e)}$ is polarized perpendicular to the k-vector and $\mathbf{D}^{(or)}$.

The z-axis is, according to definition, the optical axis with $n_a = n_b$.

Let us now derive the dispersion relation: From above we know the implicit form

$$\sum_i \frac{u_i^2}{n^2 - \epsilon_i} = \frac{1}{n^2}$$

For uniaxial crystals this leads to

$$\frac{u_1^2}{n^2 - \epsilon_{\text{or}}} + \frac{u_2^2}{n^2 - \epsilon_{\text{or}}} + \frac{u_3^2}{n^2 - \epsilon_e} = \frac{1}{n^2},$$

which can be expanded to

$$n^2 [n^2 - \epsilon_e] [n^2 - \epsilon_{\text{or}}] (u_1^2 + u_2^2) + n^2 [n^2 - \epsilon_{\text{or}}]^2 u_3^2 = [n^2 - \epsilon_e] [n^2 - \epsilon_{\text{or}}]^2.$$

A) ordinary wave

independent of direction

$$n_a^2 = \epsilon_{\text{or}} \rightarrow k_a^2 = \frac{\omega^2}{c^2} n_a^2 = k_0^2 \epsilon_{\text{or}}$$

B) extraordinary wave

dependent on direction

$$\frac{(u_1^2 + u_2^2)}{\epsilon_e} + \frac{u_3^2}{\epsilon_{\text{or}}} = \frac{1}{n_b^2}, \quad k_b^2 = \frac{\omega^2}{c^2} n_b^2 (u_1, u_2, u_3)$$

Hence for a given direction \mathbf{u} there exist two refractive indexes n_a, n_b .

The geometrical interpretation as normal surfaces is straightforward and can be done w.l.o.g. in the k_2, k_3 or y, z plane ($u_1 = 0$).

The shape of the normal surfaces can be derived in the following way.

We have with

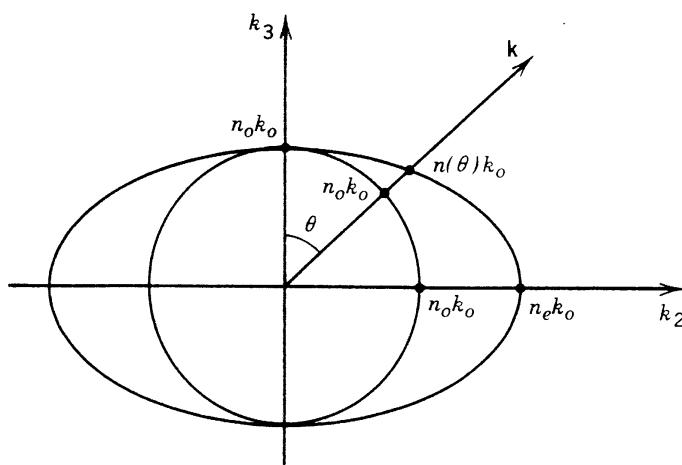
$$k_i^2 = k_0^2 n^2 u_i^2$$

A) ordinary wave

$$k_a^2 = k_1^2 + k_2^2 + k_3^2 = k_0^2 \epsilon_{\text{or}}$$

B) extraordinary wave

$$\frac{1}{\epsilon_e} \frac{(k_1^2 + k_2^2)}{k_0^2} + \frac{1}{\epsilon_{\text{or}}} \frac{k_3^2}{k_0^2} = 1$$



Normal surfaces for a uniaxial crystal.

If we introduce an angle θ , as in the figures above, to describe the propagation direction, a simple computation of $n_b^2(\theta)$ for the extraordinary wave is possible:

$$u_2 = \sin \theta, \quad u_3 = \cos \theta$$

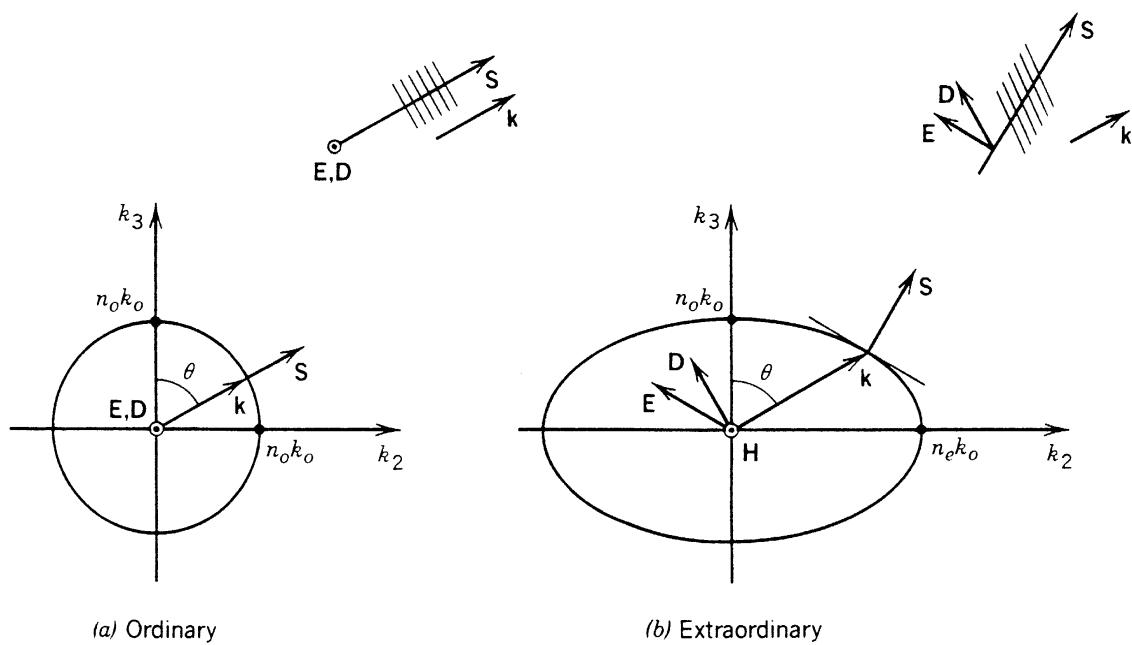
$$n_b^2(\theta) = \frac{\varepsilon_e \varepsilon_{or}}{\varepsilon_{or} \sin^2 \theta + \varepsilon_e \cos^2 \theta}$$

Determination of field components

We know from before that

$$D_1 : D_2 : D_3 = \frac{\varepsilon_{or} k_1}{\frac{\omega^2}{c^2} \varepsilon_{or} - k^2} : \frac{\varepsilon_{or} k_2}{\frac{\omega^2}{c^2} \varepsilon_{or} - k^2} : \frac{\varepsilon_e k_3}{\frac{\omega^2}{c^2} \varepsilon_e - k^2}.$$

For the extraordinary wave, all denominators are finite, and in particular $k_1 = 0$ implies $D_1^{(e)} = 0$, hence $\mathbf{D}^{(e)}$ is polarized in the y-z plane. Thus, $\mathbf{D}^{(or)} \perp \mathbf{D}^{(e)}$ implies that $\mathbf{D}^{(or)}$ is polarized in x-direction.



In summary, we find for the polarizations of the fields:

A) ordinary: \mathbf{D} perpendicular to optical axis and \mathbf{k} ,

$$\mathbf{D} \perp \mathbf{k}, \mathbf{D} \parallel \mathbf{E}$$

B) extraordinary: \mathbf{D} perpendicular to \mathbf{k} and in the plane spanned by \mathbf{k} and the optic axis

$$\mathbf{D} \perp \mathbf{k}, \mathbf{D} \nparallel \mathbf{E}, \text{ because } D_2 = \epsilon_0 \epsilon_{\text{or}} E_2, D_3 = \epsilon_0 \epsilon_{\text{e}} E_3$$

8. Optical fields in isotropic, dispersive and piecewise homogeneous media

8.1 Basics

8.1.1 Definition of the problem

Up to now, we have always treated homogeneous media. However, in the context of evanescent waves, we have already used the concept of an interface. This was a first step in the direction which we wish to pursue now. While we have treated interfaces thus far, we have never considered the effects of the interface. Instead, we have simply fixed the incident field on an interface and described its subsequent propagation in the adjacent half-space.

In this chapter, we will go further and consider the reflection and transmission properties of the following physical systems:

- interface
- layer (2 successive interfaces)
- system of layers (arbitrary number of successive interfaces)

To do this in a simplified way, we will turn back to isotropic media. However, the predominant effects which we discuss in this chapter will also be valid for anisotropic media.

Aims

- We will study the interaction of monochromatic plane waves with arbitrary multilayered systems → interferometers, dielectric mirrors, ...
- By superposition of such plane waves, we can then describe the interaction of spatio-temporally varying fields with multilayered systems.
- We will see a new effect: the “trapping” of light in systems of layers → new types of normal modes in inhomogeneous space → “guided” waves (propagation of confined light beams without diffraction).

Approach

- take Maxwell's transition condition for interfaces
- calculate field in inhomogeneous media → matrix method
- solve reflection-transmission problem for interface, layer, and system of layers,
- apply the method to consider special cases like Fabry-Perot-interferometer, 1D photonic crystals, waveguide...

Background

- Due to the orthogonality of the normal modes of homogeneous space, there is no interaction between normal modes in homogeneous space.

- The inhomogeneity breaks this orthogonality and modes interact, i.e. exchange energy.
- However, locally the concept of eigenmodes is still very useful. We will see that for the type of inhomogeneity considered (planar surfaces), the interaction at the inhomogeneities is limited to a small number of modes.

8.1.2 Decoupling of the vectorial wave equation

Before we start treating a single interface, it is worth looking again at the wave equation in homogeneous space in frequency domain

$$\text{rotrot } \bar{\mathbf{E}}(\mathbf{r}, \omega) - \frac{\omega^2}{c^2} \bar{\mathbf{E}}(\mathbf{r}, \omega) = \mathbf{i}\omega\mu_0 \bar{\mathbf{j}}(\mathbf{r}, \omega) + \mu_0\omega^2 \bar{\mathbf{P}}(\mathbf{r}, \omega).$$

In general, for isotropic media, all field components are coupled due to the **rotrot** operator. However, for problems with **translational invariance** in at least one direction (homogeneous infinite media, layers or interfaces), a simplification is possible. Without loss of generality, let us assume the translational invariance of the system in the **y direction** and propagation in the x-z-plane $\rightarrow \partial/\partial y = 0$

$$\text{rot rot } \bar{\mathbf{E}} = \mathbf{grad} \cdot \mathbf{div} \bar{\mathbf{E}} - \Delta \bar{\mathbf{E}} = \begin{bmatrix} \frac{\partial}{\partial x} \left(\frac{\partial \bar{E}_x}{\partial x} + \frac{\partial \bar{E}_z}{\partial z} \right) \\ 0 \\ \frac{\partial}{\partial z} \left(\frac{\partial \bar{E}_x}{\partial x} + \frac{\partial \bar{E}_z}{\partial z} \right) \end{bmatrix} - \begin{bmatrix} \Delta^{(2)} \bar{E}_x \\ \Delta^{(2)} \bar{E}_y \\ \Delta^{(2)} \bar{E}_z \end{bmatrix}.$$

We can then decompose the electric field as $\bar{\mathbf{E}} = \bar{\mathbf{E}}_{\perp} + \bar{\mathbf{E}}_{\parallel}$, with

$$\bar{\mathbf{E}}_{\perp} = \begin{pmatrix} 0 \\ \bar{E}_y \\ 0 \end{pmatrix}, \quad \bar{\mathbf{E}}_{\parallel} = \begin{pmatrix} \bar{E}_x \\ 0 \\ \bar{E}_z \end{pmatrix}, \nabla^{(2)} = \begin{pmatrix} \partial/\partial x \\ 0 \\ \partial/\partial z \end{pmatrix}, \Delta^{(2)} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}.$$

$\bar{\mathbf{E}}_{\perp}$ is polarized **perpendicular** to the plane of propagation, $\bar{\mathbf{E}}_{\parallel}$ is polarized **parallel** to this plane.

Common notations are:

perpendicular: \perp	$\rightarrow s$	\rightarrow TE (transverse electric)
parallel: \parallel	$\rightarrow p$	\rightarrow TM (transverse magnetic)

The two components are now decoupled and thus can be treated independently:

$$\Delta^{(2)} \bar{\mathbf{E}}_{\perp}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \bar{\mathbf{E}}_{\perp}(\mathbf{r}, \omega) = -\mathbf{i}\omega\mu_0 \bar{\mathbf{j}}_{\perp}(\mathbf{r}, \omega) - \mu_0\omega^2 \bar{\mathbf{P}}_{\perp}(\mathbf{r}, \omega)$$

$$\Delta^{(2)} \bar{\mathbf{E}}_{\parallel}(\mathbf{r}, \omega) + \frac{\omega^2}{c^2} \bar{\mathbf{E}}_{\parallel}(\mathbf{r}, \omega) - \mathbf{grad}^{(2)} \cdot \mathbf{div}^{(2)} \bar{\mathbf{E}}_{\parallel} = -\mathbf{i}\omega\mu_0 \bar{\mathbf{j}}_{\parallel}(\mathbf{r}, \omega) - \mu_0\omega^2 \bar{\mathbf{P}}_{\parallel}(\mathbf{r}, \omega)$$

From

$$\bar{\mathbf{H}}(\mathbf{r}, \omega) = -\frac{\mathbf{i}}{\omega \mu_0} \operatorname{rot} \bar{\mathbf{E}}(\mathbf{r}, \omega)$$

we can conclude that the corresponding magnetic fields are

$$\mathbf{E}_{\text{TE}} = \begin{pmatrix} 0 \\ E_y \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix}, \quad \mathbf{H}_{\text{TE}} = \begin{pmatrix} H_x \\ 0 \\ H_z \end{pmatrix},$$

$$\mathbf{E}_{\text{TM}} = \begin{pmatrix} E_x \\ 0 \\ E_z \end{pmatrix}, \quad \mathbf{H}_{\text{TM}} = \begin{pmatrix} 0 \\ H_y \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ H \\ 0 \end{pmatrix}.$$

8.1.3 Interfaces and symmetries

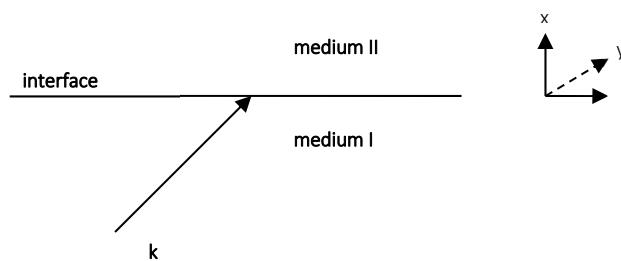
Up to now, we have treated plane waves of the form

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E} \exp[\mathbf{i}(\mathbf{k}\mathbf{r} - \omega t)]$$

- homogeneous space implies: $\exp(\mathbf{i}\mathbf{k}\mathbf{r})$
- monochromaticity leads to: $\exp(-\mathbf{i}\omega t)$

Now we break the homogeneity in x-direction by considering an interface in the y-z-plane, which is infinite in the y- and z-directions.

The symmetry of the inhomogeneity is important in determining the symmetry of the solutions considered, here the normal mode's translation symmetry in z and y.



W.l.o.g. we can assume

$$\mathbf{k} = (k_x, 0, k_z)$$

by choosing an appropriate coordinate system. Here, the x, z -plane is the plane of incidence.. Thus, the problem does not depend on the y -coordinate.

8.1.4 Transition conditions

From Maxwell's equations follow the transition conditions for the field components. Here we will use that the transverse components $\bar{\mathbf{E}}_{\perp}$ and $\bar{\mathbf{H}}_{\perp}$ are continuous at an

interface between two media since they are tangential to the interface. This implies for the:

A) Continuity of fields

- TE: $E = E_y$ and H_z continuous
- TM: E_z and $H = H_y$ continuous
- As pointed out before, we can always split any field into its TE and TM polarization components $\mathbf{E} = \mathbf{E}_{\text{TE}} + \mathbf{E}_{\text{TM}}$ and treat them separately.

B) Continuity of wave vectors

- Since we have homogeneity in the z -direction, we expect the solution on each side of the interface to vary along the z -direction as $\sim \exp(\mathbf{i}k_z z)$. Since k_z determines the rate at which the field varies along the boundary, the continuity of the transverse field components implies the continuity of the wave vector component k_z across the interface.
homogeneous in z -direction \rightarrow phase $e^{ik_z z} \rightarrow k_z$ continuous

Therefore, the electric field on both sides of the planar interface can be written as:

$$\mathbf{E}(x, z, t) = \mathbf{E}_{\text{TE}}(\mathbf{x}) \exp[\mathbf{i}(k_z z - \omega t)] + \mathbf{E}_{\text{TM}}(\mathbf{x}) \exp[\mathbf{i}(k_z z - \omega t)]$$

8.2 Fields in a layer system \rightarrow matrix method

We will now derive a powerful method to compute the electromagnetic fields in a system of layers with different dielectric properties. Essentially, this involves calculating $\mathbf{E}_{\text{TE}}(x)$ and $\mathbf{E}_{\text{TM}}(x)$ in the above equations.

8.2.1 Fields in one homogeneous layer

Let us first compute the fields in one homogeneous layer of thickness d and with dielectric function $\epsilon_f(\omega)$

- Aim: for given fields at $x=0 \rightarrow$ calculate fields at $x=d$
- Strategy:
 - Perform computation with transverse field components because they are continuous across the interfaces between different layers.
 - The normal components can be calculated later.

Here we assume monochromatic light (one Fourier component, $E(x, z; \omega)$ $H(x, z; \omega)$). In the following discussion, we will often omit the argument/parameter ω in our notation. This can then be used to describe the propagation of any pulsed beam through an arbitrary layered system by first decomposing it into its spatial and temporal frequency components. We can then apply the formalism developed below to describe the transmission/reflection of each component and superimpose them to form the transmitted/reflected pulsed beam.

TE-polarization

We have to solve the y-independent wave equation due to the translational invariance of the fields and ϵ in this direction:

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + \frac{\omega^2}{c^2} \epsilon_f \right] \mathbf{E}_{TE}(x, z) = 0.$$

We use the ansatz from above

$$\mathbf{E}_{TE}(x, z) = \mathbf{E}_{TE}(x) \exp(ik_z z) \text{ and } \mathbf{H}_{TE}(x, z) = \mathbf{H}_{TE}(x) \exp(ik_z z)$$

and substitute it into the wave equation

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\omega^2}{c^2} \epsilon_f - k_z^2 \right] \mathbf{E}_{TE}(x) = 0.$$

The magnetic field can then be calculated by

$$\mathbf{H}_{TE}(x, z) = -\frac{i}{\omega \mu_0} \mathbf{rot} \mathbf{E}_{TE}(x, z).$$

Now let us extract the equations for the transverse field components $E_y = E$, H_z :

$$\left[\frac{\partial^2}{\partial x^2} + k_{fx}^2(k_z, \omega) \right] E(x) = 0 \text{ with } k_{fx}^2(k_z, \omega) = \frac{\omega^2}{c^2} \epsilon_f(\omega) - k_z^2$$

$$H_z(x) = -\frac{i}{\omega \mu_0} \frac{\partial}{\partial x} E(x)$$

This makes sense since the wave equation for the y-component of the electric field is a second order differential equation. Hence we need to specify the field and its first derivative as initial condition at $x = 0$ to determine a unique solution.

TM-polarization

Analogous for the transverse components $H_y = H$, E_z :

$$\left[\frac{\partial^2}{\partial x^2} + k_{fx}^2(k_z, \omega) \right] H(x) = 0$$

$$E_z(x) = \frac{i}{\omega \epsilon_0 \epsilon_f} \frac{\partial}{\partial x} H(x)$$

Again, we can describe the entire dynamics in terms of transverse field components.

Now we have the following problem to solve for both TE and TM:

- Calculate fields $E(x), H(x)$ and derivatives $\frac{\partial}{\partial x} E(x), \frac{\partial}{\partial x} H(x)$ at $x = d$ from their given values at $x = 0$
- Calculate also the other field components at $x = d$
- Finally: $H_{\text{TM}} \rightarrow E_{\text{TM}} \rightarrow$ Total electric field: $E = E_{\text{TM}} + E_{\text{TE}}$

Generalized transverse fields F & G

Because the equations for TE and TM have an identical structure, we can treat them simultaneously by defining the following generalized fields

$$E, H \rightarrow F \quad \text{generalized field 1}$$

$$i\omega\mu_0 H_z, -i\omega\varepsilon_0 E_z \rightarrow G \quad \text{generalized field 2}$$

and write down the generalized problem which we have to solve:

$$\left[\frac{\partial^2}{\partial x^2} + k_{\text{fx}}^2(k_z, \omega) \right] F(x) = 0,$$

$$G(x) = \alpha_f \frac{\partial}{\partial x} F(x) \quad \text{with } \alpha_{f\text{TE}} = 1, \alpha_{f\text{TM}} = \frac{1}{\varepsilon_f}.$$

We know the general solution of this system since the problem is similar to a harmonic oscillator equation:

$$F(x) = C_1 \exp(ik_{\text{fx}}x) + C_2 \exp(-ik_{\text{fx}}x)$$

$$G(x) = \alpha_f \frac{\partial}{\partial x} F(x) = i\alpha_f k_{\text{fx}} [C_1 \exp(ik_{\text{fx}}x) - C_2 \exp(-ik_{\text{fx}}x)]$$

We have as initial conditions $F(0), G(0)$:

$$F(0) = C_1 + C_2$$

$$G(0) = i\alpha_f k_{\text{fx}} [C_1 - C_2]$$

from which we can compute the constants of integration C_1, C_2 :

$$C_1 = \frac{1}{2} \left[F(0) - \frac{i}{\alpha_f k_{\text{fx}}} G(0) \right]$$

$$C_2 = \frac{1}{2} \left[F(0) + \frac{i}{\alpha_f k_{\text{fx}}} G(0) \right]$$

By re-expressing the complex exponential functions as a sum of sines and cosines, the final solution of the initial value problem is therefore:

$$F(x) = \cos(k_{\text{fx}}x) F(0) + \frac{1}{\alpha_f k_{\text{fx}}} \sin(k_{\text{fx}}x) G(0)$$

$$G(x) = -\alpha_f k_{\text{fx}} \sin(k_{\text{fx}}x) F(0) + \cos(k_{\text{fx}}x) G(0)$$

By substituting back the physical E and H fields, we have the electromagnetic field in the layer $0 \leq x \leq d$.

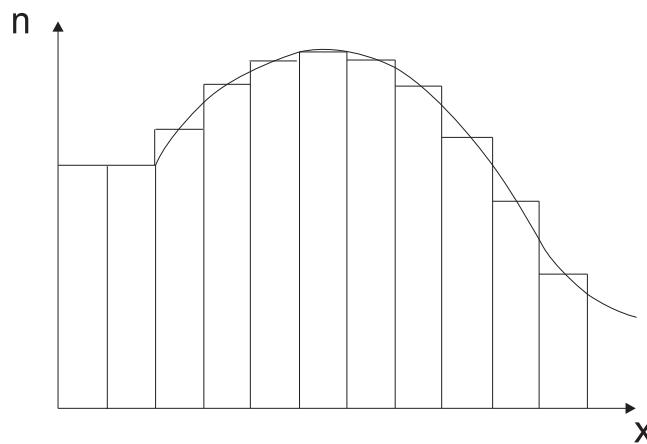
8.2.2 Fields in a system of layers

In the previous subchapter, we saw how to compute the electromagnetic field in a single dielectric layer based on the transverse field components E_y , H_z (TE) and H_y , E_z (TM) at $x = 0$. We can generalize our results to systems of dielectric layers, which are used in many optical devices:

- Bragg mirrors
- chirped mirrors for dispersion compensation
- interferometer
- multilayered waveguides
- Bragg waveguide
- metallic interfaces and layers

Based on the analytic solutions, we can even treat layer systems with layers of very different thicknesses, since the amount of computation is independent of the layer's thickness.

We can even go further and “discretize” an arbitrary inhomogeneous (in one dimension) ϵ distribution. This is important for the modelling of so-called 'GRIN' – Graded-Index Profiles.



A continuously varying ϵ distribution can be described by its discretization into a system of multiple layers each having a homogeneous ϵ .

From above, we know the fields in one layer:

$$F(x) = \cos(k_{fx}x)F(0) + \frac{1}{\alpha_f k_{fx}} \sin(k_{fx}x)G(0)$$

$$G(x) = -\alpha_f k_{fx} \sin(k_{fx}x)F(0) + \cos(k_{fx}x)G(0)$$

Using matrix notation, we can compactly write this linear relation between the incoming and outgoing fields as

$$\begin{Bmatrix} F(x) \\ G(x) \end{Bmatrix} = \hat{\mathbf{m}}(x) \begin{Bmatrix} F(0) \\ G(0) \end{Bmatrix},$$

where the 2x2-matrix $\hat{\mathbf{m}}$ describes the propagation of the fields:

$$\hat{\mathbf{m}}(x) = \begin{pmatrix} \cos(k_{fx}x) & \frac{1}{k_{fx}\alpha_f} \sin(k_{fx}x) \\ -k_{fx}\alpha_f \sin(k_{fx}x) & \cos(k_{fx}x) \end{pmatrix}$$

- To compute the fields at the end of the layer we set $x = d$.
- We assume no absorption in the layer $\rightarrow \|\hat{\mathbf{m}}(x)\| = 1$, i.e. the determinant of $\hat{\mathbf{m}}$ must be equal to 1 such that the overall field amplitudes are unchanged.
- A system of layers is characterized by their respective dielectric functions and thicknesses ϵ_i, d_i .

If multiple layers are considered, the fields within them join together continuously since the field components used for the description of the fields are the continuous tangential components.

Hence, we can directly write the total matrix describing the evolution of the fields through a multilayered system by the matrix product of the individual matrices for the constituent layers:

A) Two layers

$$\begin{Bmatrix} F \\ G \end{Bmatrix}_{d_1+d_2} = \hat{\mathbf{m}}_2(d_2) \begin{Bmatrix} F \\ G \end{Bmatrix}_{d_1} = \hat{\mathbf{m}}_2(d_2) \hat{\mathbf{m}}_1(d_1) \begin{Bmatrix} F \\ G \end{Bmatrix}_0$$

B) N layers

$$\begin{Bmatrix} F \\ G \end{Bmatrix}_{d_1+d_2+..+d_N=D} = \prod_{i=1}^N \hat{\mathbf{m}}_i(d_i) \begin{Bmatrix} F \\ G \end{Bmatrix}_0 = \hat{\mathbf{M}} \begin{Bmatrix} F \\ G \end{Bmatrix}_0 \text{ with } \hat{\mathbf{M}} = \prod_{i=1}^N \hat{\mathbf{m}}_i(d_i)$$

All matrices $\hat{\mathbf{m}}_i$ have the same form, but different $\alpha_f^i, d_i, k_{fx}^i = \sqrt{\frac{\omega^2}{c^2} \epsilon_i(\omega) - k_z^2}$.

Summary of matrix method

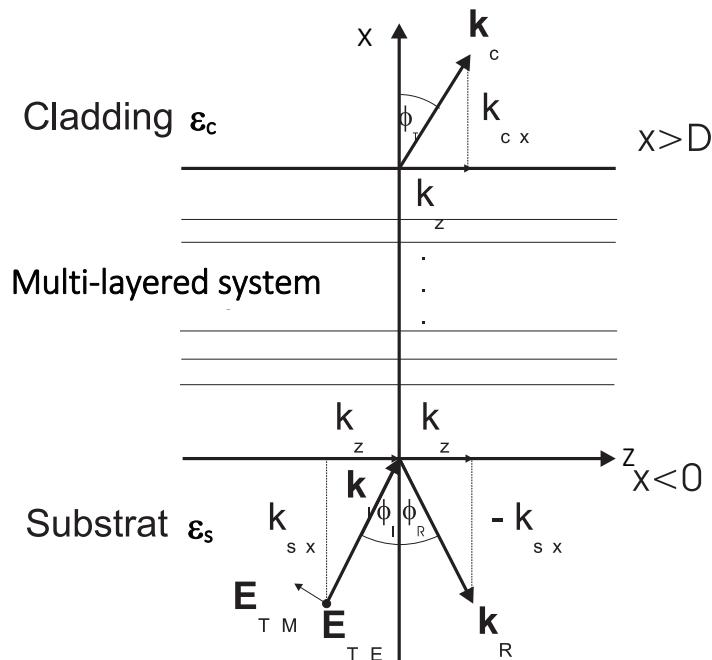
- Incoming fields $F(0)$ and $G(0)$ given (E, H_z for TE, E_z, H for TM)
- $k_z, \alpha_f^i, \epsilon_i, d_i$ given \rightarrow matrix elements
- multiplication of matrices (in the right order) \rightarrow total matrix
- Outgoing fields $F(D)$ and $G(D)$

8.3 Reflection – transmission problem for layer systems

8.3.1 General layer systems

8.3.1.1 Reflection- and transmission coefficients → generalized Fresnel formulas

In the previous chapter, we have learned how to link the electromagnetic field on one side of an arbitrary multilayered system with the field on the other side. We have seen that after splitting into the TE/TM-polarizations, the continuous (transverse) field components are sufficient to describe the whole field. What we will do now is to link those field components with the fields which are accessible in an experimental configuration, i.e. the incident, reflected, and transmitted fields. In particular, we want to solve the reflection-transmission problem, which means that we have to compute the reflected and transmitted fields for a given angle of incidence, frequency, layer system and polarization.



We introduce the wave vectors of the incident (\mathbf{k}_I), reflected (\mathbf{k}_R) and transmitted (\mathbf{k}_T) fields:

$$\mathbf{k}_I = \begin{pmatrix} k_{sx} \\ 0 \\ k_z \end{pmatrix}, \quad \mathbf{k}_R = \begin{pmatrix} -k_{sx} \\ 0 \\ k_z \end{pmatrix}, \quad \mathbf{k}_T = \begin{pmatrix} k_{cx} \\ 0 \\ k_z \end{pmatrix}$$

$$\text{with } k_{sx} = \sqrt{\frac{\omega^2}{c^2}\epsilon_s - k_z^2} = \sqrt{k_s^2(\omega) - k_z^2}, \quad k_{cx} = \sqrt{\frac{\omega^2}{c^2}\epsilon_c - k_z^2} = \sqrt{k_c^2(\omega) - k_z^2},$$

where $\epsilon_s(\omega)$ and $\epsilon_c(\omega)$ are the dielectric functions of the substrate and cladding and k_z is the tangential component of the wave vector, which is continuous throughout the layer system.

As we have seen before, the k_z component of the wave vector is conserved and $\pm k_x$ determines the propagation direction of the wave (forward or backward). The total

length of the wave vector in each layer is given by the dispersion relation for dispersive, isotropic, homogeneous media. As a consequence, the k_x component must change its value in each layer.

Remark on law of reflection and transmission (Snell's law)

It is possible to derive Snell's law just from the fact that k_z is a conserved quantity:

1. $k_s \sin \varphi_I = k_s \sin \varphi_R \curvearrowright \varphi_I = \varphi_R$ (reflection)
2. $k_s \sin \varphi_I = k_c \sin \varphi_T \curvearrowright n_s \sin \varphi_I = n_c \sin \varphi_T$ (Snell's law)

It should be noted that these formulae have been derived for the two semi-infinite half-spaces, i.e. the substrate region and the cladding region, independently of the specific layer system in between. These formulae are derived just from the continuity of the transverse wavevector component k_z at every interface.

Let us now connect the propagating waves (incident, reflected, transmitted) to the local fields at the interfaces in order to solve the reflection transmission problem:

A) Field in substrate

The fields in the substrate $F_s(x, z)$ and $G_s(x, z)$ is due to contributions from the complex amplitudes of the incident F_I and reflected field F_R :

$$F_s(x, z) = \exp(i k_z z) [F_I \exp(i k_{sx} x) + F_R \exp(-i k_{sx} x)]$$

$$G_s(x, z) = i \alpha_s k_{sx} \exp(i k_z z) [F_I \exp(i k_{sx} x) - F_R \exp(-i k_{sx} x)]$$

B) Field in layer system

The fields inside the layer system can be expressed as

$$F_f(x, z) = \exp(i k_z z) F(x)$$

$$G_f(x, z) = \exp(i k_z z) G(x)$$

where the amplitudes $F(x)$ and $G(x)$ are given as above by the matrix method as

$$\begin{pmatrix} F \\ G \end{pmatrix}_x = \hat{\mathbf{M}}(x) \begin{pmatrix} F \\ G \end{pmatrix}_0$$

C) Field in cladding

The fields in the cladding $F_c(x, z)$ and $G_c(x, z)$ are due to contributions only from the complex amplitude of the transmitted field F_T :

$$F_c(x, z) = \exp(i k_z z) F_T \exp[i k_{cx}(x - D)]$$

$$G_c(x, z) = i \alpha_c k_{cx} \exp(i k_z z) F_T \exp[i k_{cx}(x - D)]$$

Note that in the cladding we consider a forward (transmitted) wave only. Hence, we exclude reflection at any inhomogeneities after the boundary between the layer system and the cladding. For convenience, we define the phase factor in the x-direction from this interface with the last layer at $x=D$.

Reflection-transmission problem

The aim is to compute F_R and F_T for given F_I , k_z ($\sim \sin \varphi_I$), ϵ_i, d_i .

We know that F and G are continuous at each interface, in particular at $x=0$ and $x=D$. We have:

$$\begin{pmatrix} F \\ G \end{pmatrix}_D = \hat{\mathbf{M}}(D) \begin{pmatrix} F \\ G \end{pmatrix}_0.$$

Field in cladding at $x=D$ *field in substrate at $x=0$*

On the other hand, we have expressions for the fields at $x=0$ and $x=D$ from our decomposition into the incident, reflected and transmitted fields above. Hence, we can write:

$$\begin{pmatrix} F_T \\ i\alpha_c k_{cx} F_T \end{pmatrix} = \begin{pmatrix} M_{11}(D) & M_{12}(D) \\ M_{21}(D) & M_{22}(D) \end{pmatrix} \begin{pmatrix} F_I + F_R \\ i\alpha_s k_{sx} (F_I - F_R) \end{pmatrix}.$$

We consider the incident F_I as known, the reflected and transmitted fields F_R and F_T as unknown and solve for F_R and F_T :

$$F_R = \frac{(\alpha_s k_{sx} M_{22} - \alpha_c k_{cx} M_{11}) - i(M_{21} + \alpha_s k_{sx} \alpha_c k_{cx} M_{12})}{\underbrace{(\alpha_s k_{sx} M_{22} + \alpha_c k_{cx} M_{11}) + i(M_{21} - \alpha_s k_{sx} \alpha_c k_{cx} M_{12})}_N} F_I$$

$$F_T = \frac{2\alpha_s k_{sx} (M_{11} M_{22} - M_{12} M_{21})}{(\alpha_s k_{sx} M_{22} + \alpha_c k_{cx} M_{11}) + i(M_{21} - \alpha_s k_{sx} \alpha_c k_{cx} M_{12})} F_I$$

$$F_T = \frac{2\alpha_s k_{sx}}{N} F_I$$

These are the general formulae for reflected and transmitted amplitudes. Please note that the **matrix elements depend on the polarization direction** $\rightarrow M_{ij}^{\text{TE}} \neq M_{ij}^{\text{TM}}$.

Let us now transform back to the physical electric and magnetic fields, and write the solution for the reflection-transmission problem for TE and TM polarization:

A) TE-polarization

$$F = E = E_y, \quad \alpha_{\text{TE}} = 1$$

i) reflected field

$$E_R^{\text{TE}} = R_{\text{TE}} E_I^{\text{TE}}$$

with the reflection coefficient

$$R_{\text{TE}} = \frac{(k_{\text{sx}} M_{22}^{\text{TE}} - k_{\text{cx}} M_{11}^{\text{TE}}) - i(M_{21}^{\text{TE}} + k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TE}})}{(k_{\text{sx}} M_{22}^{\text{TE}} + k_{\text{cx}} M_{11}^{\text{TE}}) + i(M_{21}^{\text{TE}} - k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TE}})}$$

$\underbrace{N_{\text{TE}}}_{}$

ii) transmitted field

$$E_{\text{T}}^{\text{TE}} = T_{\text{TE}} E_{\text{I}}^{\text{TE}}$$

with the transmission coefficient

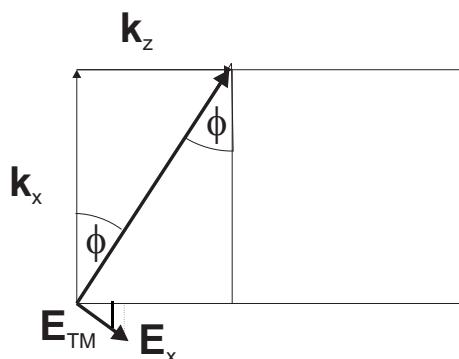
$$T_{\text{TE}} = \frac{2k_{\text{sx}}}{(k_{\text{sx}} M_{22}^{\text{TE}} + k_{\text{cx}} M_{11}^{\text{TE}}) + i(M_{21}^{\text{TE}} - k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TE}})} = \frac{2k_{\text{sx}}}{N_{\text{TE}}},$$

→ We get **complex coefficients** for reflection and transmission, which determine the amplitude and phase of the reflected and transmitted light.

B) TM-polarization

$$F = H = H_y, \quad \alpha_{\text{TM}} = \frac{1}{\epsilon}.$$

In the case of TM polarization we have the problem that an analogous calculation to TE would lead to $H_{\text{R,T}}/H_{\text{I}}$, i.e., relations between the magnetic field. However, we wish to instead find $E_{\text{R,T}}^{\text{TM}}/E_{\text{I}}^{\text{TM}}$. Therefore, we must convert the H -field to the E^{TM} -field:



As can be seen in the figure, we can express the amplitude of the E_{TM} field in terms of the E_x component:

$$\begin{aligned} \frac{E_x}{E^{\text{TM}}} &= -\sin \varphi = -\frac{k_z}{k} \\ \therefore E^{\text{TM}} &= -\frac{k}{k_z} E_x, \end{aligned}$$

With Maxwell's equations we can link E_x to H_y :

$$\mathbf{E} = -\frac{1}{\omega \epsilon_0 \epsilon} (\mathbf{k} \times \mathbf{H}) \rightarrow E_x = \frac{1}{\omega \epsilon_0 \epsilon} k_z H_y \rightarrow E^{\text{TM}} = -\frac{k}{\omega \epsilon_0 \epsilon} H_y = -\frac{1}{c \epsilon_0 \sqrt{\epsilon}} H_y$$

Result: $\frac{E_{\text{R,T}}^{\text{TM}}}{E_{\text{I}}^{\text{TM}}} = \sqrt{\frac{\epsilon_s}{\epsilon_c}} \frac{H_{\text{R,T}}}{H_{\text{I}}}$, $\rightarrow \sqrt{\epsilon_s / \epsilon_c}$ relevant only for transmission

Hence we find the following for TM polarization:

$$E_{\text{R}}^{\text{TM}} = R_{\text{TM}} E_{\text{I}}^{\text{TM}}$$

with the reflection coefficient

$$R_{\text{TM}} = \frac{(\epsilon_c k_{\text{sx}} M_{22}^{\text{TM}} - \epsilon_s k_{\text{cx}} M_{11}^{\text{TM}}) - i(\epsilon_s \epsilon_c M_{21}^{\text{TM}} + k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TM}})}{(\epsilon_c k_{\text{sx}} M_{22}^{\text{TM}} + \epsilon_s k_{\text{cx}} M_{11}^{\text{TM}}) + i(\epsilon_s \epsilon_c M_{21}^{\text{TM}} - k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TM}})}$$

$$E_{\text{T}}^{\text{TM}} = T_{\text{TM}} E_{\text{I}}^{\text{TM}}$$

with the transmission coefficient

$$T_{\text{TM}} = \frac{2\sqrt{\epsilon_s \epsilon_c} k_{\text{sx}}}{(\epsilon_c k_{\text{sx}} M_{22}^{\text{TM}} + \epsilon_s k_{\text{cx}} M_{11}^{\text{TM}}) + i(\epsilon_s \epsilon_c M_{21}^{\text{TM}} - k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TM}})} = \frac{2\sqrt{\epsilon_s \epsilon_c} k_{\text{sx}}}{N_{\text{TM}}}$$

In summary, we have found different complex coefficients for reflection and transmission for TE and TM polarization. The **resulting generalized Fresnel formulas** for multilayered systems are

$$R_{\text{TE}} = \frac{(k_{\text{sx}} M_{22}^{\text{TE}} - k_{\text{cx}} M_{11}^{\text{TE}}) - i(M_{21}^{\text{TE}} + k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TE}})}{(k_{\text{sx}} M_{22}^{\text{TE}} + k_{\text{cx}} M_{11}^{\text{TE}}) + i(M_{21}^{\text{TE}} - k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TE}})}$$

$$T_{\text{TE}} = \frac{2k_{\text{sx}}}{N_{\text{TE}}}$$

$$R_{\text{TM}} = \frac{(\epsilon_c k_{\text{sx}} M_{22}^{\text{TM}} - \epsilon_s k_{\text{cx}} M_{11}^{\text{TM}}) - i(\epsilon_s \epsilon_c M_{21}^{\text{TM}} + k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TM}})}{(\epsilon_c k_{\text{sx}} M_{22}^{\text{TM}} + \epsilon_s k_{\text{cx}} M_{11}^{\text{TM}}) + i(\epsilon_s \epsilon_c M_{21}^{\text{TM}} - k_{\text{sx}} k_{\text{cx}} M_{12}^{\text{TM}})}$$

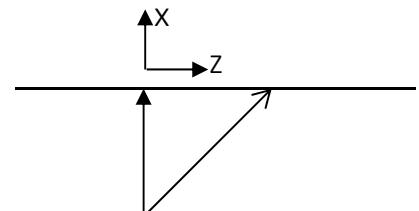
$$T_{\text{TM}} = \frac{2\sqrt{\epsilon_s \epsilon_c} k_{\text{sx}}}{N_{\text{TM}}}.$$

8.3.1.2 Reflectivity and transmissivity

In the previous chapter we have computed the coefficients of reflection and transmission, which relate the electric fields in TE and TM polarization of the incident, reflected and transmitted waves. However, in many situations it is more important to consider the relation between **energy fluxes**, the so-called **reflectivity and transmissivity**. In order to obtain information on these quantities we must compute the energy flux perpendicular to the interface:

- flux through a surface with $x = \text{const}$

For a monochromatic plane wave, it follows that



$$\langle \mathbf{S} \rangle \mathbf{e}_x = \frac{1}{2} \Re(\mathbf{E} \times \mathbf{H}^*) \mathbf{e}_x$$

With $\mathbf{H}^* = \frac{1}{\omega \mu_0} (\mathbf{k}^* \times \mathbf{E}^*)$

after evaluating the successive cross products and using the transversality of the electric field, we find that

$$\langle \mathbf{S} \rangle \mathbf{e}_x = \frac{1}{2\omega\mu_0} \Re(\mathbf{k}^* \mathbf{e}_x) |\mathbf{E}|^2 = \frac{1}{2\omega\mu_0} \Re(k_x^*) |\mathbf{E}|^2.$$

Note once more that only the real part of the wave vector contributes to energy flow. Since in an absorption-free medium the energy flux is conserved, in an absorption-free layer system the energy flux is also conserved.

In the substrate

$$k_{sx} = \sqrt{\frac{\omega^2}{c^2} \epsilon_s - k_z^2} = \sqrt{k_s^2(\omega) - k_z^2}$$

must be real-valued, since our incident wave propagates from there. The total energy flux from the substrate to the layer system, after subtracting by the reflected flux, is given by

$$\langle \mathbf{S} \rangle_s \mathbf{e}_x = \frac{1}{2\omega\mu_0} [k_{sx} |\mathbf{E}_I|^2 - k_{sx} |\mathbf{E}_R|^2]$$

In contrast, in the cladding

$$k_{cx} = \sqrt{\frac{\omega^2}{c^2} \epsilon_c - k_z^2} = \sqrt{k_c^2(\omega) - k_z^2}$$

can in general be complex-valued. The energy flux from the layer system into the cladding is

$$\langle \mathbf{S} \rangle_c \mathbf{e}_x = \frac{1}{2\omega\mu_0} \Re(k_{cx}) |\mathbf{E}_T|^2.$$

Because we have energy conservation:

$$\langle \mathbf{S} \rangle_s \mathbf{e}_x = \langle \mathbf{S} \rangle_c \mathbf{e}_x \curvearrowright$$

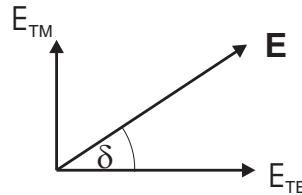
$$|\mathbf{E}_I|^2 = |\mathbf{E}_R|^2 + \frac{\Re(k_{cx})}{k_{sx}} |\mathbf{E}_T|^2$$

Now we will compute the **global** reflectivity ρ and transmissivity τ of a layer system. Of course, we will decompose into TE and TM polarizations and use the reflectivities $\rho_{TE,TM}$ and transmissivities $\tau_{TE,TM}$. Since the TE & TM polarizations are orthogonal, we can consider their energy fluxes independently:

$$\mathbf{E}_R = \mathbf{E}_R^{TE} + \mathbf{E}_R^{TM}, \quad \mathbf{E}_T = \mathbf{E}_T^{TE} + \mathbf{E}_T^{TM}$$

$$\begin{aligned} |\mathbf{E}_I|^2 &= |\mathbf{E}_R^{TE}|^2 + |\mathbf{E}_R^{TM}|^2 + \frac{\Re(k_{cx})}{k_{sx}} \left(|\mathbf{E}_T^{TE}|^2 + |\mathbf{E}_T^{TM}|^2 \right) \\ &= \left\{ |R_{TE}|^2 + \frac{\Re(k_{cx})}{k_{sx}} |T_{TE}|^2 \right\} |\mathbf{E}_I^{TE}|^2 + \left\{ |R_{TM}|^2 + \frac{\Re(k_{cx})}{k_{sx}} |T_{TM}|^2 \right\} |\mathbf{E}_I^{TM}|^2. \end{aligned}$$

Here, we simply substituted the reflected and transmitted field amplitudes by the incident amplitudes multiplied by Fresnel coefficients. Now, we decompose the **incident** field as follows:



$$E_I^{TE} = |\mathbf{E}_I| \cos \delta, \quad E_I^{TM} = |\mathbf{E}_I| \sin \delta.$$

Then, we can divide by the (arbitrary) amplitude $|\mathbf{E}_I|^2$ and write

$$\begin{aligned} 1 &= \left\{ |R_{TE}|^2 + \frac{\Re(k_{cx})}{k_{sx}} |T_{TE}|^2 \right\} \cos^2 \delta + \left\{ |R_{TM}|^2 + \frac{\Re(k_{cx})}{k_{sx}} |T_{TM}|^2 \right\} \sin^2 \delta \\ 1 &= \underbrace{\left(|R_{TE}|^2 \cos^2 \delta + |R_{TM}|^2 \sin^2 \delta \right)}_{\rho} + \underbrace{\frac{\Re(k_{cx})}{k_{sx}} \left(|T_{TE}|^2 \cos^2 \delta + |T_{TM}|^2 \sin^2 \delta \right)}_{\tau} \end{aligned}$$

The red (reflected) and blue (transmitted) terms can be identified as $1 = \rho + \tau$, which ensures energy conservation.

The global reflectivity and transmissivity are therefore given as

$$\boxed{\begin{aligned} \rho &= \rho_{TE} \cos^2 \delta + \rho_{TM} \sin^2 \delta \\ \tau &= \tau_{TE} \cos^2 \delta + \tau_{TM} \sin^2 \delta \end{aligned}}$$

with the reflectivities

$$\boxed{\rho_{TE,TM} = |R_{TE,TM}|^2, \quad \tau_{TE,TM} = \frac{\Re(k_{cx})}{k_{sx}} |T_{TE,TM}|^2.}$$

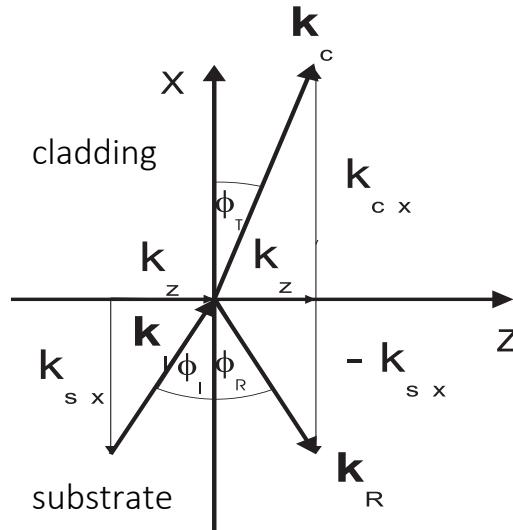
for the two polarization states TE and TM.

8.3.2 Single interface

8.3.2.1 (Classical) Fresnel formulas

Let us now consider the important example of the most simple layer system, namely the single interface. The relevant wave vectors are (as usual):

$$\mathbf{k}_I = \begin{pmatrix} k_{sx} \\ 0 \\ k_z \end{pmatrix}, \quad \mathbf{k}_R = \begin{pmatrix} -k_{sx} \\ 0 \\ k_z \end{pmatrix}, \quad \mathbf{k}_T = \begin{pmatrix} k_{cx} \\ 0 \\ k_z \end{pmatrix}.$$



The continuous component of the wave vector, expressed in terms of the angle of incidence, is

$$k_z = \frac{\omega}{c} \sqrt{\epsilon_s} \sin \varphi_I = \frac{\omega}{c} n_s \sin \varphi_I$$

Then, the discontinuous component is given by

$$\rightarrow k_{ix} = \sqrt{\frac{\omega^2}{c^2} \epsilon_i - k_z^2} = \sqrt{\frac{\omega^2}{c^2} \epsilon_i - \frac{\omega^2}{c^2} \epsilon_s \sin^2 \varphi_I} = \frac{\omega}{c} \sqrt{n_i^2 - n_s^2 \sin^2 \varphi_I}$$

$$\curvearrowright k_{sx} = \frac{\omega}{c} n_s \cos \varphi_I, \quad k_{cx} = \frac{\omega}{c} \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I} = \frac{\omega}{c} n_c \cos \varphi_T,$$

As above, we can assume that k_{sx} is always real, because otherwise we have no propagating incident wave. k_{cx} is **real** for $n_c > n_s \sin \varphi_I$, but **imaginary** for $n_c < n_s \sin \varphi_I$ (total internal reflection).

The matrix for a single interface is the unit matrix

$$\hat{\mathbf{M}} = \hat{\mathbf{m}}(d=0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and it is easy to compute coefficients for reflection and transmission, and reflectivity and transmissivity. Using the formulas from above we find:

A) TE-polarization

$$R_{TE} = \frac{(k_{sx} M_{22} - k_{cx} M_{11}) - i(M_{21} + k_{sx} k_{cx} M_{12})}{(k_{sx} M_{22} + k_{cx} M_{11}) + i(M_{21} - k_{sx} k_{cx} M_{12})}, \quad T_{TE} = \frac{2k_{sx}}{N_{TE}}$$

$\underbrace{N_{TE}}$

with: $\hat{\mathbf{M}} = \hat{\mathbf{m}}(d=0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$R_{\text{TE}} = \frac{(k_{\text{sx}} - k_{\text{cx}})}{(k_{\text{sx}} + k_{\text{cx}})} = \frac{n_s \cos \varphi_I - \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I}}{n_s \cos \varphi_I + \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I}} = \frac{n_s \cos \varphi_I - n_c \cos \varphi_T}{n_s \cos \varphi_I + n_c \cos \varphi_T}$$

$$T_{\text{TE}} = \frac{2k_{\text{sx}}}{(k_{\text{sx}} + k_{\text{cx}})} = \frac{2n_s \cos \varphi_I}{n_s \cos \varphi_I + \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I}} = \frac{2n_s \cos \varphi_I}{n_s \cos \varphi_I + n_c \cos \varphi_T}$$

$$\rho_{\text{TE}} = |R_{\text{TE}}|^2 = \frac{|k_{\text{sx}} - k_{\text{cx}}|^2}{|k_{\text{sx}} + k_{\text{cx}}|^2}$$

$$\tau_{\text{TE}} = \frac{\Re(k_{\text{cx}})}{k_{\text{sx}}} |T_{\text{TE}}|^2 = \frac{4k_{\text{sx}} \Re(k_{\text{cx}})}{|k_{\text{sx}} + k_{\text{cx}}|^2}.$$

$$\curvearrowright \rho_{\text{TE}} + \tau_{\text{TE}} = 1$$

B) TM-Polarisation

$$R_{\text{TM}} = \underbrace{\frac{(\varepsilon_c k_{\text{sx}} M_{22} - \varepsilon_s k_{\text{cx}} M_{11}) - \mathbf{i}(\varepsilon_s \varepsilon_c M_{21} + k_{\text{sx}} k_{\text{cx}} M_{12})}{(\varepsilon_c k_{\text{sx}} M_{22} + \varepsilon_s k_{\text{cx}} M_{11}) + \mathbf{i}(\varepsilon_s \varepsilon_c M_{21} - k_{\text{sx}} k_{\text{cx}} M_{12})}}_{N_{\text{TM}}} \quad T_{\text{TM}} = \frac{2\sqrt{\varepsilon_s \varepsilon_c} k_{\text{sx}}}{N_{\text{TM}}}.$$

with: $\hat{\mathbf{M}} = \hat{\mathbf{m}}(d=0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$$R_{\text{TM}} = \frac{(k_{\text{sx}} \varepsilon_c - k_{\text{cx}} \varepsilon_s)}{(k_{\text{sx}} \varepsilon_c + k_{\text{cx}} \varepsilon_s)} = \frac{n_s n_c^2 \cos \varphi_I - n_s^2 \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I}}{n_s n_c^2 \cos \varphi_I + n_s^2 \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I}} = \frac{n_c \cos \varphi_I - n_s \cos \varphi_T}{n_c \cos \varphi_I + n_s \cos \varphi_T}$$

$$T_{\text{TM}} = \frac{2k_{\text{sx}} \sqrt{\varepsilon_c \varepsilon_s}}{(k_{\text{sx}} \varepsilon_c + k_{\text{cx}} \varepsilon_s)} = \frac{2n_s^2 n_c \cos \varphi_I}{n_s n_c^2 \cos \varphi_I + n_s^2 \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I}} = \frac{2n_s \cos \varphi_I}{n_c \cos \varphi_I + n_s \cos \varphi_T},$$

$$\rho_{\text{TM}} = |R_{\text{TM}}|^2 = \frac{|k_{\text{sx}} \varepsilon_c - k_{\text{cx}} \varepsilon_s|^2}{|k_{\text{sx}} \varepsilon_c + k_{\text{cx}} \varepsilon_s|^2},$$

$$\tau_{\text{TM}} = \frac{\Re(k_{\text{cx}})}{k_{\text{sx}}} |T_{\text{TM}}|^2 = \frac{4k_{\text{sx}} \Re(k_{\text{cx}}) \varepsilon_s \varepsilon_c}{|k_{\text{sx}} \varepsilon_c + k_{\text{cx}} \varepsilon_s|^2}$$

$$\curvearrowright \rho_{\text{TM}} + \tau_{\text{TM}} = 1$$

Remark

It may seem that we have a problem for $\varphi_I = 0$. For $\varphi_I = 0$, the TE and TM polarizations should be equivalent, since the fields are always polarized parallel

to the interface. However, formally we have $R_{\text{TE}} = -R_{\text{TM}}$, $T_{\text{TE}} = T_{\text{TM}}$. The “strange” behavior of the coefficient of reflection can be explained by the following figures:



The positive direction of the TM field is defined by the propagation direction through the right-hand rule for cross products. Even though for $\varphi_I = 0$ the TE and TM polarizations are equivalent, the definitions of axes for their respective reflected components differ. Thus there is a relative sign between the reflection coefficients.

8.3.2.2 Total internal reflection (TIR) for $\epsilon_s > \epsilon_c$

Let us now consider the special case when all incident light is reflected from the interface. This means that the reflectivity is unity.

$$\rho_{\text{TE}} = \frac{|k_{sx} - k_{cx}|^2}{|k_{sx} + k_{cx}|^2} \quad \rho_{\text{TM}} = \frac{|k_{sx}\epsilon_c - k_{cx}\epsilon_s|^2}{|k_{sx}\epsilon_c + k_{cx}\epsilon_s|^2}$$

With the normal component of the wavevector $k_{cx} = \frac{\omega}{c} \sqrt{n_c^2 - n_s^2 \sin^2 \varphi_I} = 0$ we can compute the smallest angle of incidence with $\rho_{\text{TE,TM}} = 1$:

$$k_{cx} = 0 \curvearrowright n_c = n_s \sin \varphi_{I\text{tot}}$$

$$\boxed{\sin \varphi_{I\text{tot}} = \frac{n_c}{n_s}.}$$

For angles of incidence larger than this critical angle, $\varphi_I > \varphi_{I\text{tot}}$ we have

$$k_{cx} = i \frac{\omega}{c} \sqrt{n_s^2 \sin^2 \varphi_I - n_c^2} = i \mu_c = i \sqrt{k_z^2 - \frac{\omega^2}{c^2} \epsilon_c} \curvearrowright \text{imaginary}$$

$$\rightarrow \Re(k_{cx}) = 0 \rightarrow \text{TIR}$$

The intuition here is that above the critical angle, the continuous tangential component k_z exceeds the cladding wavenumber as determined by the corresponding dispersion relation. As such, only evanescent waves can be excited in the cladding and thus the incident wave is entirely reflected.

Obviously, we find the same critical angle of TIR for TE and TM polarization since here we have only considered a condition on the wavevector, not the polarizations. The energy fluxes are given as (here TE, same result for TM):

$$\rho_{\text{TE}} = \frac{|k_{sx} - i\mu_c|^2}{|k_{sx} + i\mu_c|^2} = 1 \quad \tau_{\text{TE}} = \frac{4k_{sx}\Re(k_{cx})}{|k_{sx} + k_{cx}|^2} = 0.$$

Remark

For metals in visible range (below the plasma frequency) we have always TIR, because: $\Re(\epsilon_c) < 0 \rightarrow k_{cx} = \frac{\omega}{c} \sqrt{\epsilon_c - n_s^2 \sin^2 \varphi_I}$ is always imaginary.

In the case of TIR the modulus of the coefficient of reflection is one, but the coefficient itself is complex \rightarrow nontrivial phase shift for reflected light:

A) TE-polarization

As we found above, the normal component of the wave number in the cladding is purely imaginary when total internal reflection occurs. The expression for the reflection coefficient thus becomes:

$$R_{\text{TE}} = 1 \cdot \exp(i\Theta_{\text{TE}}) = \frac{k_{sx} - i\mu_c}{k_{sx} + i\mu_c} = \frac{Z}{Z^*} = \frac{\exp(i\alpha)}{\exp(-i\alpha)} = \exp(2i\alpha)$$

$$\curvearrowright \tan \alpha = \tan \frac{\Theta_{\text{TE}}}{2} = -\frac{\mu_c}{k_{sx}} = -\frac{\sqrt{n_s^2 \sin^2 \varphi_I - n_c^2}}{n_s \cos \varphi_I} = -\frac{\sqrt{\sin^2 \varphi_I - \sin^2 \varphi_{I\text{tot}}}}{\cos \varphi_I}.$$

B) TM-polarization

$$R_{\text{TM}} = 1 \cdot \exp(i\Theta_{\text{TM}}) = \frac{k_{sx}\epsilon_c - i\mu_c\epsilon_s}{k_{sx}\epsilon_c + i\mu_c\epsilon_s} = \frac{Z}{Z^*} = \frac{\exp(i\alpha)}{\exp(-i\alpha)} = \exp(2i\alpha)$$

$$\curvearrowright \tan \alpha = \tan \frac{\Theta_{\text{TM}}}{2} = -\frac{\mu_c\epsilon_s}{k_{sx}\epsilon_c} = \frac{\epsilon_s}{\epsilon_c} \tan \frac{\Theta_{\text{TE}}}{2},$$

In conclusion, we have seen that the phase shifts of the reflected light at TIR is different for TE and TM polarization, and because $\epsilon_s > \epsilon_c$, the phase shift is larger than for the TM polarization

$$|\Theta_{\text{TM}}| > |\Theta_{\text{TE}}|,$$

As a consequence, incident linearly polarized light in general becomes elliptically polarized after TIR \curvearrowright Fresnel prism

Remark

The field in the cladding is **evanescent** $\sim \exp(i k_{xc} x) = \exp(-\mu_c x)$.

\curvearrowright The averaged energy flux in the cladding normal to the interface vanishes.

$$\langle \mathbf{S} \rangle_x = \frac{1}{2\omega\mu_0} \Re(\mathbf{k}|\mathbf{E}|^2)_x = \frac{1}{2\omega\mu_0} \underbrace{\Re(k_x)}_{=0} |\mathbf{E}|^2 = 0.$$

8.3.2.3 The Brewster angle

There exists another special angle with interesting reflection properties. For TM polarization incident at the Brewster angle φ_B we find $R_{\text{TM}} = 0$:

$$\rho_{\text{TM}} = \frac{|k_{sx}\varepsilon_c - k_{cx}\varepsilon_s|^2}{|k_{sx}\varepsilon_c + k_{cx}\varepsilon_s|^2} = 0,$$

$$\rightarrow k_{sx}\varepsilon_c = k_{cx}\varepsilon_s$$

$$\varepsilon_c^2 (\varepsilon_s - \sin^2 \varphi_B \varepsilon_s) = \varepsilon_s^2 (\varepsilon_c - \sin^2 \varphi_B \varepsilon_s)$$

$$\sin^2 \varphi_B = \frac{\varepsilon_s \varepsilon_c (\varepsilon_s - \varepsilon_c)}{\varepsilon_s (\varepsilon_s^2 - \varepsilon_c^2)} = \frac{\varepsilon_c}{(\varepsilon_s + \varepsilon_c)}$$

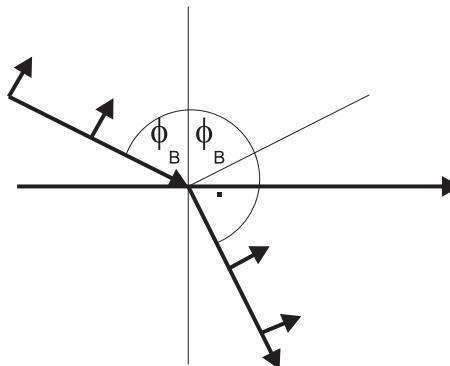
$$\cos^2 \varphi_B = 1 - \sin^2 \varphi_B = 1 - \frac{\varepsilon_c}{(\varepsilon_s + \varepsilon_c)} = \frac{\varepsilon_s}{(\varepsilon_s + \varepsilon_c)}$$

With the last two lines we can write the final result for the Brewster angle:

$$\boxed{\tan \varphi_B = \sqrt{\frac{\varepsilon_c}{\varepsilon_s}}}.$$

The Brewster angle exists only for TM polarization, but for any $n_s \leq n_c$.

There is a simple physical explanation for why there is no reflection from the interface at the Brewster angle.



$$\tan \varphi_B = \frac{\sin \varphi_B}{\cos \varphi_B} = \frac{n_c}{n_s}$$

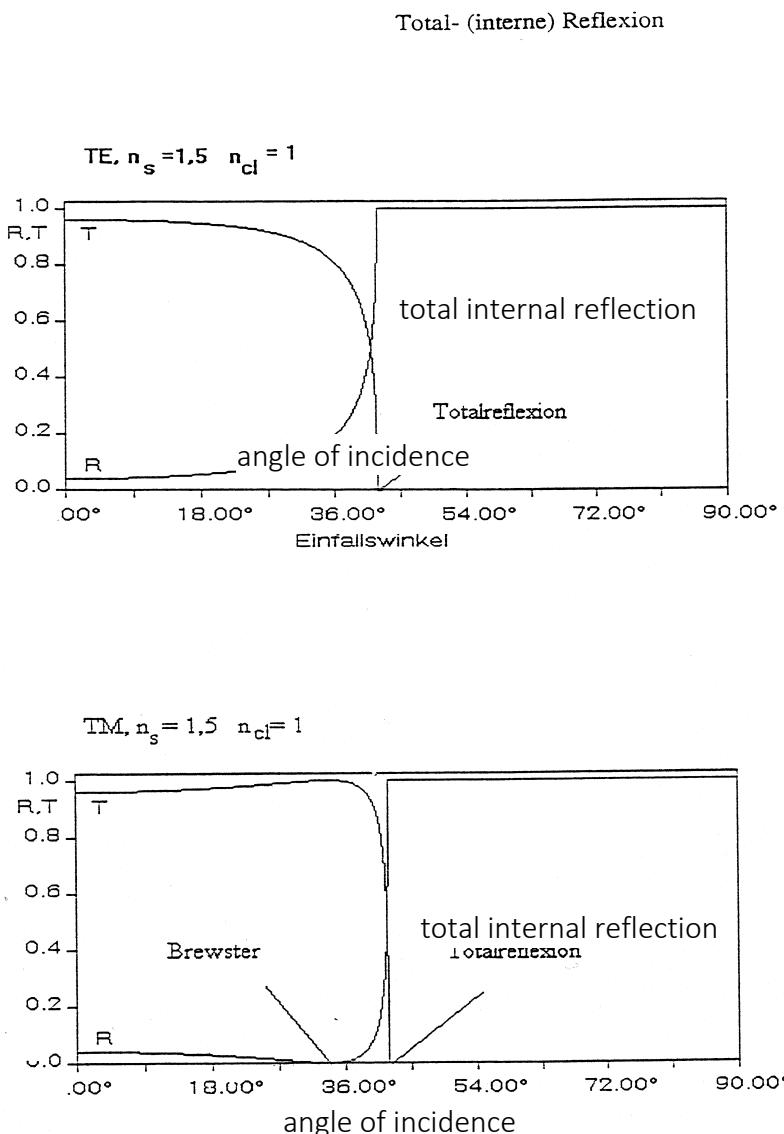
$$\curvearrowright n_s \sin \varphi_B = n_c \cos \varphi_B = n_c \sin\left(\frac{\pi}{2} - \varphi_B\right)$$

But from Snell's law, the angle of the transmitted light is always

$$n_s \sin \varphi_B = n_c \sin \varphi_T \curvearrowright \varphi_T = \frac{\pi}{2} - \varphi_B,$$

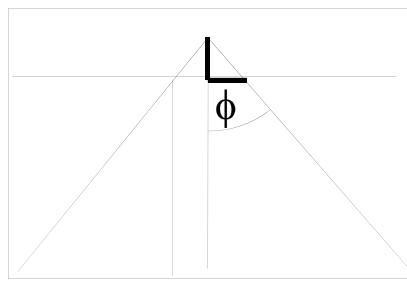
Hence, at the Brewster angle, any reflected and transmitted waves would propagate in perpendicular directions. If we interpret any reflected light as an emission from oscillating dipoles in the cladding which are induced by the transmitted light, no reflected wave can occur for TM polarization since there can be no radiation in the direction of dipole oscillation.

In summary, we have the following results for reflectivity and transmissivity at a single interface with $\epsilon_s > \epsilon_c$.



8.3.2.4 The Goos-Hänchen-Shift

The Goos-Hänchen shift is a direct consequence of the nontrivial phase shift of the reflected light during TIR. It occurs when beams undergo total internal reflection at an interface. The reflected beam appears to be shifted along the interface. As a result, it appears as if the beam penetrates the cladding and reflection occurs at a plane parallel to the interface at a certain depth, the so-called penetration depth. For the sake of simplicity, here we will treat TE-polarization only.



Let us start with an incident plane wave in TE polarization:

$$E_I(x, z) = E_I \exp[i(k_{sx}x + k_z z)] \rightarrow E_I(x, z) = E_I \exp[i(\alpha z + \gamma_s x)]$$

with

$$\text{transverse wave number: } \alpha = \frac{\omega}{c} n_s \sin \varphi_I$$

$$\text{longitudinal wave number: } \gamma_s = \sqrt{\frac{\omega^2}{c^2} n_s^2 - \alpha^2}$$

This gives rise to a reflected plane wave (minus sign associated with the longitudinal wave number):

$$E_R(x, z) = E_I \exp[i(\alpha z - \gamma_s x)] \exp[i\Theta(\alpha)].$$

The reflected plane wave receives a phase shift $\Theta(\alpha)$, which depends on the angle of incidence (here characterized by the transverse wave number α).

Now we want to treat beams, which we can express as a superposition of plane waves with Fourier amplitude $e_I(\alpha)$:

$$E_I(x, z) = \int d\alpha e_I(\alpha) \exp[i(\alpha z + \gamma_s(\alpha)x)]$$

We assume a mean angle of incidence φ_{I0} for this superposition of plane waves:

$$\alpha_0 = \frac{\omega}{c} n_s \sin \varphi_{I0} \text{ mean angular frequency}$$

$$\alpha = \alpha_0 + \varepsilon$$

In the Fourier integral, we only have to integrate over angular frequencies with non-zero amplitudes $e_I(\alpha)$, i.e. $e_I(\alpha_0 + \varepsilon) \neq 0$ for $-\Delta \leq \varepsilon \leq \Delta$ only.

$$E_I(x, z) = \exp(i\alpha_0 z) \int_{-\Delta}^{\Delta} e_I(\alpha_0 + \varepsilon) \exp[i(\varepsilon z + \gamma_s(\alpha_0 + \varepsilon)x)] d\varepsilon$$

$$E_R(x, z) = \exp(i\alpha_0 z) \int_{-\Delta}^{\Delta} e_I(\alpha_0 + \varepsilon) \exp[i(\varepsilon z - \gamma_s(\alpha_0 + \varepsilon)x)] |R(\alpha_0 + \varepsilon)| \exp[i\Theta(\alpha_0 + \varepsilon)] d\varepsilon$$

Let us make the following further assumptions:

- small divergence of beam (narrow spectrum, $\Delta \ll \frac{\omega}{c} n_s$)

- all Fourier component undergo TIR ($\Theta(\alpha) > \Theta_{tot}$) $\rightarrow \alpha > \alpha_{tot} \rightarrow$ the reflection coefficient $|R(\alpha_0 + \varepsilon)| = 1$ in the equation above

It is then justified to expand the phase shift $\Theta(\alpha)$ of the reflected wave in a Taylor series up to first order:

$$\Theta(\alpha_0 + \varepsilon) \approx \Theta(\alpha_0) + \frac{\partial \Theta}{\partial \alpha} \Big|_{\alpha_0} \varepsilon = \Theta(\alpha_0) + \Theta' \varepsilon$$

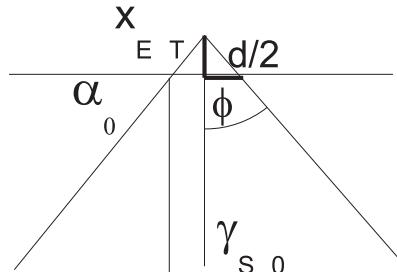
Thus, the reflected beam at the interface at $x=0$ is given by:

$$E_R(0, z) = \exp\left\{i[\alpha_0 z + \Theta(\alpha_0)]\right\} \int_{-\Delta}^{\Delta} d\varepsilon e_i(\alpha_0 + \varepsilon) \exp[i(z + \Theta')\varepsilon]$$

The first factor corresponds to a carrier wave at the mean angular frequency, after undergoing a phase change due to the reflection. This is modulated by the remaining integral in red, which we can identify as a shifted version of the incident beam profile at $x=0$ (with an additional spatially independent phase shift)

$$E_R(0, z) = \exp\left\{i[\Theta(\alpha_0) - \alpha_0 \Theta']\right\} E_i(z + \Theta').$$

Thus, the reflected beam appears shifted by $d = -\Theta'$ (Goos-Hänchen Shift).



Let us finally compute the shift $d = -\Theta'$. We know from before that the phase shift for TIR is given by :

$$\tan \frac{\Theta_{TE}}{2} = -\frac{\mu_c}{k_{sx}} = -\frac{\mu_c}{\gamma_s}$$

$$\Theta = -2 \arctan \frac{\mu_c}{k_{sx}} = -2 \arctan \frac{\sqrt{\alpha^2 - \frac{\omega^2}{c^2} n_c^2}}{\sqrt{\frac{\omega^2}{c^2} n_s^2 - \alpha^2}}$$

$$\Theta' = \frac{\partial \Theta}{\partial \alpha} = -2 \times \frac{1}{1 + \frac{\mu_c^2}{k_{sx}^2}} \times \frac{\frac{2\alpha}{2\mu_c} k_{sx} - \frac{-2\alpha}{2k_{sx}} \mu_c}{k_{sx}^2} = -2 \frac{\frac{\alpha(k_{sx}^2 + \mu_c^2)}{\mu_c k_{sx}}}{k_{sx}^2 + \mu_c^2} = -2 \frac{1}{\mu_c} \frac{\alpha}{k_{sx}}$$

$$\boxed{\Theta'|_{\alpha_0} = -d = -2x_{ET} \tan \varphi_{IO}} \text{ with } x_{ET} = \frac{1}{\mu_c} = \frac{1}{\sqrt{\alpha_0^2 - \frac{\omega^2}{c^2} n_c^2}} \text{ and } \tan \varphi_{IO} = \frac{\alpha}{k_{sx}} = \frac{\alpha}{\gamma_s}$$

$\rightarrow x_{ET}$ depth of penetration

8.3.3 Periodic multi-layer systems – Bragg mirrors – 1D photonic crystals

In the previous chapters, we have learnt how to treat (finite) arbitrary multi-layered systems. Beyond this, interesting effects occur when these multi-layered systems become semi-infinite periodic stacks of layers, so-called Bragg mirrors. The reflectivities of such mirrors is almost 100 % in certain frequency ranges; the more layers we have, the closer we approach this ideal value. Bragg mirrors are important for building resonators (lasers, interferometers). Furthermore, periodic structures are of general importance in physics (lattices, crystals, atomic chains etc.). We can learn much about the general features of such periodic systems by looking at the optical properties of periodic (dielectric) multi-layered systems.

In our theoretical approach, we will assume these layer systems to be infinite, i.e. consisting of an infinite number of layers, and we treat them as so-called one-dimensional photonic crystals. We will discuss effects like band gaps, dispersion and diffraction in such periodic media and gain an understanding of the basics of Bragg reflection and the physics of photonic crystals.

In order to keep things simple we will treat:

- semi-infinite periodic multi-layered systems with

$$[x > 0, (\varepsilon_1, d_1), (\varepsilon_2, d_2), (\varepsilon_1, d_1), (\varepsilon_2, d_2), \dots]$$
- placed on a substrate $[x < 0, \varepsilon_s]$
- TE-polarization only
- monochromatic light

At the interface between the substrate and the Bragg-mirror ($x=0$), we have the incident and the reflected electric field:

$$E_0 = E_R + E_I \quad \text{and} \quad \left. \frac{\partial E}{\partial x} \right|_0 = E'_0 = ik_{sx} (E_I - E_R)$$

where the minus sign associated with the reflected field is obtained by differentiating the corresponding phase factor.

Alternatively, the incident and reflected field can be expressed in terms of the field at the interface and its derivative as

$$E_I = \frac{E_0}{2} - \frac{iE'_0}{2k_{sx}} \quad \text{and} \quad E_R = \frac{E_0}{2} + \frac{iE'_0}{2k_{sx}}.$$

In chapter 8.2, we developed a matrix formalism involving the generalized fields F and G . Because here we only treat the TE polarization, we can directly use the electric field amplitude E and its derivative E' with respect to x , because

$$E = F \quad \text{and} \quad i\omega\mu_0 H_z = G = \frac{\partial E}{\partial x} = E'.$$

Let us now calculate the field in the multi-layered system. From before, we know how to treat finite systems with the matrix method. Here we want to treat an **infinite periodic medium** (like a one-dimensional crystal).

As a particular example, we will investigate an infinite system consisting of just two periodically repeated layers. The two layers should consist of homogeneous materials with ϵ_1 and ϵ_2 and have a thickness of d_1 and d_2 , respectively. Hence we have:

$$\epsilon(x) = \epsilon(x + \Lambda) \text{ with the period } \Lambda = d_1 + d_2$$

For infinite periodic media, we can make use of the so-called Bloch's theorem to find the generalized normal modes (Bloch modes or Bloch waves). We seek solutions in the form of:

$$E(x, z; \omega) = \exp\left\{\mathbf{i}\left[k_x(k_z, \omega)x + k_z z\right]\right\} E_{k_x}(x)$$

with

$$E_{k_x}(x + \Lambda) = E_{k_x}(x)$$

In other words, $E_{k_x}(x)$ is a periodic function and we are looking for solutions with the same periodicity. In particular, the solutions should have the same amplitude after one period of the layered system, though they can have a different phase. This phase difference cannot assume arbitrary values but must instead be given by $\sim \exp\left\{\mathbf{i}\left[k_x(k_z, \omega)x\right]\right\}$. Here k_x is the **Bloch vector**, which is currently unknown. Because in this simple example we deal with a one-dimensional problem, the Bloch vector becomes a scalar.

In the following, we will find a dispersion relation for the Bloch waves $k_x(k_z, \omega)$, in complete analogy to the DR for plane waves $k_x^2 = \frac{\omega^2}{c^2} \epsilon(\omega) - k_z^2$ in homogeneous media. In order to highlight the difference from the homogeneous case, we change the notation for the Bloch vector to:

$$k_x \rightarrow K.$$

According to Bloch's theorem (our ansatz), we have a relation for E and E' when we advance by one period of the multi-layered system (from period N to period $N+1$):

$$\begin{pmatrix} E \\ E' \end{pmatrix}_{(N+1)\Lambda} = \exp(\mathbf{i}K\Lambda) \begin{pmatrix} E \\ E' \end{pmatrix}_{N\Lambda}.$$

On the other hand, we know from our matrix method that the fields before and after the two layers that make up the period are connected by:

$$\begin{pmatrix} E \\ E' \end{pmatrix}_{(N+1)\Lambda} = \hat{\mathbf{M}} \begin{pmatrix} E \\ E' \end{pmatrix}_{N\Lambda} \text{ with } \hat{\mathbf{M}} = \hat{\mathbf{m}}(d_2) \hat{\mathbf{m}}(d_1) \rightarrow M_{ij} = \sum_k m_{ik}^{(2)} m_{kj}^{(1)}.$$

If the Bloch wave is a solution to our problem, we can set the two expressions equal:

$$\{\hat{\mathbf{M}} - \underbrace{\exp(iK\Lambda)}_{\mu} \hat{\mathbf{I}}\} \begin{pmatrix} E \\ E' \end{pmatrix}_{N\Lambda} = 0.$$

Defining $\mu = \exp(iK\Lambda)$, we can formulate the following eigenvalue problem:

$$\{\hat{\mathbf{M}} - \mu(K) \hat{\mathbf{I}}\} \begin{pmatrix} E \\ E' \end{pmatrix}_{N\Lambda} = 0.$$

This eigenvalue problem determines the Bloch vector K and will finally give our dispersion relation.

As usual, we use the determinant condition $\det\{\hat{\mathbf{M}} - \mu \hat{\mathbf{I}}\} = 0$ to compute the dispersion relation expressed in terms of μ . Hence, we still need to compute K afterwards with $\mu = \exp(iK\Lambda)$, as

$$\mu_{\pm} = \exp(iK_{\pm}\Lambda) = \frac{(M_{11} + M_{22})}{2} \pm \sqrt{\left\{ \frac{(M_{11} + M_{22})}{2} \right\}^2 - 1}. \quad (*)$$

Note that we assumed non-absorbing media ($\epsilon''_1 = 0, \epsilon''_2 = 0$), such that $\det\{\hat{\mathbf{M}}\} = 1$. This explains why the off-diagonal elements of the matrix do not appear in the formula. Moreover, because of $\det\{\hat{\mathbf{M}}\} = 1$, we have the product of eigenvalues $\mu_+ \mu_- = 1$.

The corresponding eigenvectors (field and its derivative at $x = N\Lambda$) can be computed from

$$\begin{aligned} & \{\hat{\mathbf{M}} - \exp(iK\Lambda) \hat{\mathbf{I}}\} \begin{pmatrix} E \\ E' \end{pmatrix}_{N\Lambda} = 0 \\ & \Rightarrow \begin{pmatrix} M_{11} - \mu & M_{12} \\ M_{21} & M_{22} - \mu \end{pmatrix} \begin{pmatrix} E \\ E' \end{pmatrix}_{N\Lambda} = 0. \end{aligned}$$

From the first row, the following condition can be derived:

$$(M_{11} - \mu)E + M_{12}E' = 0$$

Since the investigated system is linear and phase-invariant, the absolute amplitude and phase of the E -component of the eigenvector can be chosen arbitrarily. Here we take $E = 1$ and obtain for the full eigenvector

$$\begin{pmatrix} E \\ E' \end{pmatrix}_{N\Lambda} = \begin{Bmatrix} 1 \\ (\mu - M_{11}) / M_{12} \end{Bmatrix} E_{N\Lambda}.$$

If we seek the values of the field of the Bloch mode inside the layers, i.e. the function $E_{k_x}(x + \Lambda) = E_{k_x}(x)$, they can be computed using the matrix formalism and the above eigenvector $(E, E')_{N\Lambda}$.

Physical properties of infinite multi-layered systems

We are interested in the reflection properties of an infinite Bragg mirror. Based on the electric field E_0 and its derivative E'_0 at the interface ($x=0$), we can express the reflectivity of the Bragg mirror as

$$\rho = \left| \frac{E_R}{E_I} \right|^2 \text{ with } E_R = \frac{E_0}{2} + \frac{iE'_0}{2k_{sx}} \text{ and } E_I = \frac{E_0}{2} - \frac{iE'_0}{2k_{sx}} \text{ from before}$$

$$\rightarrow \rho = \left| \frac{k_{sx}E_0 + iE'_0}{k_{sx}E_0 - iE'_0} \right|^2$$

With our knowledge of the eigenvector from above, we can compute the reflectivity:

$$E'_0 = \frac{\mu - M_{11}}{M_{12}} E_0 \rightarrow \rho = \left| \frac{k_{sx}E_0 + iE'_0}{k_{sx}E_0 - iE'_0} \right|^2 = \left| \frac{k_{sx} + i\frac{\mu - M_{11}}{M_{12}}}{k_{sx} - i\frac{\mu - M_{11}}{M_{12}}} \right|^2$$

According to this formula two scenarios are possible:

A) total internal reflection $\rho = 1$

Hence, μ must be real. From equation (*), the following condition results

$$\rightarrow \left| \frac{(M_{11} + M_{22})}{2} \right| \geq 1$$

with [obtained by matrix multiplication for our example $(\varepsilon_1, \varepsilon_2, d_1, d_2)$]

$$M_{11} = \cos(k_{1x}d_1)\cos(k_{2x}d_2) - \frac{k_{2x}}{k_{1x}}\sin(k_{1x}d_1)\sin(k_{2x}d_2)$$

$$M_{22} = \cos(k_{1x}d_1)\cos(k_{2x}d_2) - \frac{k_{1x}}{k_{2x}}\sin(k_{1x}d_1)\sin(k_{2x}d_2).$$

This defines the so-called **band gap**, i.e. frequencies of excitation for which no propagating solutions exist.

B) propagating normal modes

Hence, μ must be **complex**, which results in the condition

$$\rightarrow \left| \frac{(M_{11} + M_{22})}{2} \right| < 1$$

We can compute a more explicit version of the dispersion relation by starting from

$$\mu = \exp(iK\Lambda) = \frac{(M_{11} + M_{22})}{2} \pm \sqrt{\left\{ \frac{(M_{11} + M_{22})}{2} \right\}^2 - 1}.$$

and splitting the exponential into

$$\mu = \exp(iK\Lambda) = \cos(K\Lambda) + i\sin(K\Lambda) = \cos(K\Lambda) \pm \sqrt{\{\cos(K\Lambda)\}^2 - 1}.$$

Here the plus-minus sign takes into account that the sine function can be positive or negative.

By comparing the two expressions we get

$$\rightarrow \cos\{K(k_z, \omega)\Lambda\} = \frac{(M_{11} + M_{22})}{2}.$$

Thus, the Bloch wave is a solution to Maxwell's equations in infinite periodic media only if the Bloch vector K fulfills this DR. This is in complete analogy to plane waves in homogeneous media with the DR $k_x^2 = \frac{\omega^2}{c^2}\epsilon(\omega) - k_z^2$.

Interpretation

- For the case of total internal reflection (μ real, $\left|\frac{(M_{11} + M_{22})}{2}\right| \geq 1$), the Bloch vector K is imaginary/complex, $\mu = \exp(iK\Lambda) = \exp(i\Re(K)\Lambda)\exp(-\Im(K)\Lambda)$.

Hence $\Re\{K(k_z, \omega)\Lambda\} = n\pi$ and

$$\Im\{K(k_z, \omega)\Lambda\} = -\ln\left((-1)^n \sqrt{\left(\frac{(M_{11} + M_{22})}{2}\right)^2 - 1}\right).$$

simply by taking the logarithm of the above solution for μ .

The \pm accounts for exponentially damped and growing solution, as we would expect in the case of complex wave vectors and evanescent waves.

- There is an infinite number of so-called **band gaps** or **forbidden bands**, because $n=1\dots\infty$ can take any integer values. These band gaps are interesting for Bragg mirrors and Bragg waveguides as they correspond to "forbidden" frequency ranges where no propagating solution exists.
- The limits of the *bands* are given by

$$\Re\{K(k_z, \omega)\Lambda\} = n\pi \text{ and } \Im\{K(k_z, \omega)\Lambda\} = 0$$

$$\curvearrowright K(k_z, \omega) = n\pi / \Lambda$$

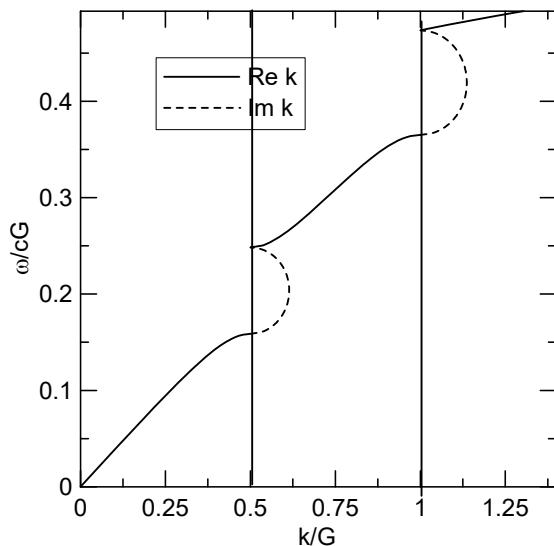
- Outside the band gaps, i.e. inside the bands, we find propagating solutions, which have different properties from the normal modes in homogeneous media (different dispersion relation).
→ We can exploit the strong curvature, i.e. frequency dependence, of the DR for various applications, e.g. dispersion compensation or diffraction free propagation

Special case: normal incidence

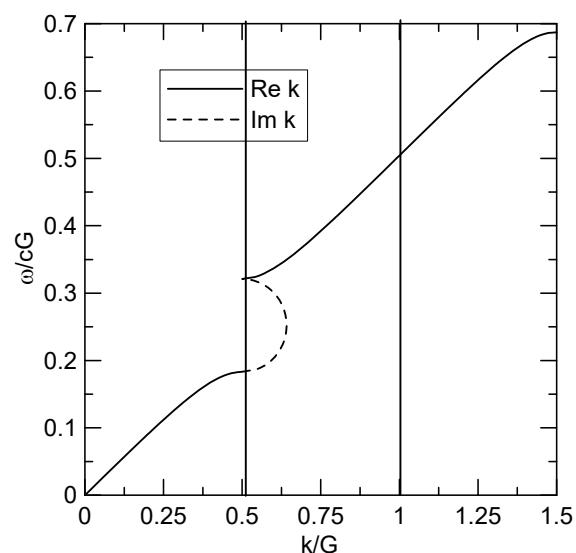
In general, there is a complex interplay between the angle of incidence and the frequency of light which determines the reflection properties of multi-layered systems. Therefore, let us have a look at the simpler case of normal incidence ($k_z = 0$). In a graphical representation of the dispersion relation for $k_z = 0$, it is common to use the following dimensionless quantities

$$\frac{\omega}{cG} \text{ and } \frac{K}{G} \text{ with the scaling constant } G = \frac{2\pi}{\Lambda}$$

Examples for normal incidence:

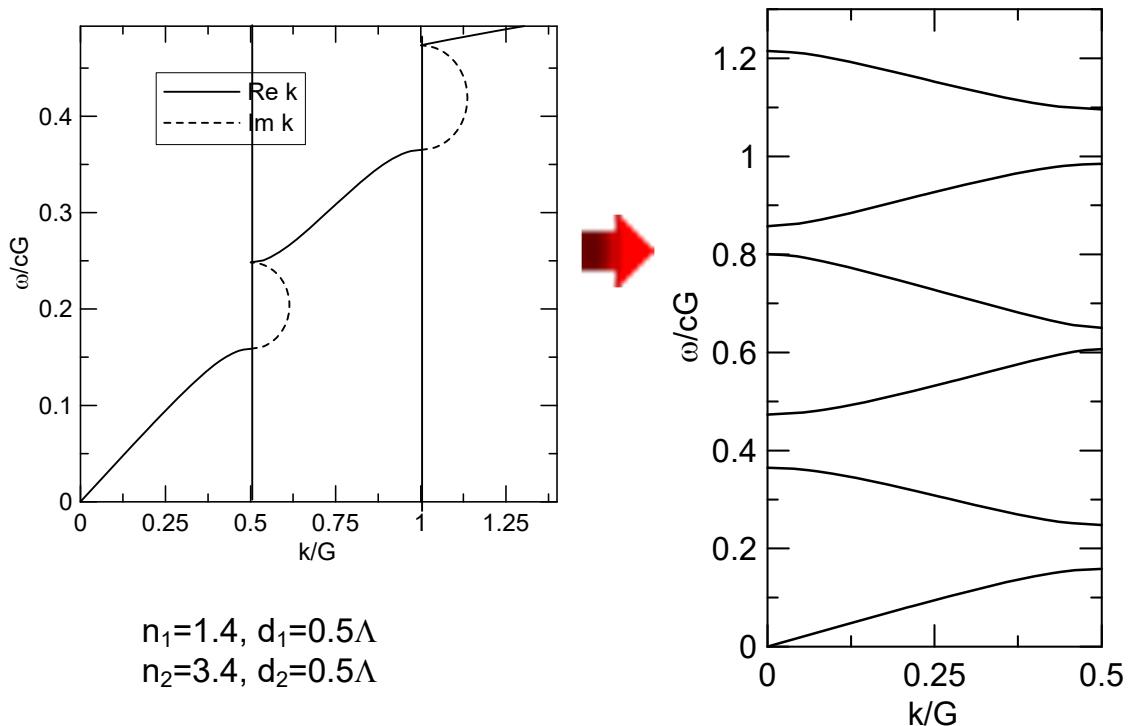


$$\begin{aligned} n_1 &= 1.4, d_1 = 0.5\Lambda \\ n_2 &= 3.4, d_2 = 0.5\Lambda \end{aligned}$$



$$\begin{aligned} n_1 &= 1.4, d_1 = 17/24 \Lambda \\ n_2 &= 3.4, d_2 = 7/24 \Lambda \end{aligned}$$

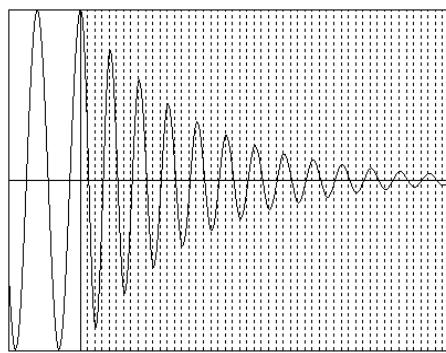
It is common to use the reduced band structure, where the information for all possible Bloch vectors is mapped onto the Bloch vectors in the interval $-0.5 \leq (k / G) \leq 0.5$, which is the so-called first Brillouin zone.



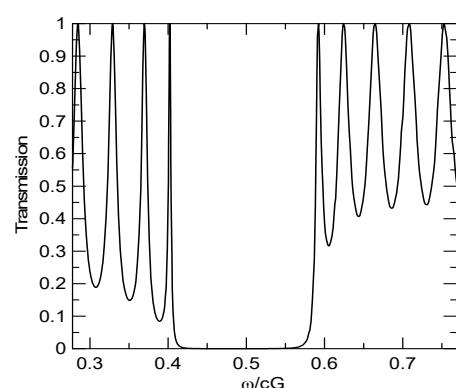
Because of the periodicity of $e^{iK\Lambda}$, it is sufficient to describe the dispersion relation in the first Brillouin zone (one period) of the space of spatial frequencies $\frac{K}{G}$. Thus we need only $-\pi \leq K\Lambda \leq \pi \rightarrow \left| \frac{K}{G} \right| \leq 0.5$ to describe the entire dispersion relation for all spatial frequencies.

Inside the band gap, we find damped solutions:

$$n_0=1.0, n_1=1.4, n_2=1.6, d_1/d_2=1$$



Field in a Bragg-mirror



Transmission

Let us quantify the damping. In our example (n_1, n_2, d_1, d_2) , we have

$$\frac{(M_{11} + M_{22})}{2} = \cos\left(\frac{\omega}{c} n_1 d_1\right) \cos\left(\frac{\omega}{c} n_2 d_2\right) - \frac{1}{2} \left(\frac{n_2}{n_1} + \frac{n_1}{n_2} \right) \sin\left(\frac{\omega}{c} n_1 d_1\right) \sin\left(\frac{\omega}{c} n_2 d_2\right)$$

In the middle of the first band gap (optimal configuration for high reflection), we have $\frac{\omega_B}{c} n_1 d_1 = \frac{\omega_B}{c} n_2 d_2 = \frac{\pi}{2}$, with ω_B being the Bragg frequency, and therefore $\frac{(M_{11} + M_{22})}{2} = -\frac{1}{2} \left(\frac{n_2}{n_1} + \frac{n_1}{n_2} \right) < -1$. If we substitute this (for the first gap $n = 1$) into our expression for $\Im(K)$ and assume a small index contrast $|n_2 - n_1| \ll (n_2 + n_1)$, we find

$$\Lambda \Im(K)_{\max} \approx 2 \frac{n_2 - n_1}{n_2 + n_1} \quad (\text{do derivation as an exercise})$$

→ Damping is proportional to the index contrast of the subsequent layers $|n_2 - n_1|$

The spectral width of the gap $\left(\left| \frac{(M_{11} + M_{22})}{2} \right| \geq 1 \right)$ is then

$$\Delta \omega_{\text{gap}} \approx \frac{2\omega_B}{\pi} \Lambda \Im(K)_{\max} \quad (\text{do derivation as an exercise})$$

↷ → Spectral width is proportional to index contrast $|n_2 - n_1|$ as well.

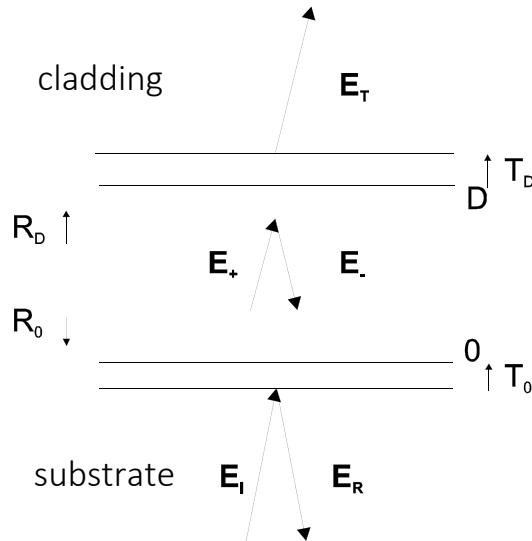
8.3.4 Fabry-Perot resonators

In this chapter we will treat a special multi-layered system, the so-called Fabry-Perot resonator. To construct a Fabry-Perot resonator, one can start with a highly reflecting periodic multi-layered system (Bragg reflector). If one changes just a single layer in the middle of the otherwise periodic layer system, a so-called cavity is formed. We are interested in the forward and backward propagating fields in the entire layered system. While the single layer, which is distinguished from the periodic stack, forms the cavity, the other layers function as mirrors and may be periodic multilayered-systems or metallic films. Fabry-Perot resonators are very important in optics, as they appear in:

- Fabry-Perot interferometers
- lasers with planar mirrors → Fabry-Perot resonators with active media inside the cavities
- nonlinear optics → high intensities inside the cavity → nonlinear optical effects in spite of low-intensity incident light:
 - bistability
 - modulational instability
 - pattern formation, solitons

Here, we want to compute the transmission properties of the resonator for arbitrary planar mirrors. This task could be achieved by employing the matrix method which we developed in the previous sections. However, we will take a different approach to obtain deeper physical insights into the cavity's behavior.

For simplicity, we will restrict ourselves to the TE-polarization. The figure shows our setup with two mirrors at $x=0$ and $x=D$, characterized by coefficients of reflection and transmission R_0 , T_0 , R_D , and T_D .



- E_i, E_r and E_t , \rightarrow amplitudes of incident, reflected and transmitted external fields in substrate and cladding
- E_+, E_- \rightarrow amplitudes of forward and backward propagating internal fields inside the cavity

Using the known coefficients of reflection and transmission of the two mirrors, we can eliminate E_+ and E_- by relating the field amplitudes:

- A) At the **lower mirror inside** the cavity, the forward propagating wave E_+ in the cavity arises due to transmission of the incident wave as well as reflection of the backward propagating wave E_- :

$$T_0 E_i + R_0 E_-(0) = E_+(0).$$

- B) At the **upper mirror outside** the cavity, the transmitted wave is only due to transmission of the forward propagating wave:

$$E_t = T_D E_+(D)$$

Since the forward propagating wave simply evolves as $E_+(D) = E_+(0) \exp(i k_{fx} D)$, we can relate the transmitted field and the forward propagating wave at the lower mirror:

$$E_+(0) = \frac{E_t}{T_D} \exp(-i k_{fx} D).$$

- C) At the **upper mirror inside** the cavity, the backward propagating wave is only due to reflection of the forward propagating wave:

$$E_-(D) = R_D E_+(D)$$

$$\text{with } E_+(D) = \frac{E_t}{T_D} \text{ from B}$$

and with $E_{-}(0) = E_{-}(D) \exp(i k_{fx} D)$

$$E_{-}(0) = \frac{R_D}{T_D} E_T \exp(i k_{fx} D)$$

D) we substitute $E_{+}(0)$ and $E_{-}(0)$ into A) and solve for the incident field

$$T_0 E_I + R_0 \frac{R_D}{T_D} E_T \exp(i k_{fx} D) = \frac{E_T}{T_D} \exp(-i k_{fx} D)$$

$$\rightarrow E_I = \frac{1}{T_0 T_D} \{ \exp(-i k_{fx} D) - R_0 R_D \exp(i k_{fx} D) \} E_T.$$

Thus, the coefficient of transmission for the whole FP resonator is expressed in terms of the coefficients of the mirrors (R_0 , T_0 , R_D , and T_D), the cavity properties and the

angle of incidence (D , $k_{fx} = \sqrt{\frac{\omega^2}{c^2} \epsilon_f - k_z^2}$):

$$T_{TE} = \frac{E_T}{E_I} = \frac{T_0 T_D \exp(i k_{fx} D)}{1 - R_0 R_D \exp(2 i k_{fx} D)}.$$

This is the general transmission function of a lossless Fabry-Perot resonator. In general, the mirror coefficients $(R, T)_{0,D} = (|R|, |T|)_{0,D} \exp(i \phi_{0,D}^{R,T})$ are complex and the fields receive certain phase shifts. These phase shifts induced by the coefficients of reflection R_0, R_D are important in determining the transmissivity of the FP resonator $\tau_{TE} \sim |T_{TE}|^2$.

For given values of $|R_0|, |R_D|$, $|T_0|, |T_D|$ and ϕ_0^R, ϕ_D^R , the general transmissivity of a lossless Fabry-Perot resonator reads:

$$|T_{TE}|^2 = |T|^2 = \frac{|T_0|^2 |T_D|^2}{1 + |R_0|^2 |R_D|^2 - 2 |R_0| |R_D| \underbrace{\cos(2k_{fx} D + \phi_0^R + \phi_D^R)}_{\delta}}$$

$$\tau = \frac{k_{cx}}{k_{sx}} |T|^2.$$

Here we introduce the phase-shift δ which the field acquires in one round-trip of the cavity.

Discussion

Depending on whether the two mirrors have identical properties, we distinguish between symmetric and asymmetric FP-resonators.

a) asymmetric FP-resonator

$$\tau = \frac{k_{cx}}{k_{sx}} \frac{|T_0|^2 |T_D|^2}{1 + |R_0|^2 |R_D|^2 - 2 |R_0| |R_D| \cos \delta}$$

Since we assume no losses to be present, we can use energy conservation (reflectivity plus transmittivity must add to one) at each mirror to eliminate $T_{0,D}$ (recall the relation between the transmission coefficients and transmittivity):

$$\begin{aligned}|T_0|^2 |T_D|^2 &= \frac{k_{sx}}{k_{fx}} (1 - |R_0|^2) \frac{k_{fx}}{k_{cx}} (1 - |R_D|^2) \\&= \frac{k_{sx}}{k_{cx}} (1 - |R_0|^2) (1 - |R_D|^2)\end{aligned}$$

Note: τ and ρ for a lossless mirror are the same for both sides of the mirror. For lossy mirrors, only τ is the same while ρ is side-dependent.

For discussing the effect of the phase shift δ we rewrite

$$\cos \delta = \cos^2 \frac{\delta}{2} - \sin^2 \frac{\delta}{2} = 1 - 2 \sin^2 \frac{\delta}{2}$$

Substituting everything in, we get

$$\begin{aligned}\tau &= \frac{(1 - |R_0|^2)(1 - |R_D|^2)}{1 + |R_0|^2 |R_D|^2 - 2|R_0||R_D|(1 - 2 \sin^2 \frac{\delta}{2})} \\&= \left\{ \frac{(1 - |R_0||R_D|)^2}{(1 - |R_0|^2)(1 - |R_D|^2)} + \frac{4|R_0||R_D|}{(1 - |R_0|^2)(1 - |R_D|^2)} \sin^2 \frac{\delta}{2} \right\}^{-1}.\end{aligned}$$

and with

$$\rho_0 = |R_0|^2, \rho_D = |R_D|^2$$

$$\boxed{\tau = \left\{ \frac{(1 - \sqrt{\rho_0 \rho_D})^2}{(1 - \rho_0)(1 - \rho_D)} + \frac{4\sqrt{\rho_0 \rho_D}}{(1 - \rho_0)(1 - \rho_D)} \sin^2 \frac{\delta}{2} \right\}^{-1}.}$$

b) symmetric FP-resonator (Airy-formula for transmissivity)

Assuming that the two mirrors have identical reflection properties:

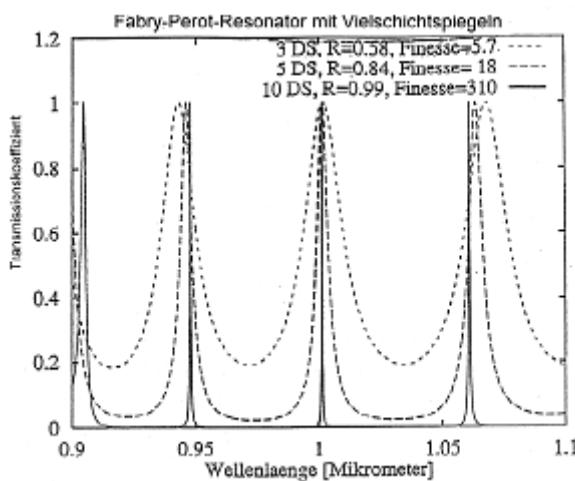
$$\rho_m = |R_0|^2 = |R_D|^2 = \rho_0 = \rho_D, \quad \varphi^R = \varphi_0^R = \varphi_D^R$$

$$\tau = \frac{(1 - \rho_m)^2}{(1 - \rho_m)^2 + 4\rho_m \sin^2 \frac{\delta}{2}} = \left\{ \underbrace{\frac{(1 - \rho_m)^2}{(1 - \rho_m)^2}}_{=1} + \frac{4\rho}{(1 - \rho_m)^2} \sin^2 \frac{\delta}{2} \right\}^{-1}$$

$$\tau = \left\{ 1 + F \sin^2 \frac{\delta}{2} \right\}^{-1} \quad \text{Airy-formula}$$

with $F = \frac{4\rho_m}{(1-\rho_m)^2}$ and $\frac{\delta}{2} = k_{fx} D + \varphi$

The **Airy-formula** (see also the Labworks script, where phase shifts φ due to the mirrors are not considered) gives the transmissivity of a symmetric, lossless Fabry-Perot resonator. Only in this case can we obtain the maximum transmissivity $\tau = 1$ for $\delta/2 = n\pi$.



Remarks and conclusions

- We can perform an analogous calculation for the TM-polarization $\rightarrow R^{\text{TM}}$ resp. ρ^{TM}
- Resonances of the cavity with $\tau_{\text{MAX}} = 1$ occur for $\delta/2 = k_{fx} D_{\text{MAX}} + \varphi = m\pi$, with the normal component of the wave vector in the resonator given by

$$\begin{aligned} k_{fx} &= \frac{2\pi}{\lambda} \sqrt{n_f^2 - n_s^2 \sin^2 \varphi_I} \\ \curvearrowright D_{\text{MAX}} &= \frac{m\pi - \varphi}{k_{fx}} = \frac{\lambda}{2} \frac{\left(m - \frac{\varphi}{\pi}\right)}{\sqrt{n_f^2 - n_s^2 \sin^2 \varphi_I}} \\ \curvearrowright \frac{\lambda}{2} &\text{ cavity} \end{aligned}$$

where φ_I is the angle of incidence in the substrate

- transmission properties of a given resonator depend on φ_I and λ .
- minimum transmission is given as $\tau_{\text{MIN}} = \frac{1}{1+F}$
- it is favorable to have large F , e.g.

$$100 = F = \frac{4\rho_m}{(1-\rho_m)^2}$$

$$\rho_m = 1 - \tau_m \curvearrowright \frac{4(1-\tau_m)}{\tau_m^2} \approx \frac{4}{\tau_m^2} \approx 100 \curvearrowright \tau_m = 0.2 \curvearrowright \rho_m = 0.8.$$

- pulses and beams can be treated efficiently by multiplication with the transfer function in the Fourier domain:
→ e.g. TE: $E_{\text{TE}}(\alpha, \beta, \omega) = T_{\text{TE}}(\alpha, \beta, \omega)E_{\text{I}}(\alpha, \beta, \omega)$
Fourier back transformation: $E_{\text{I}}(x, y, t) = \text{FT}^{-1}[E_{\text{TE}}(\alpha, \beta, \omega)]$
- since beams always contain a certain range of incident angles φ_{I} , they produce interference rings in the far-field output (or image of lens, like in Labworks).
- a quantity often used to characterize a resonator is the finesse:

$$\Phi = \frac{\text{distance between resonance}}{\text{full width at half maximum of resonance}} = \frac{\Delta}{\varepsilon}$$

where ε is the FWHM of the resonances in radians and $\Delta = \pi$ is the distance in radians between two successive resonances.

To calculate the FWHM ε , we set the expression for the transmittivity (in terms of the phase accumulated in the resonator) equal to:

$$\left\{ 1 + F \sin^2 \left(m\pi \pm \frac{\varepsilon}{2} \right) \right\}^{-1} \doteq \frac{1}{2}$$

For narrow resonances (small line width ε), we can make the small-angle approximation

$$\left\{ 1 + F \left(\frac{\varepsilon}{2} \right)^2 \right\}^{-1} \approx \frac{1}{2} \rightarrow F \left(\frac{\varepsilon}{2} \right)^2 = 1 \curvearrowright \varepsilon = \frac{2}{\sqrt{F}}$$

$$\boxed{\Phi = \frac{\Delta}{\varepsilon} = \frac{\pi}{2} \sqrt{F} = \pi \frac{\sqrt{\rho_m}}{1 - \rho_m}}$$

- The line width ε (FWHM) is inversely proportional to the finesse Φ .
- The Airy-formula can be expressed in terms of the finesse

$$\boxed{\tau = \left\{ 1 + \left(\frac{2\Phi}{\pi} \right)^2 \sin^2 \frac{\delta}{2} \right\}^{-1}}$$

- A Fabry-Perot resonator can be used as a spectroscope. We can then calculate its resolution (assuming normal incidence).

Resonances (maximum transmission) occur at:

$$kD + \varphi = m\pi$$

FWHM of the transmission in terms of the wavenumber:

$$\rightarrow kD + \varphi \pm \frac{\Delta k}{2} D = m\pi \pm \frac{\varepsilon}{2} = m\pi \pm \frac{\pi}{2\Phi}$$

$$\curvearrowright |\Delta k| = \frac{2\pi}{\lambda^2} n_f |\Delta \lambda| = \frac{\pi}{\Phi D}$$

$$\curvearrowright \left| \frac{\Delta \lambda}{\lambda} \right| = \frac{\lambda}{n_f \Phi D} \sim \frac{\lambda}{\Phi D} \sim \lambda \frac{\varepsilon}{D}$$

where we have used the definition of the **finesse** Φ .

Example: $\lambda = 5 \cdot 10^{-7} \text{ m}$, $\Phi = 30$, $n_f D = 4 \cdot 10^{-3} \text{ m}$ $\curvearrowright \Delta \lambda = 2 \cdot 10^{-12} \text{ m}$

- The field amplitudes (here forward field) inside the cavity are given by $E_T = T_D E_+(D)$. Because $|T|^{-2} \sim 1/(1-\rho) \sim \Phi$, these intra-cavity fields can be very high \curvearrowright important for nonlinear effects
- Lifetime of photons in cavity: from the "uncertainty relation": $\Delta \omega T_c = \text{const.} \approx 1$. It is then possible to define a lifetime of photons inside the cavity by expressing the resolution in terms of the frequency:

$$|\Delta k| = \frac{1}{c} n_f \Delta \omega = \frac{\pi}{\Phi D}$$

$$\Delta \omega = \frac{c}{n_f \Phi D} \quad \curvearrowright \quad T_c = \frac{1}{\Delta \omega} = \frac{n_f \Phi D}{c \pi} \sim \frac{D}{\varepsilon}.$$

8.4 Guided waves in layer systems

Finally, we want to explore our layered systems as waveguides. For many applications, it is interesting to have waves which propagate **without diffraction**. This is crucial for **integrated optics**, where we want to guide light in very small (micrometric or smaller) dielectric layers (film, fiber), or **optical communication technology**, where information encoded in light is transported over long distances. Moreover, waveguides are important in nonlinear optics, where nonlinear effects become important due to the strong confinement of light over long propagation distances (since then there is no diffractive spreading as in free space). Here, we will treat waveguides in one dimension since we restrict ourselves to layer systems, but the general concepts developed here can be transferred to other settings with two dimensional confinement, e.g. where the waveguide is a fiber.

8.4.1 Field structure of guided waves

Let us first look at some general consideration about the field structure of guided waves. We want to find guided waves in a layered system. In such systems, we have solved the reflection and transmission problem: for given $k_z, \mathbf{E}_T, d_i, \epsilon_i$, we calculated $\mathbf{E}_R, \mathbf{E}_T$.

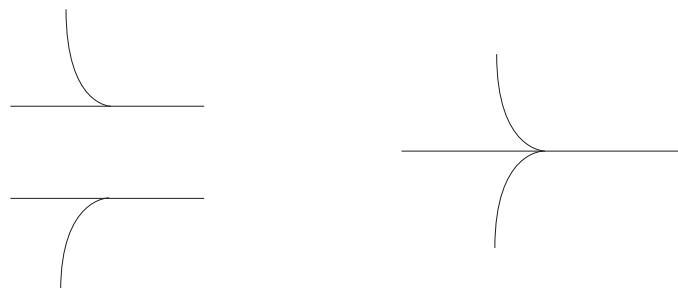
Inside each layer, we have plane waves

$$\sim \mathbf{E}_\alpha(k_z, \omega) \exp[i(k_{\alpha x}x + k_z z - \omega t)].$$

The question is: how can we trap (or guide) waves within a finite layered system? A possible hint is given by the effect of **total internal reflection (TIR)**, where the transmitted field at the interface of the multi-layered system to the cladding is:

$$\mathbf{E}_T(x, z) = \mathbf{E}_T \exp(ik_z z) \exp(-\mu_c x).$$

Obviously, in the case of TIR we have no energy flux in the cladding medium. If TIR is the key mechanism for guiding light, we can have TIR on two sides, i.e. to the cladding and to the substrate. Perhaps we can also have a similar process at a single interface.



We will first concentrate our discussion on a system of layers, which we will call the core of the waveguide, and which we place between a semi-infinite substrate and cladding. The single interface, where the core is absent, will be considered at a later stage.

According to our discussion, the guided waves should have the following field structure:

- plane wave in the propagation direction $\sim \exp(ik_z z)$
- evanescent waves in substrate and cladding

$$\sim \exp[-\mu_c(x - D)] \quad \text{cladding}$$

$$\sim \exp(\mu_s x) \quad \text{substrate}$$

$$\text{with } \mu_{s,c} = \sqrt{k_z^2 - \frac{\omega^2}{c^2} \epsilon_{s,c}(\omega)} > 0$$

$$\curvearrowright \text{Condition 1: } k_z^2 > \frac{\omega^2}{c^2} \max \{ \epsilon_{s,c}(\omega) \}$$

- oscillating solution (standing wave) in the core (layers, fiber core, film)

$$\sim A \sin(k_{fx} x) + B \cos(k_{fx} x)$$

$$\text{with } k_{fx} = \sqrt{\frac{\omega^2}{c^2} \epsilon_i(\omega) - k_z^2} > 0$$

$$\curvearrowright \text{Condition 2: } k_z^2 < \frac{\omega^2}{c^2} \max_i \{ \epsilon_i(\omega) \}$$

Note that this second condition is not obvious at this stage. It appears due to boundaries conditions at the interfaces with substrate and cladding. The fact that the fields in the substrate and cladding are evanescent requires that both fields are exponentially decreasing away from their respective interfaces with the core. The boundary conditions at the interfaces impose the continuity of the tangential components of the fields and their first derivatives normal to the surface. This fact was shown during the derivation of the transmission and reflection properties in a multi-layered system. Since the field/derivatives of the fields in the core must continuously connect with the fields/derivatives of the fields in the substrate and cladding, it must have a non-monotonic profile. This condition excludes evanescent field profiles in the core and leaves only oscillating (radiation) fields as a possible solution in the core of a waveguide. For waveguide cores consisting of multiple layers, the field must be of radiating nature in at least one of the core layers.

In summary, the z -component of the wave vector of guided waves must fulfill:

$$\max\left(\frac{\omega}{c}n_{s,c}\right) < k_z < \max_i\left(\frac{\omega}{c}n_i\right)$$

In summary, the field structure in the substrate and cladding is given by:

$$\begin{aligned} \mathbf{E}_c(x,z) &= \mathbf{E}_t \exp(i k_z z) \exp[-\mu_c(x - D)] & x > D \\ \mathbf{E}_s(x,z) &= \mathbf{E}_R \exp(i k_z z) \exp(\mu_s x) & x < 0 \end{aligned}$$

and in the core by

$$\mathbf{E}_{fi}(x,z) = \mathbf{E}_{fi} \exp(i k_z z) \exp(i k_{xi} x) \quad D \geq x \geq 0$$

8.4.2 Dispersion relation for guided waves

If we compare the shapes of the fields for the guided wave to our usual reflection/transmission problem, we see that for guided waves the reflected and transmitted (evanescent) field exist even for zero incident field. Please note that $E_t, E_R \neq 0$ even though there is no energy transport connected to these fields due to their evanescent natures. In the following, we will use this condition ($E_t, E_R \neq 0$ for $E_I \rightarrow 0$) to derive the dispersion relation for guided waves.

Consequently the coefficients for reflection and transmission are

$$T^{TE,TM} = \left(\frac{E_t}{E_I} \right)^{TE,TM}, \quad R^{TE,TM} = \left(\frac{E_R}{E_I} \right)^{TE,TM} \quad \text{with } E_I \rightarrow 0$$

Therefore we must have $R, T \rightarrow \infty$ in the case of guided waves. In this sense, guided waves can be considered as **resonances** of the multi-layered system. To appreciate this concept, let us compare with a driven harmonic oscillator

$$x = \frac{F}{\omega^2 - \omega_0^2}, \quad x = \text{effect}, F = \text{cause}$$

In the case of resonance ($\omega = \omega_0$) we observe an effect for an infinitesimal cause. Hence, we can obtain the dispersion relation of guided waves by looking for a vanishing denominator in the expressions for R, T . This reasoning is a general principle in physics: the **poles of the response function** (or Green's function) are the **resonances** of the system. Furthermore, a resonance of a system is equivalent to an eigenmode of the system, which is localized in our case.

From before, we have the coefficients of reflection and transmission for a layered system, which have the same denominator:

$$R = \frac{F_R}{F_I} = \frac{(\alpha_s k_{sx} M_{22} - \alpha_c k_{cx} M_{11}) - i(M_{21} + \alpha_s k_{sx} \alpha_c k_{cx} M_{12})}{(\alpha_s k_{sx} M_{22} + \alpha_c k_{cx} M_{11}) + i(M_{21} - \alpha_s k_{sx} \alpha_c k_{cx} M_{12})}$$

The pole is then given by:

$$\rightarrow (\alpha_s k_{sx} M_{22} + \alpha_c k_{cx} M_{11}) + i(M_{21} - \alpha_s k_{sx} \alpha_c k_{cx} M_{12}) \doteq 0$$

Because we have evanescent waves in the substrate and the cladding, the imaginary valued k_{sx} and k_{cx} can be expressed as

$$k_{sx} = i\mu_s = i\sqrt{k_z^2 - \frac{\omega^2}{c^2}\epsilon_s(\omega)}, \quad k_{cx} = i\mu_c = i\sqrt{k_z^2 - \frac{\omega^2}{c^2}\epsilon_c(\omega)}$$

Now we can write the **general dispersion relation of guided modes** in an **arbitrary layered system** as

$$M_{11}^{\text{TE,TM}} + \alpha_s \mu_s M_{12}^{\text{TE,TM}} + \frac{1}{\alpha_c \mu_c} M_{21}^{\text{TE,TM}} + \frac{\alpha_s \mu_s}{\alpha_c \mu_c} M_{22}^{\text{TE,TM}} \doteq 0$$

Here, as before, we have $\alpha_{\text{TE}} = 1$, $\alpha_{\text{TM}} = 1/\epsilon$ for the two independent polarizations.

This waveguide dispersion relation gives a discrete set of solutions, which are the so-called waveguide **modes**. For given ϵ_i, d_i, ω , we obtain $k_{zv}(\omega)$.

We can see that the dispersion relation of guided modes depends on the material's dispersion in the individual layers as well as in the substrate and the cladding as given by the dispersion relations which would apply in homogeneous space

$$k_i^2(\omega) = \frac{\omega^2}{c^2} \epsilon_i(\omega) = k_{ix}^2 + k_z^2.$$

However, in addition to the material's dispersion in the layers, the dispersion relation of guided modes depends on the geometry of the waveguide's layered system

$$k_z(\omega, \text{geometry})$$

In the case of guided waves it is easy to compute the field (mode profile) inside the layered system:

- take k_z from dispersion relation
- in the substrate we have:

$$F(x) = F \exp(\mu_s x), \quad G(x) = \alpha \frac{\partial}{\partial x} \{ F \exp(\mu_s x) \}$$

Hence, for a given $F(0)$ we obtain $G(0) = \alpha \mu_s F(0)$ ($F(0) \rightarrow$ free parameter)

$$\begin{pmatrix} F \\ G \end{pmatrix}_x = \hat{\mathbf{M}}(x) \begin{pmatrix} F \\ G \end{pmatrix}_0 = \hat{\mathbf{M}}(x) \begin{pmatrix} 1 \\ \alpha \mu_s \end{pmatrix} F(0).$$

Analogy of optics to the stationary Schrödinger equation in QM

Optics (e.g. TE-polarization)

$$\left\{ \frac{d^2}{dx^2} + \frac{\omega^2}{c^2} \epsilon(x) \right\} E(x) = k_z^2 E(x) \quad \leftrightarrow$$

guided waves

$$k_z^2 > \frac{\omega^2}{c^2} \max \{ \epsilon_{s,c} \} \quad \leftrightarrow$$

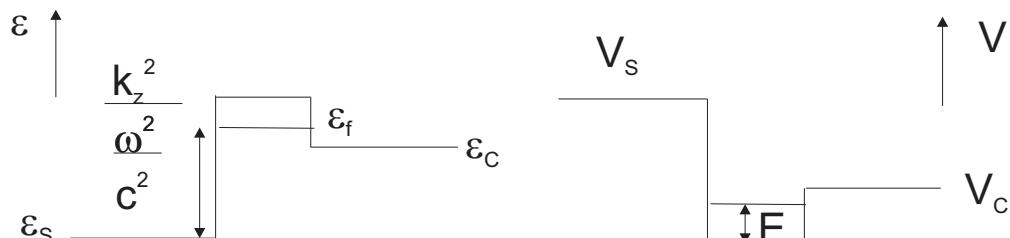
Quantum mechanics

$$\left\{ \frac{d^2}{dx^2} - \frac{2m}{\hbar^2} V(x) \right\} \psi(x) = -\frac{2m}{\hbar^2} E \psi(x)$$

discrete set of energy eigenfunctions

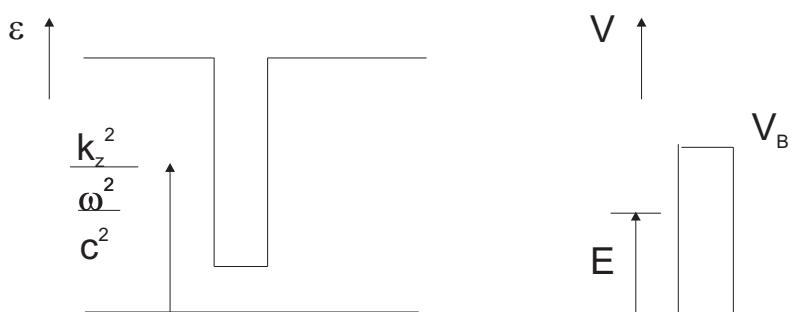
$$E < V_{\text{ext}}$$

Note that in this analogy, a high dielectric constant plays the role of an attractive potential in quantum mechanics. This is to be expected since light tends to travel more slowly/spend more time in regions of higher refractive index, just as an electron tends to be found near the potential minimum.



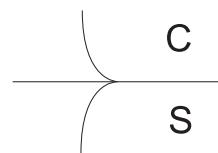
Tunnel effect

$$k_z^2 > \frac{\omega^2}{c^2} \epsilon_{\text{Film}} \quad \leftrightarrow \quad E < V_{\text{Barriere}}$$



8.4.3 Guided waves at interface - surface polariton

Let us first have a look at the most simple case where the guiding layer structure is just an interface.



Our condition for guided waves is that on both sides of the interface we have evanescent waves: $k_z^2 > \frac{\omega^2}{c^2} \epsilon_{c,s}$, because

$$\mu_{s,c} = \sqrt{k_z^2 - \frac{\omega^2}{c^2} \epsilon_{c,s}} > 0$$

The general dispersion relation we derived before reads

$$M_{11}^{\text{TE,TM}} + \alpha_s \mu_s M_{12}^{\text{TE,TM}} + \frac{1}{\alpha_c \mu_c} M_{21}^{\text{TE,TM}} + \frac{\alpha_s \mu_s}{\alpha_c \mu_c} M_{22}^{\text{TE,TM}} \doteq 0$$

and with the matrix for a single interface: $\hat{\mathbf{M}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

we get the dispersion relation $1 + \frac{\alpha_s \mu_s}{\alpha_c \mu_c} = 0$

A) TE-polarization ($\alpha = 1$)

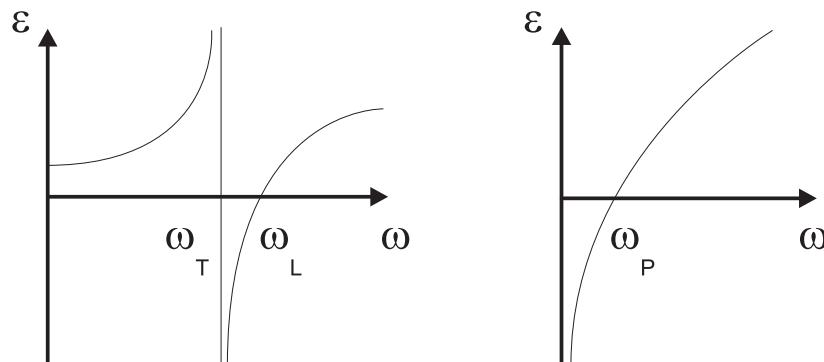
$$\mu_s + \mu_c = 0, \rightarrow \text{no solution because } \mu_c, \mu_s > 0$$

B) TM-polarization ($\alpha = 1/\epsilon$)

$$\frac{\mu_c}{\epsilon_c} + \frac{\mu_s}{\epsilon_s} = 0$$

with $\mu_c, \mu_s > 0 \curvearrowright \epsilon_c \cdot \epsilon_s < 0$,

\rightarrow one of the media has to have negative ϵ (dielectric near resonance or metal)



In dielectrics $\omega_{0(\text{r})} < \omega < \omega_L$ we can find surface-phonon-polaritons.

In metals $\omega < \omega_P$ we can find surface-plasmon-polaritons.

Remark

Surface polaritons occur in TM polarization only, similar to the phenomenon of Brewster-angle (no reflection for $\frac{k_{cx}}{\epsilon_c} - \frac{k_{sx}}{\epsilon_s} = 0$,)

Let us now compute the explicit dispersion relation for surface polaritons. W.o.l.g. we assume $\epsilon_s(\omega) < 0$, and because $\epsilon_s(\omega)$ is near a resonance it will show a much stronger ω dependence than ϵ_c .

$$(\mu_c \epsilon_s)^2 = (\mu_s \epsilon_c)^2$$

$$\epsilon_s^2(\omega) \left\{ k_z^2 - \frac{\omega^2}{c^2} \epsilon_c^2 \right\} = \epsilon_c^2 \left\{ k_z^2 - \frac{\omega^2}{c^2} \epsilon_s(\omega) \right\}$$

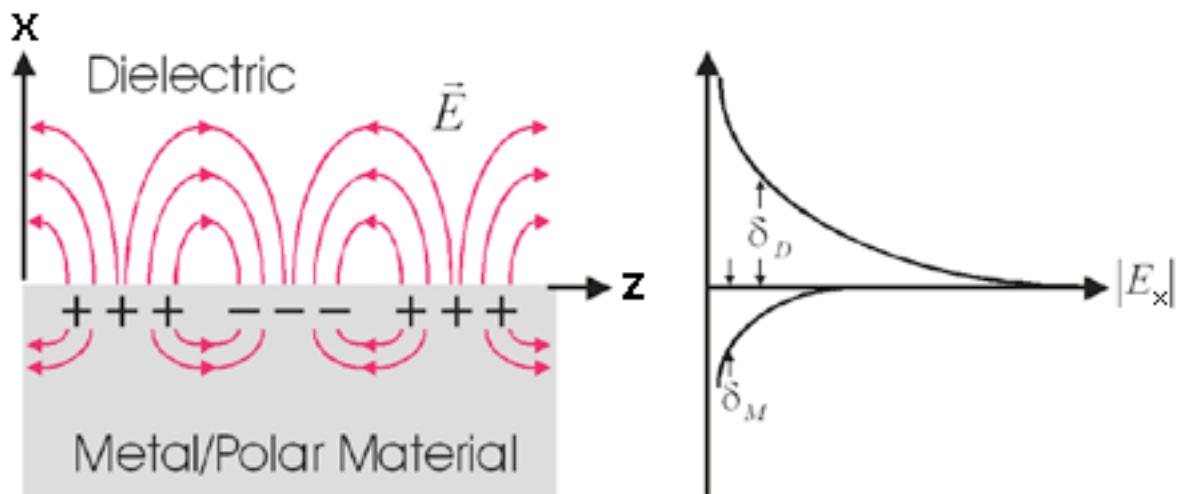
$$k_z(\omega) = \frac{\omega}{c} \sqrt{\frac{\epsilon_s(\omega) \epsilon_c}{\epsilon_c + \epsilon_s(\omega)}}$$

There is a second condition for existence of surface polaritons: $\epsilon_c + \epsilon_s(\omega) < 0$

Conclusion:

$$k_z(\omega) = \frac{\omega}{c} \sqrt{\frac{\epsilon_s(\omega) \epsilon_c}{\epsilon_s(\omega) + \epsilon_c}}$$

$$\text{TM polarization} \quad \epsilon_s(\omega) < 0 \quad \epsilon_c > 0 \quad |\epsilon_s(\omega)| > \epsilon_c$$

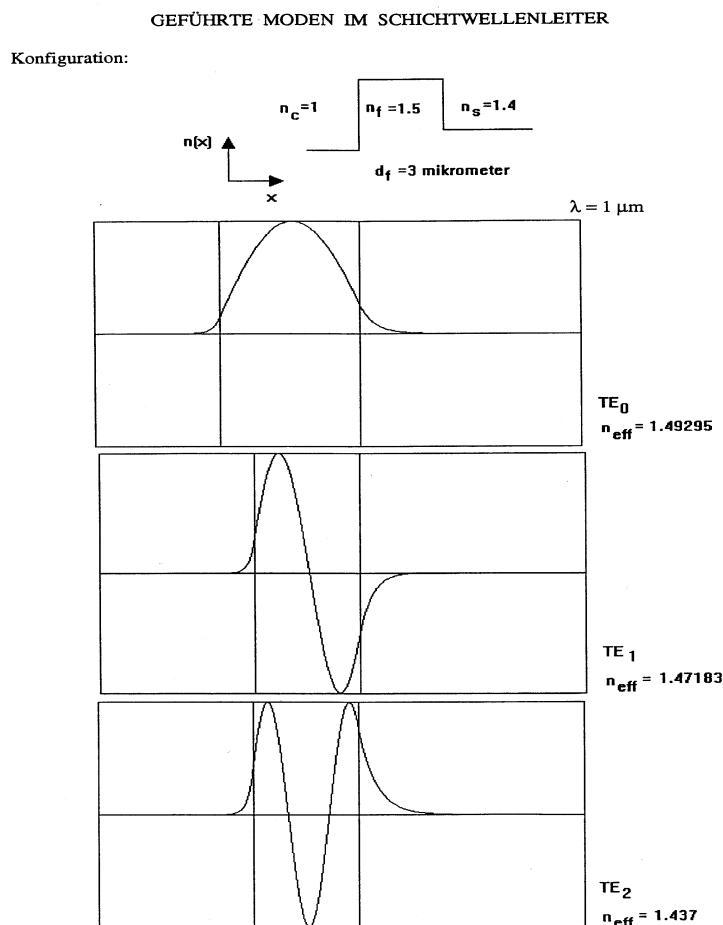


Surface polaritons may have very small effective wavelengths in z-direction:

$$\epsilon_c = 1, |\epsilon_s(\omega)| \approx 1 \rightarrow k_z(\omega) = \frac{2\pi}{\lambda} \sqrt{\frac{\epsilon_s(\omega) \epsilon_c}{\epsilon_s(\omega) + \epsilon_c}} = \frac{2\pi}{\lambda_{\text{eff}}} \rightarrow \lambda_{\text{eff}} \ll \lambda$$

8.4.4 Guided waves in a layer – film waveguide

The prototype of a waveguide is the film waveguide, where the waveguide consists of one guiding layer with $\frac{\omega^2}{c^2} \epsilon_f(\omega) > k_z^2(\omega)$.



Such film waveguides are the basis of integrated optics. Typical parameters are:

$$d \approx \text{a few wavelengths}$$

$$\Delta\epsilon = \epsilon_f - \epsilon_s \approx 10^{-3} - 10^{-1}$$

Fabrication of film waveguides can be achieved by coating, diffusion or ion implantation.

The matrix of a single layer (film) is given as:

$$\hat{\mathbf{M}}^{\text{TE,TM}} = \hat{\mathbf{m}}^{\text{TE,TM}}(d) = \begin{pmatrix} \cos(k_{fx}d) & \frac{1}{\alpha_f k_{fx}} \sin(k_{fx}d) \\ -\alpha_f k_{fx} \sin(k_{fx}d) & \cos(k_{fx}d) \end{pmatrix}$$

From this matrix we can compute the dispersion relation for guided modes:

$$M_{11} + \alpha_s \mu_s M_{12} + \frac{1}{\alpha_c \mu_c} M_{21} + \frac{\alpha_s \mu_s}{\alpha_c \mu_c} M_{22} \doteq 0$$

$$\cos(k_{fx}d) + \frac{\alpha_s \mu_s}{\alpha_f k_{fx}} \sin(k_{fx}d) - \frac{\alpha_f k_{fx}}{\alpha_c \mu_c} \sin(k_{fx}d) + \frac{\alpha_s \mu_s}{\alpha_c \mu_c} \cos(k_{fx}d) = 0$$

$$\frac{\sin(k_{fx}d)}{\cos(k_{fx}d)} = \tan(k_{fx}d) = \frac{1 + \frac{\alpha_s \mu_s}{\alpha_c \mu_c}}{\frac{\alpha_f k_{fx}}{\alpha_c \mu_c} - \frac{\alpha_s \mu_s}{\alpha_f k_{fx}}} = \frac{\alpha_f k_{fx} (\alpha_s \mu_s + \alpha_c \mu_c)}{\alpha_f^2 k_{fx}^2 - \alpha_c \alpha_s \mu_c \mu_s}$$

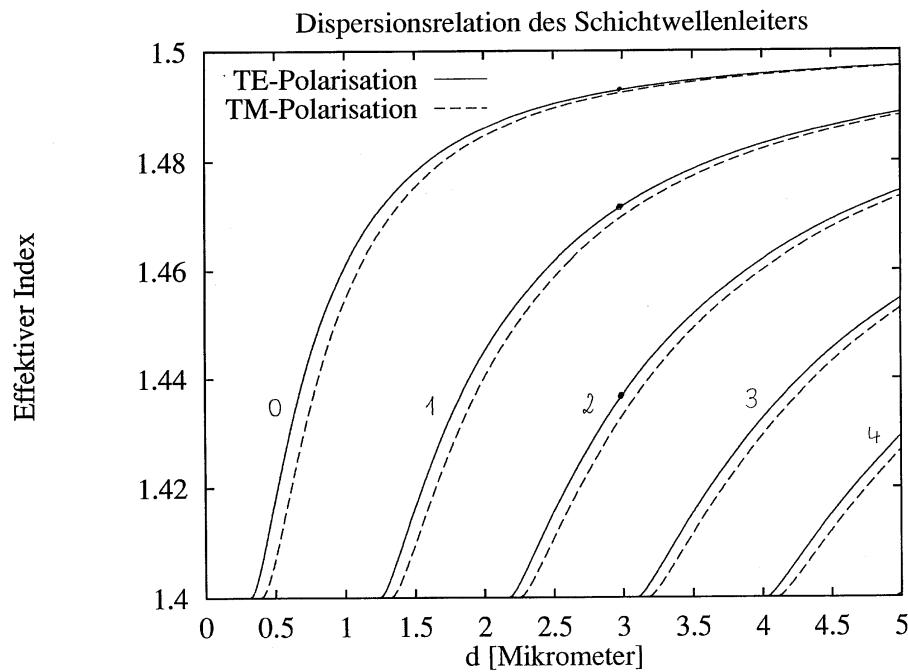
$$\boxed{\tan(k_{fx}d) = \frac{\frac{k_{fx}}{\alpha_f} \left(\frac{\mu_s}{\alpha_c} + \frac{\mu_c}{\alpha_s} \right)}{\frac{k_{fx}^2}{\alpha_c \alpha_s} - \frac{\mu_c \mu_s}{\alpha_f^2}}}$$

Here: TE-Polarisation ($\alpha = 1$)

$$\boxed{\tan(k_{fx}d) = \frac{k_{fx}(\mu_s + \mu_c)}{k_{fx}^2 - \mu_c \mu_s},}$$

This waveguide dispersion relation is an implicit equation for k_z . For given frequency ω and thickness d we get several solutions with index k_{zv}

Here is an example for fixed frequency ω , the effective index $n_{\text{eff}} = k_z / \left(\frac{\omega}{c} \right)$ versus the thickness d :



We can see in the figure that for large thickness d we have many modes. If we decrease d , more and more modes vanish at a certain **cut-off** thickness.

Definition of cut-off:

a guided mode vanishes → **cut-off** (here w.o.l.g. $\epsilon_c < \epsilon_s$)

The idea of the cut-off is that a mode is not guided anymore. Guiding means evanescent fields in the substrate and cladding, so cut-off means

$$\mu_s = \sqrt{k_z^2 - \frac{\omega^2}{c^2} \epsilon_s} = 0 \rightarrow \text{no guiding} \curvearrowright k_z^2 = \frac{\omega^2}{c^2} \epsilon_s$$

We can plug this cut-off condition in the DR: $\tan(k_{fx}d) = \frac{k_{fx}(\mu_s + \mu_c)}{k_{fx}^2 - \mu_c \mu_s}$,

$$\tan\left(\frac{\omega}{c} \sqrt{\epsilon_f - \epsilon_s} d\right) = \frac{\sqrt{\epsilon_f - \epsilon_s} \sqrt{\epsilon_s - \epsilon_c}}{\epsilon_f - \epsilon_s} = \sqrt{\frac{\epsilon_s - \epsilon_c}{\epsilon_f - \epsilon_s}}$$

$$(\omega d)_{co}^{TE} = \frac{c}{\sqrt{\epsilon_f - \epsilon_s}} \left\{ \arctan \sqrt{\frac{\epsilon_s - \epsilon_c}{\epsilon_f - \epsilon_s}} + v\pi \right\}$$

$$\rightarrow (\omega d)_{co}^{TE} = \frac{c}{\sqrt{\epsilon_f - \epsilon_s}} \left\{ \underbrace{\arctan a}_{\max \pi/2} + v\pi \right\}$$

$$\epsilon_s \approx \epsilon_c \quad a \rightarrow 0$$

with parameter of asymmetry a : $\epsilon_s \approx \epsilon_f \quad a \rightarrow \infty$

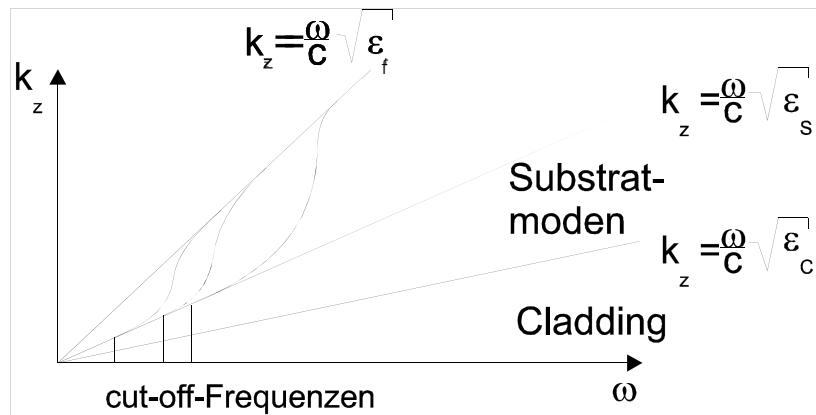
$$\epsilon_s \approx \epsilon_c$$

→ we can define a **cut-off frequency** for $k_z(\omega)$ when we keep d fix

→ we can define a **cut-off thickness** for $k_z(d)$ when we keep ω fix

In a **symmetric** waveguide the fundamental mode ($v = 0$) has **cut-off = 0!**

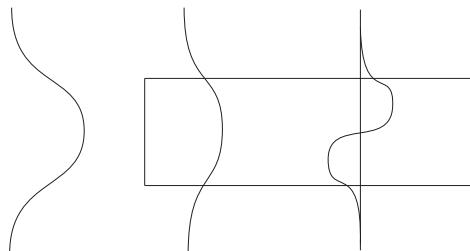
If we plot the dispersion curves for each mode we get a graphical representation of the dispersion relation:



8.4.5 Excitation of guided waves

Finally, we want to address the question how we can excite guided waves. In principle, there are two possibilities, we can adapt the field profile or the wave vector (k_z)

A) adaption of field → front face coupling



Then, inside the waveguide we have (without radiative modes):

$$E(x, z) = \sum_v a_v E_v(x) \exp(i k_{zv} z)$$

$$\leadsto E(x, 0) \approx \sum_v a_v E_v(x) \quad \left| \int E_\mu(x)$$

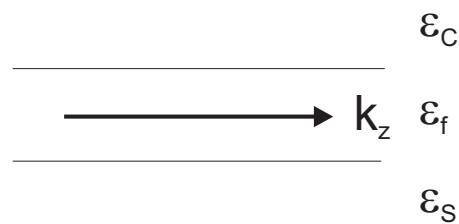
with: $P_v = \frac{k_{zv}}{2\omega\mu_0} \int_{-\infty}^{\infty} |E_v(x)|^2 dx$

$$a_v = \frac{k_{zv}}{2\omega\mu_0 P_v} \int_{-\infty}^{\infty} E_{in}(x) E_v(x) dx,$$

→ mode v couples to the incident field $E_{in}(0)$ with amplitude a_v .

→ Gauss-beam couples very good to the fundamental mode

B) adaption of wave vector → coupling through the interface



We know that k_z is continuous at interface. The condition for the existence of guided modes is

$$k_z > \frac{\omega}{c} \sqrt{\epsilon_{c,s}}$$

but dispersion relation for waves in bulk media dictates

$$k_z = \sqrt{\frac{\omega^2}{c^2} \epsilon_{c,s} - k_{s,cx}^2} < \frac{\omega}{c} \sqrt{\epsilon_{c,s}}$$

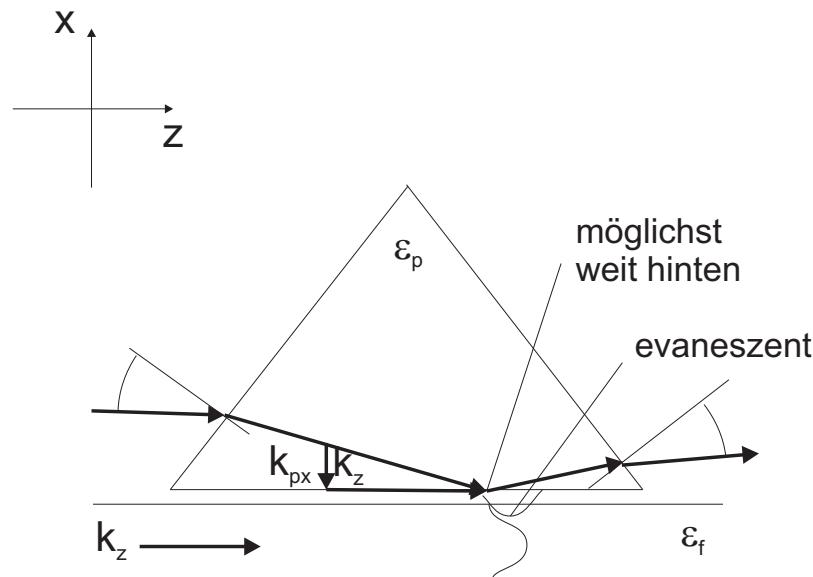
↷ We got a problem!

There are two solutions:

i) coupling by prism

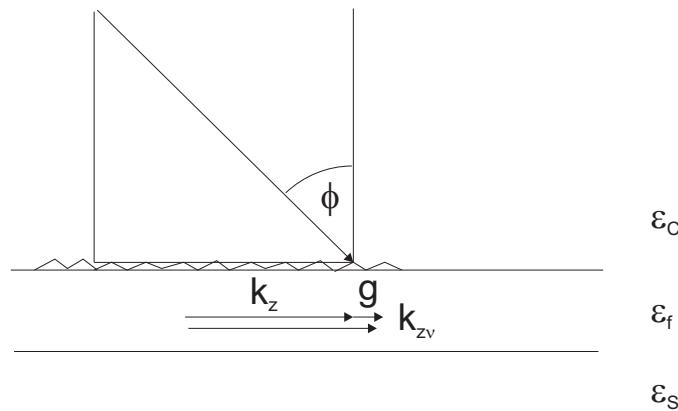
we bring a medium with $\epsilon_p > \epsilon_f$ (prism) near the waveguide

$$\sim k_z < \frac{\omega}{c} \sqrt{\epsilon_f} \sim k_z < \frac{\omega}{c} \sqrt{\epsilon_p} \sim k_{px} = \sqrt{\frac{\omega^2}{c^2} \epsilon_p - k_z^2} > 0.$$



→ light can couple to the waveguide via optical tunneling: ATR ('attenuated total reflection')

ii) coupling by grating



grating on waveguide (modulated thickness of layer d):

$$d(z) = d + \zeta(z)$$

$$\zeta(z) = A \sin(gz) \quad \text{mit} \quad g = \frac{2\pi}{P} \dots_{P-period}$$

coupling works for m'th diffraction order:

$$k_{zv} = k_z + mg$$

$$= \frac{\omega}{c} n_s \sin \phi + mg.$$

9. Ray optics- geometrical optics (covered by lecture Introduction to Optical Modeling)

The topic of “Ray optics – geometrical optics” is not covered in the course “Fundamentals of modern optics”. This topic will be covered rather by the course “Introduction to optical modeling”. The following part of the script, which is devoted to this topic, is just included in the script for consistency.

9.1 Introduction

- Ray optics or geometrical optics is the simplest theory for doing optics.
- In this theory, propagation of light in various optical media can be described by simple geometrical rules.
- Ray optics is based on a very rough approximation ($\lambda \rightarrow 0$, no wave phenomena), but we can explain almost all daily life experiences involving light (shadows, mirrors, etc.).
- In particular, we can describe *optical imaging* with ray optics approach.
- In isotropic media, the direction of rays corresponds to the direction of energy flow.

What is covered in this chapter?

- It gives fundamental postulates of the theory.
- It derives simple rules for propagation of light (rays).
- It introduces simple optical components.
- It introduces light propagation in inhomogeneous media (graded-index (GRIN) optics).
- It introduces paraxial matrix optics.

9.2 Postulates

- A) Light propagates as rays. Those rays are emitted by light-sources and are observable by optical detectors.
- B) The optical medium is characterized by a function $n(r)$, the so-called refractive index ($n(r) \geq 1$ - meta-materials $n(r) < 0$)

$$n = \frac{c}{c_n} \quad c_n - \text{speed of light in the medium}$$

- C) optical path length ~ delay
 - i) homogeneous media

nl

- ii) inhomogeneous media

$$\int_A^B n(\mathbf{r}) ds$$

D) Fermat's principle

$$\delta \int_A^B n(\mathbf{r}) ds = 0$$

Rays of light choose the optical path with the shortest delay.

9.3 Simple rules for propagation of light

A) Homogeneous media

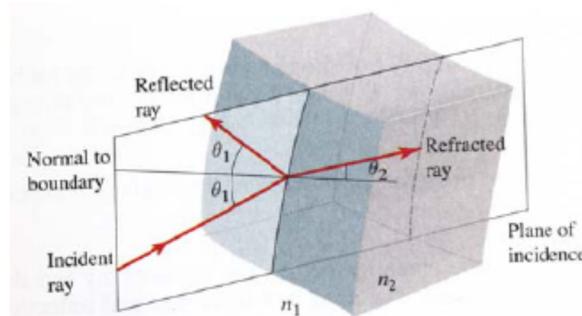
- $n = \text{const.} \rightarrow \text{minimum delay} = \text{minimum distance}$
- Rays of light propagate on straight lines.

B) Reflection by a mirror (metal, dielectric coating)

- The reflected ray lies in the plane of incidence.
- The angle of reflection equals the angle of incidence.

C) Reflection and refraction by an interface

- Incident ray \rightarrow reflected ray plus refracted ray
- The reflected ray obeys b).
- The refracted ray lies in the plane of incidence.



- The angle of refraction θ_2 depends on the angle of incidence θ_1 and is given by Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

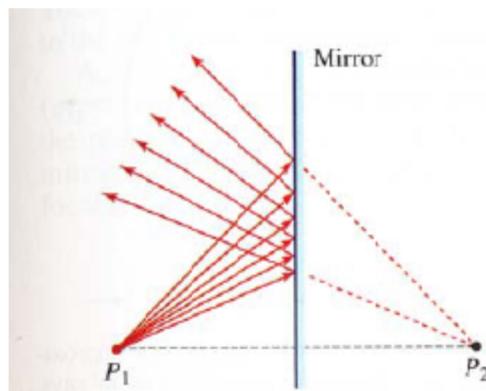
- no information about amplitude ratio.

9.4 Simple optical components

A) Mirror

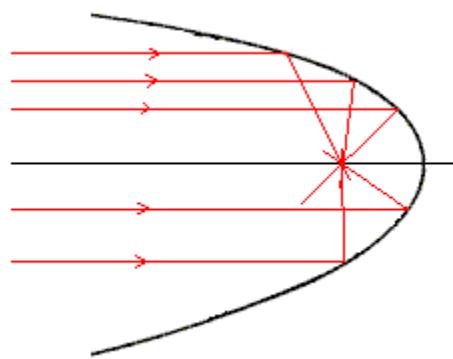
i) Planar mirror

- Rays originating from P_1 are reflected and seem to originate from P_2 .



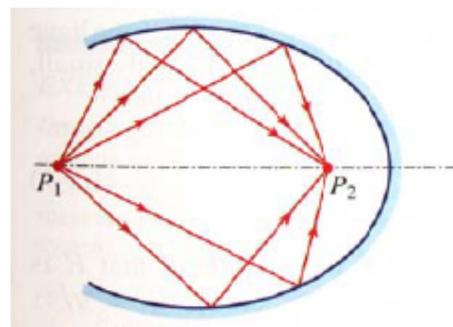
ii) Parabolic mirror

- Parallel rays converge in the focal point (focal length f).
- Applications: Telescope, collimator



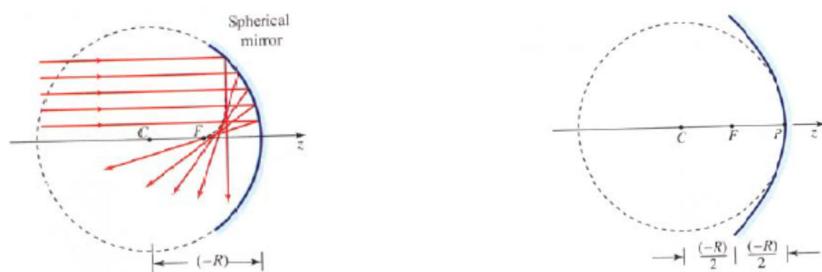
iii) Elliptic mirror

- Rays originating from focal point P_1 converge in the second focal point P_2



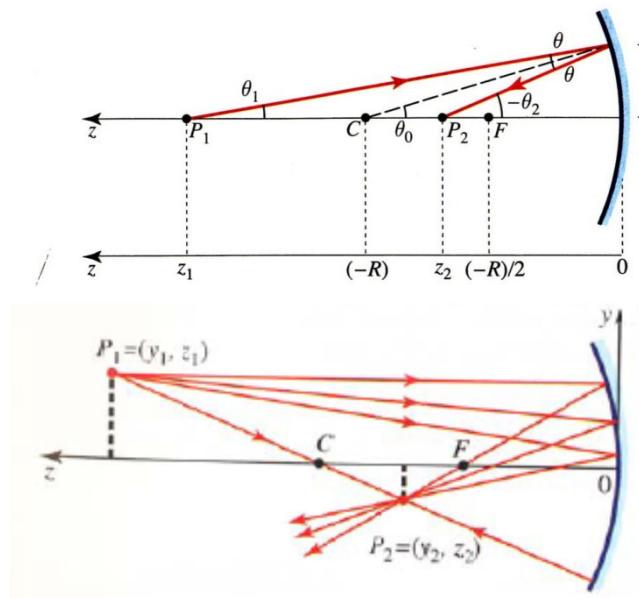
iv) Spherical mirror

- Neither imaging like elliptical mirror nor focusing like parabolic mirror
- parallel rays cross the optical axis at different points
- connecting line of intersections of rays → *caustic*



- parallel, paraxial rays converge to the focal point $f = (-R)/2$

- convention: $R < 0$ - concave mirror; $R > 0$ - convex mirror.
- for paraxial rays the spherical mirror acts as a focusing as well as an imaging optical element. paraxial rays emitted in point P_1 are reflected and converge in point P_2



$$\frac{1}{z_1} + \frac{1}{z_2} \approx \frac{2}{(-R)} \quad (\text{imaging formula})$$

paraxial imaging: imaging formula and magnification

$$m = -z_2/z_1 \quad (\text{proof given in exercises})$$

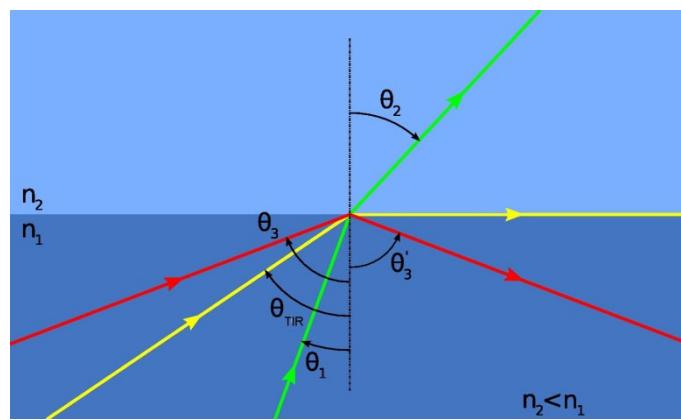
B) Planar interface

Snell's law: $n_1 \sin \theta_1 = n_2 \sin \theta_2$

for paraxial rays: $n_1 \theta_1 = n_2 \theta_2$

- external reflection ($n_1 < n_2$): ray refracted away from the interface
- internal reflection ($n_1 > n_2$): ray refracted towards the interface
- total internal reflection (TIR) for:

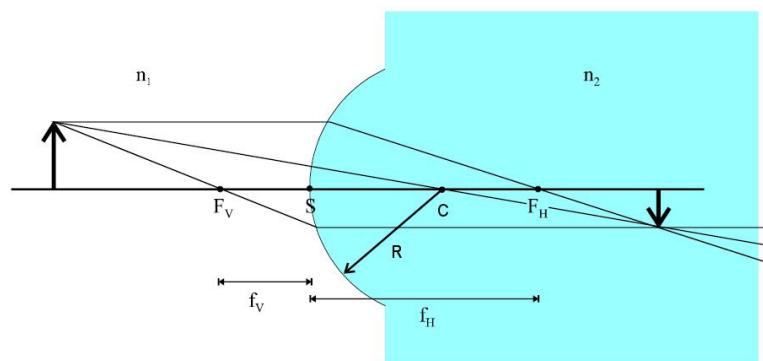
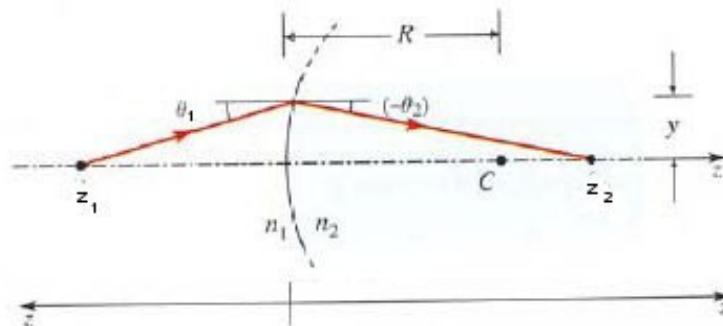
$$\theta_2 = \frac{\pi}{2} \rightarrow \sin \theta_1 = \sin \theta_{\text{TIR}} = \frac{n_2}{n_1}$$



C) Spherical interface (paraxial)

- paraxial imaging

$$\theta_2 \approx \frac{n_1}{n_2} \theta_1 - \frac{n_2 - n_1}{n_2} \frac{y}{R} (*)$$



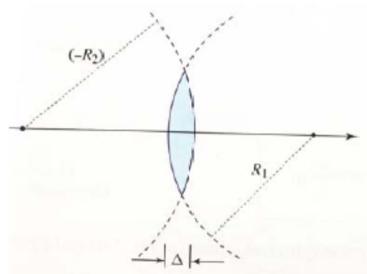
$$\frac{n_1}{z_1} + \frac{n_2}{z_2} \approx \frac{n_2 - n_1}{R} \quad (\text{imaging formula})$$

$$m = -\frac{n_1}{n_2} \frac{z_2}{z_1} \quad (\text{magnification})$$

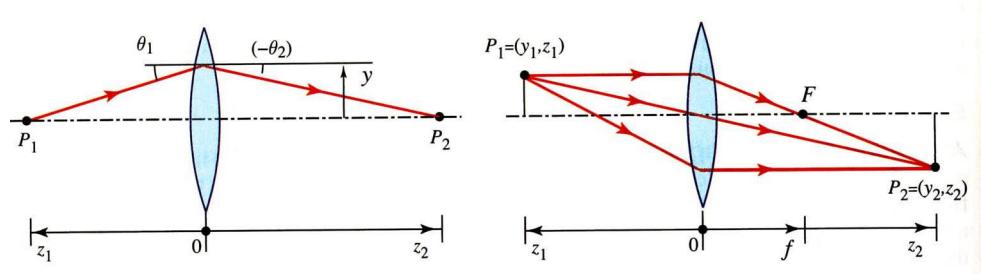
(Proof: exercise)

- if paraxiality is violated \rightarrow aberration
- rays coming from one point of the object do not intersect in one point of the image (caustic)

D) Spherical thin lens (paraxial)



- two spherical interfaces (R_1, R_2, Δ) apply (*) two times and assume $y=\text{const}$ (Δ small)



$$\theta_2 \approx \theta_1 - \frac{y}{f} \text{ with focal length: } \frac{1}{f} = (n-1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$$

$$\frac{1}{z_1} + \frac{1}{z_2} \approx \frac{1}{f} \text{ (imaging formula)} \quad m = -\frac{z_2}{z_1} \text{ (magnification)}$$

(compare to spherical mirror)

9.5 Ray tracing in inhomogeneous media (graded-index - GRIN optics)

- $n(\mathbf{r})$ - continuous function, fabricated by, e.g., doping
- curved trajectories → graded-index layer can act as, e.g., a lens

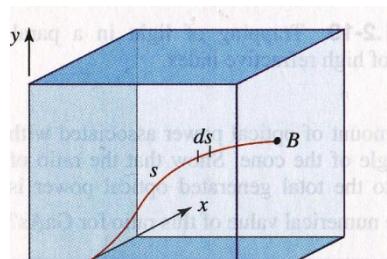
9.5.1 Ray equation

Starting point: we minimize the optical path or the delay (Fermat)

$$\delta \int_A^B n(\mathbf{r}) ds = 0$$

computation:

$$L = \int_A^B n[\mathbf{r}(s)] ds$$



variation of the path: $\mathbf{r}(s) + \delta\mathbf{r}(s)$

$$\delta L = \int_A^B \delta n ds + \int_A^B n \delta ds$$

$$\delta n = \text{grad } n \cdot \delta \mathbf{r}$$

$$\begin{aligned} \delta ds &= \sqrt{(d\mathbf{r} + d\delta\mathbf{r})^2} - \sqrt{(d\mathbf{r})^2} \\ &= \sqrt{(d\mathbf{r})^2 + 2d\mathbf{r} \cdot d\delta\mathbf{r} + (d\delta\mathbf{r})^2} - \sqrt{(d\mathbf{r})^2} \\ &\approx ds \sqrt{1 + 2 \frac{d\mathbf{r}}{ds} \cdot \frac{d\delta\mathbf{r}}{ds}} - ds \\ &\approx ds \left(1 + \frac{d\mathbf{r}}{ds} \cdot \frac{d\delta\mathbf{r}}{ds} \right) - ds \\ &= ds \frac{d\mathbf{r}}{ds} \cdot \frac{d\delta\mathbf{r}}{ds} \end{aligned}$$

$$\begin{aligned} \delta L &= \int_A^B \left(\text{grad } n \cdot \delta \mathbf{r} + n \frac{d\mathbf{r}}{ds} \cdot \frac{d\delta\mathbf{r}}{ds} \right) ds \\ &= \int_A^B \left(\text{grad } n - \frac{d}{ds} \left(n \frac{d\mathbf{r}}{ds} \right) \right) \cdot \delta \mathbf{r} ds \end{aligned} \quad \text{integration by parts and A,B fix}$$

$\delta L = 0$ for arbitrary variation

$$\text{grad } n = \frac{d}{ds} \left(n \frac{d\mathbf{r}}{ds} \right)$$

ray equation

Possible solutions:

A) trajectory

$$x(z), y(z) \text{ and } ds = dz \sqrt{1 + (dx/dz)^2 + (dy/dz)^2}$$

- solve for $x(z), y(z)$
- paraxial rays $\rightarrow (ds \approx dz)$

$$\frac{d}{dz} \left[n(x, y, z) \frac{dx}{dz} \right] \approx \frac{dn}{dx}$$

$$\frac{d}{dz} \left[n(x, y, z) \frac{dy}{dz} \right] \approx \frac{dn}{dy}$$

- B) homogeneous media
– straight lines
- C) graded-index layer $n(y)$ - paraxial, SELFOC

$$\text{paraxial} \rightarrow \frac{dy}{dz} \ll 1 \text{ and } dz \approx ds$$

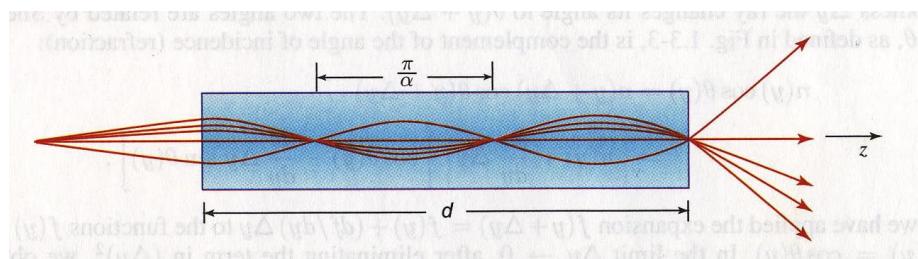
$$n^2(y) = n_0^2 \left(1 - \alpha^2 y^2\right) \Rightarrow n(y) \approx n_0 \left(1 - \frac{1}{2} \alpha^2 y^2\right) \text{ for } \alpha \ll 1$$

$$\frac{d}{ds} \left[n(y) \frac{dy}{ds} \right] \approx \frac{d}{dz} \left[n(y) \frac{dy}{dz} \right] \approx n(y) \frac{d^2 y}{dz^2} \Rightarrow \frac{d^2 y}{dz^2} = \frac{1}{n(y)} \frac{dn(y)}{dy}$$

for $n(y) - n_0 \ll 1$: $\frac{d^2 y}{dz^2} = -\alpha^2 y \quad \rightarrow$

$$y(z) = y_0 \cos \alpha z + \frac{\theta_0}{\alpha} \sin \alpha z$$

$$\theta(z) = \frac{dy}{dz} = -y_0 \alpha \sin \alpha z + \theta_0 \cos \alpha z$$



9.5.2 Eikonal equation

- bridge between geometrical optics and wave
- eikonal $S(\mathbf{r}) = \text{constant} \rightarrow$ planes perpendicular to rays
- from $S(\mathbf{r})$ we can determine direction of rays $\sim \text{grad } S(\mathbf{r})$ (like potential)

$$[\text{grad } S(\mathbf{r})]^2 = n(\mathbf{r})^2$$

Remark: it is possible to derive Fermat's principle from eikonal equation

- geometrical optics: Fermat's or eikonal equation

$$S(\mathbf{r}_B) - S(\mathbf{r}_A) = \int_A^B |\text{grad } S(\mathbf{r})| ds = \int_A^B n(\mathbf{r}) ds$$

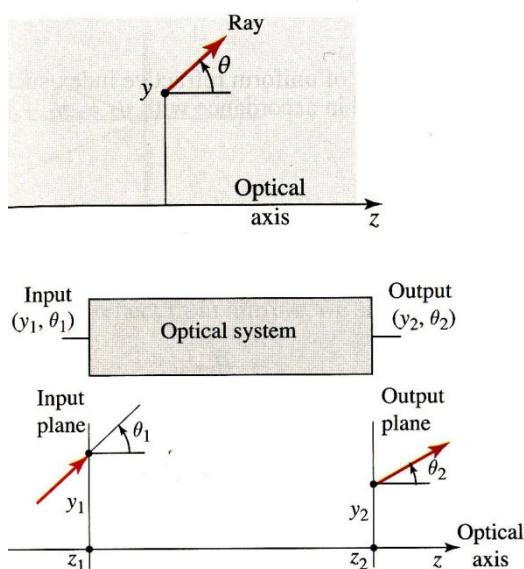
eikonal \rightarrow optical path length \sim phase of the wave

9.6 Matrix optics

- technique for paraxial ray tracing through optical systems
- propagation in a single plane only
- rays are characterized by the distance to the optical axis (y) and their inclination (θ) → two algebraic equations → 2×2 matrix

Advantage: we can trace a ray through an optical system of many elements by multiplication of matrices.

9.6.1 Ray-transfer-matrix



in paraxial approximation:

$$\begin{aligned} y_2 &= Ay_1 + B\theta_1 \\ \theta_2 &= Cy_1 + D\theta_1 \end{aligned} \rightarrow \begin{bmatrix} y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \rightarrow \mathbf{M} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

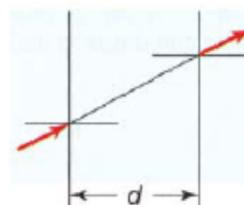
$A=0$: same θ_1 → same y_2 → focusing

$D=0$: same y_1 → same θ_2 → collimation

9.6.2 Matrices of optical elements

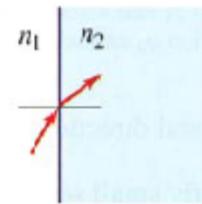
A) free space

$$\mathbf{M} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$$



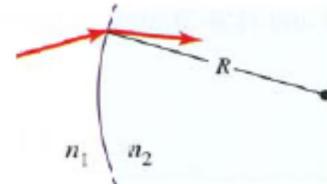
B) refraction on planar interface

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix}$$



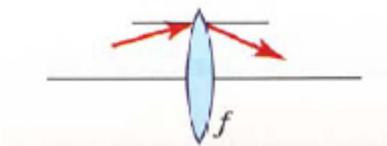
C) refraction on spherical interface

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -(n_2 - n_1)/n_2 R & n_1/n_2 \end{bmatrix}$$



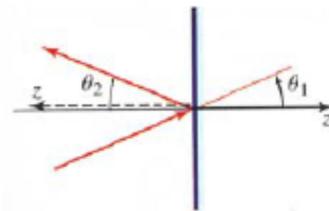
D) thin lens

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}$$



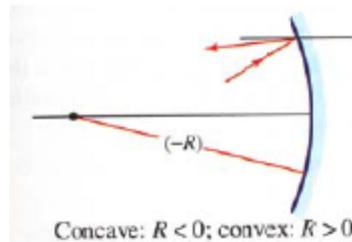
E) reflection on planar mirror

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



F) reflection on spherical mirror (compare to lens)

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 2/R & 1 \end{bmatrix}$$



9.6.3 Cascaded elements

$$\begin{bmatrix} y_{N+1} \\ \theta_{N+1} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} y_1 \\ \theta_1 \end{bmatrix} \rightarrow \mathbf{M} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad \mathbf{M} = \mathbf{M}_N \dots \mathbf{M}_2 \mathbf{M}_1$$