Script for the lecture on

Theoretical Optics

Institute of Theoretical Solid State Physics
Institute of Nanotechnology
Karlsruhe Institute of Technology

Prof. Dr. Carsten Rockstuhl
Carsten.Rockstuhl@kit.edu
http://www.tfp.kit.edu

Please note, this is a pre-release version and is likely incomplete. Feedback and suggestions on what to improve are highly welcomed.

Version: 4. April 2022

I would like to mention that the script and the course are partially based on lectures on theoretical optics that have been given by ***Prof. F. Lederer*** at the Friedrich-Schiller-Universität Jena. Parts of the present script, i.e. those that deal with quantum optics, have been typed by ***Erik Hebestreit*** based on a script I have developed for a lecture on "Quantum Optics".

Suggested literature:
"*Principles of Optics*" by E. Wolf (referential but sometimes too complete)
"*Fundamentals of Photonics*" by B. E. A. Saleh and M. C. Teich (excellent overview over the basics)
"*Classical Electrodynamics*" by J. D. Jackson
"*Theoretical Optics: An Introduction*" by H. Römer
"*Introduction to Fourier Optics*" by J. W. Goodman (chapter on diffraction)
"*Introduction to the Theory of Coherence and Polarization of Light*" by E. Wolf (chapter on coherence properties of light)
"*The Quantum Theory of Light*" by R. Loudon (chapter on quantum optics)

1. Review of Electromagnetism

1.1 Maxwell's equations

1.1.1 Maxwell's equations in optics

The starting point for all our future considerations are Maxwell's equations in all their beauty. They provide the theoretical foundation for our lecture on theoretical optics. Details are discussed in previous courses such as *"Fundamentals of Optics and Photonics"* from Prof. Kalt and *"Electromagnetics and Numerical Calculation of Fields"* from Prof. Dössel. Maxwell's equations in time and space domain read as

$$\mathbf{rot}\,\mathbf{E}(\mathbf{r},t) = -\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t} \qquad \mathbf{div}\,\mathbf{D}(\mathbf{r},t) = \rho_{\text{ext}}(\mathbf{r},t)$$

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

The quantities whose functional dependencies are governed by those equations are the electric field $\mathbf{E}(\mathbf{r},t)$ given in units of $\mathrm{V/m}$, the magnetic field $\mathbf{B}(\mathbf{r},t)$ given in units of $\mathrm{Vs/m^2}$ or Tesla. In addition we have the free charge density $\rho_{\text{ext}}(\mathbf{r},t)$ in units of $\mathrm{As/m^3}$ and the macroscopic current density $\mathbf{j}_{\text{macr}}(\mathbf{r},t)$ in units of $\mathrm{A/m^2}$. Furthermore we have two auxiliary fields commonly known as the (electric) displacement field $\mathbf{D}(\mathbf{r},t)$ given in units of $\mathrm{As/m^2}$ and the magnetizing or the H-field $\mathbf{H}(\mathbf{r},t)$ given in units of $\mathrm{A/m}$. To solve those equations, we have to define the functional dependency that links the electric and the magnetic field to their respective auxiliary fields. These functional dependencies are known as the

$$\mathbf{D}(\mathbf{r},t) = \varepsilon_0\,\mathbf{E}(\mathbf{r},t) + \mathbf{P}(\mathbf{r},t)$$

$$\mathbf{H}(\mathbf{r},t) = \frac{1}{\mu_0}[\,\mathbf{B}(\mathbf{r},t) - \mathbf{M}(\mathbf{r},t)]$$

constitutive relations:

Here, the polarization $\mathbf{P}(\mathbf{r},t)$ is in units of $\mathrm{As/m^2}$ and the magnetization $\mathbf{M}(\mathbf{r},t)$ is in units of $\mathrm{Vs/m^2}$. They reflect the impact of matter. The two natural constants, which appear in these equations, are the electric constant (vacuum permittivity) that is

given by $\varepsilon_0 = \frac{1}{\mu_0 c^2} \approx 8.854 \times 10^{-12}\mathrm{As/Vm}$ and the magnetic constant (vacuum permeability) that is given by $\mu_0 = 4\pi \times 10^{-7}\mathrm{Vs/Am}$. The polarization and the magnetization, in principle, are functions that depend on the electric and magnetic field. These dependencies are called constitutive relations. While neglecting effects due to electro-magnetic coupling that would suggest that an electric field can induce a magnetization and vice versa, we can write the constitutive relations generically as

$$\mathbf{P}(\mathbf{r},t) = f[\mathbf{E}(\mathbf{r},t)] \text{ and } \mathbf{M}(\mathbf{r},t) = f[\mathbf{B}(\mathbf{r},t)]$$

In optics, we usually deal with media that does not have a magnetic response. Therefore, we can safely assume that $\mathbf{M}(\mathbf{r},t) = 0$ holds.

Please note, there are exceptions to these assumptions particularly in the contemporary field of research on metamaterials. There, it is principally possible that $\mathbf{M}(\mathbf{r},t) \neq 0$. Eventually, in the field of metamaterials more complicated functional dependencies can be perceived where the polarization depends not just on the electric field but in addition also on the magnetic field. The same holds for the magnetization that may also depend on the electric field. These constitutive relations are called bi-anisotropic.

Sources of electromagnetic radiation:

- free charge density $\rho_{\mathrm{ext}}(\mathbf{r},t)$
- macroscopic current density: $\mathbf{j}_{\mathrm{makr}}(\mathbf{r},t) = \mathbf{j}_{\mathrm{cond}}(\mathbf{r},t) + \mathbf{j}_{\mathrm{conv}}(\mathbf{r},t)$
- conductive current density: $\mathbf{j}_{\mathrm{cond}}(\mathbf{r},t) = f[\mathbf{E}(\mathbf{r},t)]$
- convective current density: $\mathbf{j}_{\mathrm{conv}}(\mathbf{r},t) = \rho_{\mathrm{ext}}(\mathbf{r},t)\mathbf{v}$

In optics we do, generally, not have free charges:

$$\rho_{\mathrm{ext}}(\mathbf{r},t) = 0 \rightarrow \mathbf{j}_{\mathrm{conv}}(\mathbf{r},t) = 0 \rightarrow \mathbf{j}_{\mathrm{cond}}(\mathbf{r},t) \equiv \mathbf{j}(\mathbf{r},t)$$

Maxwell's equations in the field of optics in the time domain therefore read as

$$\mathbf{rot}\,\mathbf{E}(\mathbf{r},t) = -\mu_0 \frac{\partial \mathbf{H}(\mathbf{r},t)}{\partial t} \qquad \varepsilon_0 \mathbf{div}\,\mathbf{E}(\mathbf{r},t) = -\mathbf{div}\,\mathbf{P}(\mathbf{r},t)$$

$$\mathbf{rot}\,\mathbf{H}(\mathbf{r},t) = \mathbf{j}(\mathbf{r},t) + \frac{\partial \mathbf{P}(\mathbf{r},t)}{\partial t} + \varepsilon_0 \frac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t} \qquad \mathbf{div}\,\mathbf{H}(\mathbf{r},t) = 0$$

1.1.2 Temporal dependencies of the fields

A) The simplest situation concerns monochromatic, stationary fields.

- They have a time harmonic dependency according to $\mathrm{e}^{-i\omega_0 t}$→fixed phase relation, perfect coherence, infinite extent of the wave train

  $\mathbf{E}(\mathbf{r},t) = \bar{\mathbf{E}}(\mathbf{r},\omega_0)\mathrm{e}^{-i\omega_0 t}$
- Please note, this is an artificial assumption since an infinite wave will carry infinite energy that is not feasible, the assumption, however, works very well in most situations that are important to us.

B) Polychromatic, non-stationary fields

- They are described as a fully coherent finite or an infinite number of different frequencies contribute to form the signal.
- To treat such systems, we perform a Fourier decomposition of an arbitrary field that allows a description in terms of infinite time harmonic wave trains.

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

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

  Please note, if you look into the literature you may find different definitions for the Fourier transform. This concerns the way the pre-factor occurs in the equations but also the sign for the exponential function is chosen differently (a different sign convention often coincides with the use of j as the imaginary unit).

Finally, we deal in this part of the course with perfectly coherent light, i.e. it can be described in all its properties (amplitude, phase, frequency for each spectral component) precisely. Many sources of light do not obey these requirements but require a partially coherent or an incoherent treatment. This will be reflected in a devoted section in this course.

1.1.3 Maxwell's equations in frequency space

The transition from time to frequency domain is done by Fourier transforming the equations and herewith all quantities of interest. The partial derivative that occur in Maxwell's equations turn out to be algebraic expressions:

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

$$= \frac{1}{2\pi}\,\mathbf{E}(\mathbf{r},t)\mathrm{e}^{i\omega t}\Big|_{t=-\infty}^{t=\infty} - i\omega\frac{1}{2\pi}\int_{-\infty}^{\infty} dt\mathbf{E}(\mathbf{r},t)\mathrm{e}^{i\omega t}$$

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

Therefore, in Maxwell's equations the transition from time to frequency domain is made by the following replacement:

$$\frac{\partial}{\partial t} \overset{FT}{\rightarrow} -i\omega$$

$$\mathbf{rot}\,\bar{\mathbf{E}}(\mathbf{r},\omega) = i\omega\mu_0\bar{\mathbf{H}}(\mathbf{r},\omega) \qquad \varepsilon_0\mathbf{div}\,\bar{\mathbf{E}}(\mathbf{r},\omega) = -\mathbf{div}\,\bar{\mathbf{P}}(\mathbf{r},\omega)$$

$$\mathbf{rot}\,\bar{\mathbf{H}}(\mathbf{r},\omega) = \bar{\mathbf{j}}(\mathbf{r},\omega) - i\omega\bar{\mathbf{P}}(\mathbf{r},\omega) - i\omega\varepsilon_0\bar{\mathbf{E}}(\mathbf{r},\omega) \qquad \mathbf{div}\,\bar{\mathbf{H}}(\mathbf{r},\omega) = 0$$

Please note the subtle but important difference in the notation. Quantities in frequency domain are highlighted here by a bar across the quantity. At some point of the course I will drop this to make the expressions more light.

1.1.4 Wave equation

A) Time domain

$$\mathbf{rot}\,\mathbf{rot}\,\mathbf{E}(\mathbf{r},t) = -\mu_0\mathbf{rot}\frac{\partial\mathbf{H}(\mathbf{r},t)}{\partial t} = -\mu_0\frac{\partial}{\partial t}\left[\mathbf{j}(\mathbf{r},t) + \frac{\partial\mathbf{P}(\mathbf{r},t)}{\partial t} + \varepsilon_0\frac{\partial\mathbf{E}(\mathbf{r},t)}{\partial t}\right]$$

$$\mathbf{rot}\,\mathbf{rot}\,\mathbf{E}(\mathbf{r},t) + \frac{1}{c_0^2}\frac{\partial^2\mathbf{E}(\mathbf{r},t)}{\partial t^2} = -\mu_0\frac{\partial\mathbf{j}(\mathbf{r},t)}{\partial t} - \mu_0\frac{\partial^2\mathbf{P}(\mathbf{r},t)}{\partial t^2}$$

and:

$$\mathbf{div}[\,\varepsilon_0\mathbf{E}(\mathbf{r},t) + \mathbf{P}(\mathbf{r},t)] = 0$$

magnetic field:

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

$$\mathbf{rot}\,\mathbf{rot}\,\bar{\mathbf{E}}(\mathbf{r},\omega) - \frac{\omega^2}{c_0^2}\,\bar{\mathbf{E}}(\mathbf{r},\omega) = i\omega\mu_0\bar{\mathbf{j}}(\mathbf{r},\omega) + \omega^2\mu_0\bar{\mathbf{P}}(\mathbf{r},\omega)$$

B) Frequency domain

and

$$\mathbf{div}[\,\varepsilon_0\bar{\mathbf{E}}(\mathbf{r},\omega) + \bar{\mathbf{P}}(\mathbf{r},\omega)] = 0$$

magnetic field:

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

Stationary fields in time domain correspond to the solution above and just multiplied with $\mathrm{e}^{-i\omega t}$. For non-stationary fields and linear media it can be written as a superposition of individual solutions: $\mathbf{E}(\mathbf{r},t) = \int_{-\infty}^{\infty} \bar{\mathbf{E}}(\mathbf{r},\omega)\mathrm{e}^{-i\omega t}\, d\omega$.

1.2 Material properties

Having exact information on the functional dependency of $\mathbf{P}(\mathbf{E})$ and $\mathbf{j}(\mathbf{E})$ is the solution of a complicated many body problem. The discussion is too complicated and often phenomenological models are fully sufficient. The functional dependency between cause (electric field) and action (induced polarization or current) is governed here by linear response theory. This assumption breaks down for larger field strength. Effects thereof are subject to the course on "*Nonlinear optics*" by Prof. Koos. The description of the functional dependency is possible in time or frequency domain. We distinguish the terms *response function* for the time domain and *transfer function* for the frequency domain. This terminology applies to many situations in this but also in other lectures.

$$\mathbf{E}(\mathbf{r},t) \rightarrow \textbf{medium}\ (\text{response function}) \rightarrow \mathbf{P}(\mathbf{r},t)$$
$$\bar{\mathbf{E}}(\mathbf{r},\omega) \rightarrow \textbf{medium}\ (\text{transfer function}) \rightarrow \bar{\mathbf{P}}(\mathbf{r},\omega)$$

1.2.1 Basic properties

For a general biaxial anisotropic material the response and the transfer function read as

$P_i(\mathbf{r},t) = \varepsilon_0 \sum_{j=1}^{3} \int_{-\infty}^{t} R_{ij}^{(1)}(\mathbf{r},t-t')E_j(\mathbf{r},t')dt'$ $\quad$ $R(\mathbf{r},\tau)$ - response function

$\bar{P}_i(\mathbf{r},\omega) = \varepsilon_0 \sum_{j=1}^{3} \chi_{ij}^{(1)}(\mathbf{r},\omega)\bar{E}_j(\mathbf{r},\omega)$ $\quad$ $\chi^{(1)}(\mathbf{r},\omega)$- susceptibility

Response and transfer function are linked by a Fourier transform

$$\chi^{(1)}(\mathbf{r},\omega) \overset{FT}{\leftrightarrow} R(\mathbf{r},\tau)$$

For the simplest case of a linear, homogenous, isotropic and dispersive media this reads as

$$R(t) = \int_{-\infty}^{\infty} \chi^{(1)}(\omega)\mathrm{e}^{-i\omega t}\, d\omega$$

In this section and the following we will strictly deal with these isotropic, dispersive materials. Anisotropic materials are subject to an individual section.

To describe the properties of materials on phenomenological grounds, we can use physical insights and can develop basic models to express their properties. The

description of the material response is much more convenient in frequency domain since it is an algebraic expression. From previous lectures we know that there are two basic functional dependencies that can be used to describe phenomenologically the material response.

We can distinguish between contributions from free electrons (that will lead to a **Drude model**) and contributions from bound electrons, lattice vibrations or similar effect (that will lead to a **Lorentz model**). These models, which we will shortly discuss in the following, lead to a permittivity that is composed of one Drude term and a countable finite number of Lorentz terms. Each of these terms is characterized by a few free parameters only that have to be found suitably for each material.

1.2.2 Drude model

The Drude model is canonical to describe the contributions of free electrons. Such assumption is fully justified to express the optical properties of metals and excited semi-conductors. The free electrons eventually lead to the model of a free electron gas, which we develop in the following. The main feature of the Drude model is the acceleration of the electrons by the electric field that causes a current. Contrary to the Lorentz model, there is no restoring force and the electrons are simply accelerated. Interaction of the electrons with the ionic background (that assures charge neutrality) and the interaction between electrons suggest that the electrons are not accelerated arbitrarily but some damping will occur. This damping is taken into account as well on phenomenological grounds.

To describe the equation of motion of the charges, we use the following Newtonian equation:

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

Please note, it will be important further on, this corresponds to a driven harmonic oscillator without restoring force. m is here the mass of the electrons that is in a semiconductor an effective mass. g is the phenomenologically introduced damping and $\mathbf{s}(\mathbf{r},t)$ describes the displacement of the electron due to an electric field $\mathbf{E}(\mathbf{r},t)$.

The induced current density for an electron density of N is

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

which leads to an expression as

$$\frac{\partial}{\partial t}\mathbf{j}(\mathbf{r},t)+g\mathbf{j}(\mathbf{r},t)=\frac{e^2N}{m}\mathbf{E}(\mathbf{r},t)=\varepsilon_0 f\mathbf{E}(\mathbf{r},t)=\varepsilon_0\omega_p^2\mathbf{E}(\mathbf{r},t)$$

where the plasma frequency has been introduced as

$$\omega_p^2=f=\frac{1}{\varepsilon_0}\frac{e^2N}{m}$$

The solution of the above differential equation in time domain is best expressed in frequency domain by Fourier transforming the equation:

$$-i\omega\bar{\mathbf{j}}(\mathbf{r},\omega)+g\bar{\mathbf{j}}(\mathbf{r},\omega)=\varepsilon_0\omega_p^2\bar{\mathbf{E}}(\mathbf{r},\omega)$$

This equation can be recast to express the induced current density depending on the electric field, which is just one form of writing the constitutive relation.

$$\bar{\mathbf{j}}(\mathbf{r},\omega)=\frac{\varepsilon_0\omega_p^2}{g-i\omega}\bar{\mathbf{E}}(\mathbf{r},\omega)=\sigma(\mathbf{r},\omega)\bar{\mathbf{E}}(\mathbf{r},\omega)$$

Here, the complex conductivity had been introduced that collects all the pre-factors.

1.2.4 Lorentz model

The Lorentz model expresses the contribution of bound electrons and lattice vibrations. Here, the starting point is again a physically motivated expression for the displacement of a specific kind of particle but contrary to the free electrons there will be now a restoring force. This leads to a driven, damped harmonic oscillator in the phenomenological description according to:

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

Here, ω_0 is the resonance frequency of the harmonic oscillator and we have generalized the charge to q. The harmonic oscillator is damped as well by a factor g. $\mathbf{s}(\mathbf{r},t)$ is best understood as the displacement from a fix point. The simplest physical picture you can and should potentially use is that of a negative charge displaced from its ionic core. Displacing the negative charge from the positively charged ionic core induces an electric dipole moment

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

Making the transition to a dipole density or electrical polarization, we reach an expression that links the polarization to the displacement via

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

Plugging this expression into the equation of motion we obtain a differential equation for the polarization that reads as

$$\frac{\partial^2}{\partial t^2}\mathbf{P}(\mathbf{r},t) + g\frac{\partial}{\partial t}\mathbf{P}(\mathbf{r},t) + \omega_0^2\mathbf{P}(\mathbf{r},t) = \frac{q^2N}{m}\mathbf{E}(\mathbf{r},t) = \varepsilon_0 f\mathbf{E}(\mathbf{r},t)$$

where the oscillator strength has been introduced according to

$$f = \frac{1}{\varepsilon_0}\frac{q^2N}{m}$$

Solving the equation in frequency space

$$-\omega^2\bar{\mathbf{P}}(\mathbf{r},\omega) - ig\omega\bar{\mathbf{P}}(\mathbf{r},\omega) + \omega_0^2\bar{\mathbf{P}}(\mathbf{r},\omega) = \varepsilon_0 f\bar{\mathbf{E}}(\mathbf{r},\omega)$$

leads to the final expression of

$$\bar{\mathbf{P}}(\mathbf{r},\omega) = \frac{\varepsilon_0 f}{(\omega_0^2 - \omega^2) - ig\omega}\bar{\mathbf{E}}(\mathbf{r},\omega) = \varepsilon_0\chi(\omega)\bar{\mathbf{E}}(\mathbf{r},\omega)$$

where the susceptibility had been introduced. With these expressions, all materials can be described in terms of uncoupled, driven, damped, harmonic oscillators + one Drude term.

1.2.4 Complex dielectric function

Constitutive relations $\bar{\mathbf{j}}(\mathbf{r},\omega)$and $\bar{\mathbf{P}}(\mathbf{r},\omega)$ are plugged into wave equations

$$\mathbf{rot}\,\mathbf{rot}\,\bar{\mathbf{E}}(\mathbf{r},\omega) - \frac{\omega^2}{c_0^2}\bar{\mathbf{E}}(\mathbf{r},\omega) = \omega^2\mu_0\bar{\mathbf{P}}(\mathbf{r},\omega) + i\omega\mu_0\bar{\mathbf{j}}(\mathbf{r},\omega)$$

$$= [\mu_0\varepsilon_0\omega^2\chi(\omega) + i\omega\mu_0\sigma(\mathbf{r},\omega)]\bar{\mathbf{E}}(\mathbf{r},\omega)$$

$$\mathbf{rot}\,\mathbf{rot}\,\bar{\mathbf{E}}(\mathbf{r},\omega) = \frac{\omega^2}{c_0^2}\left\{\mathbf{1} + \chi(\omega) + \frac{i}{\omega\varepsilon_0}\sigma(\mathbf{r},\omega)\right\}\bar{\mathbf{E}}(\mathbf{r},\omega)$$

$$= \frac{\omega^2}{c_0^2}\varepsilon(\omega)\bar{\mathbf{E}}(\mathbf{r},\omega)$$

leading to the generalized complex dielectric function

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

$$\varepsilon(\omega) = 1 + \sum_j \frac{f_j}{\left(\omega_{0j}^2 - \omega^2\right) - ig_j\omega} + \frac{\omega_p^2}{-\omega^2 - ig\omega}$$

Please note, this is what is usually called the relative permittivity of the medium. The entire permittivity requires a multiplication with the vacuum permittivity.

1.2.5 Kramers-Kronig dispersion relation

Statement: real (dispersion) and imaginary (absorption) part of the transfer function are linked by an integral transformation!

Here: dielectric susceptibility or dielectric function

Applies when the response function is:

- time invariant,
- real valued,
- when causality applies.

We will explicitly take advantage of each of these properties in the derivation of the Kramers-Kronig dispersion relation.

Causality is a fundamental property and suggests that the polarization shall not depend on some future electric field

$$\mathbf{P}(r,t) = \varepsilon_0 \int_{-\infty}^{t} R(t-t')\mathbf{E}(r,t')dt' \quad \leftrightarrow \quad \mathbf{P}(r,t) = \varepsilon_0 \int_{0}^{\infty} R(\tau)\mathbf{E}(r,t-\tau)d\tau$$

The response function is ***real valued*** as a direct consequence from Maxwell's equations that are equally real valued

$$R(\tau) = \int_{-\infty}^{\infty} \chi(\omega)\mathrm{e}^{-i\omega\tau}\, d\omega = \int_{-\infty}^{\infty} \chi^*(\omega)\mathrm{e}^{i\omega\tau}\, d\omega \;\rightarrow\; \chi(\omega) = \chi^*(-\omega)$$

Causality of the response function requires to write it as

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

Now we have:

$$\chi(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R(\tau)\mathrm{e}^{i\omega\tau}\, d\tau = \frac{1}{2\pi}\int_{-\infty}^{\infty} \theta(\tau)y(\tau)\mathrm{e}^{i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} \bar{\theta}(\omega-\bar{\omega})\bar{y}(\bar{\omega})\, d\bar{\omega}$$

since a product in time space is a convolution in frequency space (and vice versa).

What is the Fourier transform of the Heaviside distribution? Details on the mathematics of the Fourier transform of the Heaviside distribution will be given in the tutorial but the final result will read as

$$2\pi\bar{\theta}(\omega) = \int_{-\infty}^{\infty} \theta(t)e^{i\omega t}dt = \lim_{\varepsilon\to 0}\frac{i}{\omega + i\varepsilon} = P\frac{i}{\omega} + \pi\delta(\omega)$$

The latter two terms are only defined in terms of integrals. Eventually, the Fourier transform consists of two parts that are defined:

Delta distribution:

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

Cauchy principal value:

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

Using these two definitions we can explicitly calculate the susceptibility by performing the convolution for the two individual terms with the function

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

Now we exploit the fact that the function $y(\tau)$ can be chosen freely at negative times without affecting the susceptibility. Therefore we chose, for example, the function to be either an even or an odd function.

(1) $y(-\tau) = y(\tau)$ even function

(2) $y(-\tau) = -y(\tau)$ odd function.

Now let us discuss the consequences on its Fourier transform!

(1) Considering the function $y(\tau)$ to be an even function

$$\bar{y}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau \, y(\tau) e^{i\omega\tau} = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau \, y(\tau) e^{-i\omega\tau} = \bar{y}^*(\omega)$$

Then $\bar{y}(\omega)$ is real valued as well. Therefore, we can conclude that the complex conjugate from the susceptibility as discussed above reads as

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

Based on these two expressions we can compute now the real and the imaginary part of the susceptibility:

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

$$\chi(\omega) - \chi^*(\omega) = \cdots = \frac{1}{\pi} P \int_{-\infty}^{\infty} d\bar{\omega} \, \frac{i \, \bar{y}(\bar{\omega})}{\omega - \bar{\omega}} \quad \text{(b)}$$

Now by inserting (a) into the right hand side of (b) we eventually obtain the first Kramers-Kronig relation:

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

Please note, in the last step the arguments of the denominator where interchanged to adhere to a usual notation. The equations tells us that once we know the real part of the susceptibility, that tells us something on the dispersion, we can compute the imaginary part of the susceptibility, that tells us something on the absorption. However, it has to be stressed that for an exact analysis the dispersion needs to be known across the entire frequency domain, which is technically difficult, if not to say impossible to obtain. However, the integrand sufficiently decays such that the expression can be often used in an approximate sense.

(2) In a second step we consider the function $y(\tau)$ to be an odd function and discuss the consequences on how the function $\bar{y}(\omega)$ and its conjugate are interlinked.

$$\bar{y}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, y(\tau) e^{i\omega\tau} = -\frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, y(\tau) e^{-i\omega\tau} = -\bar{y}^*(\omega)$$

Therefore, we can conclude that the complex conjugate from the susceptibility as discussed above reads as

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

We perform the same procedure as before and compute the real and the imaginary part of the susceptibility:

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

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

By inserting (a) into the lhs of (b) we obtain the second Kramers-Kronig relation:

By using the following previously discussed functional dependencies

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

$$\chi(\omega) = \chi^*(-\omega) \quad \rightarrow \quad \chi'(\omega) = \chi'(-\omega) \quad \text{and} \quad \chi''(\omega) = -\chi''(-\omega)$$

and

$$\chi(\omega) = \varepsilon(\omega) - 1 = [\varepsilon'(\omega) - 1] + i\varepsilon''(\omega)$$

it follows for the two Kramers-Kronig relations for the permittivity that

$$\chi'(\omega) = \frac{1}{\pi} P\int_{-\infty}^{\infty} d\bar{\omega}\, \frac{\chi''(\bar{\omega})}{\bar{\omega} - \omega} = \frac{1}{\pi} P\int_{-\infty}^{0} d\bar{\omega}\, \frac{\chi''(\bar{\omega})}{\bar{\omega} - \omega} + \frac{1}{\pi} P\int_{0}^{\infty} d\bar{\omega}\, \frac{\chi''(\bar{\omega})}{\bar{\omega} - \omega}$$

$$= \frac{1}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\chi''(-\bar{\omega})}{-\bar{\omega}-\omega} + \frac{1}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\chi''(\bar{\omega})}{\bar{\omega}-\omega}$$

$$= -\frac{1}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\chi''(\bar{\omega})(\bar{\omega}-\omega)}{(-\bar{\omega}-\omega)(\bar{\omega}-\omega)} + \frac{1}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\chi''(\bar{\omega})(-\bar{\omega}-\omega)}{(\bar{\omega}-\omega)(-\bar{\omega}-\omega)}$$

$$= -\frac{1}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\chi''(\bar{\omega})(\bar{\omega}-\omega)}{\omega^2-\bar{\omega}^2} + \frac{1}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\chi''(\bar{\omega})(-\bar{\omega}-\omega)}{\omega^2-\bar{\omega}^2}$$

$$= \frac{1}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\omega\chi''(\bar{\omega}) - \bar{\omega}\chi''(\bar{\omega}) - \omega\chi''(\bar{\omega}) - \bar{\omega}\chi''(\bar{\omega})}{\omega^2-\bar{\omega}^2}$$

$$= \frac{2}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\bar{\omega}\chi''(\bar{\omega})}{\bar{\omega}^2-\omega^2}$$

$$\varepsilon'(\omega) - 1 = \frac{2}{\pi} P \int_0^\infty d\bar{\omega}\, \frac{\bar{\omega}\varepsilon''(\bar{\omega})}{\bar{\omega}^2-\omega^2}$$

$$\varepsilon''(\omega) = -\frac{2}{\pi}\omega P \int_0^\infty d\bar{\omega}\, \frac{[\varepsilon'(\bar{\omega})-1]}{\bar{\omega}^2-\omega^2}$$

In a last step we have only plugged in the real and imaginary part of the susceptibility in terms of the permittivity. The second equation can be derived in an analogous manner.

An example for the immediate application of the Kramers-Kronig relation is that of an arbitrary narrow absorption line at a discrete frequency. This absorption line can be written using a delta-distribution according to $\varepsilon''(\omega) \sim \delta(\omega - \omega_0)$. It follows from the equation above that only a contribution to the integral is made at the frequency $\bar{\omega} = \omega_0$. The real part can be evaluated according to $\varepsilon'(\omega) - 1 \sim \frac{\omega_0}{\omega_0^2 - \omega^2}$. This corresponds to a Lorentzian line with an arbitrary narrow Absorption line.

More details and a good source for a reference is the following book: V. Lucarini J.J. Saarinen, K.-E. Peiponen E.M. Vartiainen 'Kramers–Kronig Relations in Optical Materials Research' Springer 2004**.**

1.3 Maxwell's stress tensor

From Maxwell's equations it follows straight that electric charge and electric current density are related by a continuity equation. This reads as

$$\frac{\partial \rho(\mathbf{r},t)}{\partial t} + \nabla \cdot \mathbf{j}(\mathbf{r},t) = 0$$

It suggests that the rate of decrease of the charge density $\rho(\mathbf{r},t)$ is equal to the divergence of the current density. Charge is a conserved quantity that cannot vanish at one point and instantly appear somewhere else in space. Instead, charges have to move continuously along a path in space where they can create a current. We can therefore also think of the current density as being given by $\mathbf{j}(\mathbf{r},t) = \rho(\mathbf{r},t)\mathbf{v}(\mathbf{r},t)$.

To define basic terminology, we can use an analogy to mechanics.

The same type of continuity holds for momentum that can be written as

$\frac{\partial \mathbf{p}(\mathbf{r},t)}{\partial t} + \nabla \cdot \mathbf{J}_{\mathrm{M}}(\mathbf{r},t) = 0$

where $\mathbf{p}(\mathbf{r},t)$ is the momentum density and $\mathbf{J}_{\mathrm{M}}(\mathbf{r},t)$ is the momentum density tensor. It can be understood as the outer product of the momentum density and velocity, i.e. $\mathbf{J}_{\mathrm{M}}(\mathbf{r},t) = \mathbf{p}(\mathbf{r},t)\mathbf{v}(\mathbf{r},t)$. The rate of change of momentum per volume is the force per volume

$\mathbf{f}(\mathbf{r},t) = \frac{\partial \mathbf{p}(\mathbf{r},t)}{\partial t} = -\nabla \cdot \mathbf{J}_{\mathrm{M}}(\mathbf{r},t)$

The total force acting on the body needs to be integrated across the volume such that

$\mathbf{F}(\boldsymbol{t}) = \int \mathbf{f}(\mathbf{r},t)\, dV = -\int \nabla \cdot \mathbf{J}_{\mathrm{M}}(\mathbf{r},t)\, dV = -\int \mathbf{J}_{\mathrm{M}}(\mathbf{r},t) \cdot d\mathbf{A}$

The last expression in the integral, i.e. $-\mathbf{J}_{\mathrm{M}}(\mathbf{r},t) \cdot d\mathbf{A}$ is the force per area which defines the stress $\mathbf{T}(\mathbf{r},t)$ tensor as $\mathbf{T}(\mathbf{r},t) = -\mathbf{J}_{\mathrm{M}}(\mathbf{r},t)$.

Now we proceed with the electromagnetic problem and discuss the force per volume acting on free charges in the presence of electromagnetic fields. This will lead to the derivation of the Maxwell Stress Tensor (or actually the Minkowski Stress Tensor since Maxwell only considered free space).

The force acting on a charge density is given by the Lorentz force that reads as

$$\mathbf{f}(\mathbf{r},t) = \rho(\mathbf{r},t)\mathbf{E}(\mathbf{r},t) + \rho(\mathbf{r},t)\mathbf{v}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t) = \rho(\mathbf{r},t)\mathbf{E}(\mathbf{r},t) + \mathbf{j}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)$$

Now using Maxwell's equations

$$\nabla \cdot \mathbf{D}(\mathbf{r},t) = \rho(\mathbf{r},t)$$

and

$$\nabla \times \mathbf{H}(\mathbf{r},t) - \frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} = \mathbf{j}(\mathbf{r},t)$$

we obtain

$$\mathbf{f}(\mathbf{r},t) = \mathbf{E}(\mathbf{r},t)[\nabla \cdot \mathbf{D}(\mathbf{r},t)] - \mathbf{B}(\mathbf{r},t) \times \nabla \times \mathbf{H}(\mathbf{r},t) - \frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} \times \mathbf{B}(\mathbf{r},t)$$

By keeping in mind that

$$\frac{\partial}{\partial t}[\mathbf{D}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)] = \frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t} \times \mathbf{B}(\mathbf{r},t) + \mathbf{D}(\mathbf{r},t) \times \frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t}$$

we obtain

$$\mathbf{f}(\mathbf{r},t) = \mathbf{E}(\mathbf{r},t)[\nabla \cdot \mathbf{D}(\mathbf{r},t)] - \mathbf{B}(\mathbf{r},t) \times \nabla \times \mathbf{H}(\mathbf{r},t) + \mathbf{D}(\mathbf{r},t) \times \frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t} - \frac{\partial}{\partial t}[\mathbf{D}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)]$$

Substituting for the second last time derivative $\nabla \times \mathbf{E}(\mathbf{r},t) = -\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t}$ gives

$$\mathbf{f}(\mathbf{r},t) = \mathbf{E}(\mathbf{r},t)[\nabla \cdot \mathbf{D}(\mathbf{r},t)] - \mathbf{B}(\mathbf{r},t) \times \nabla \times \mathbf{H}(\mathbf{r},t) - \mathbf{D}(\mathbf{r},t) \times \nabla \times \mathbf{E}(\mathbf{r},t) - \frac{\partial}{\partial t}[\mathbf{D}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)]$$

By taking advantage that $\nabla \cdot \mathbf{B}(\mathbf{r},t) = 0$, we can insert a suitable zero into the expression to obtain a highly symmetric expression that reads as

$$\mathbf{f}(\mathbf{r},t) = \mathbf{E}(\mathbf{r},t)[\nabla \cdot \mathbf{D}(\mathbf{r},t)] + \mathbf{H}(\mathbf{r},t)[\nabla \cdot \mathbf{B}(\mathbf{r},t)] - \mathbf{B}(\mathbf{r},t) \times \nabla \times \mathbf{H}(\mathbf{r},t) - \mathbf{D}(\mathbf{r},t) \times \nabla \times \mathbf{E}(\mathbf{r},t) - \frac{\partial}{\partial t}[\mathbf{D}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)]$$

This expression now needs to be simplified. Please note that for the tensor that mediates the response between electric field and displacement it can be shown (in the tutorial) that

$$\mathbf{E}(\mathbf{r},t)[\nabla \cdot \mathbf{D}(\mathbf{r},t)] - \mathbf{D}(\mathbf{r},t) \times \nabla \times \mathbf{E}(\mathbf{r},t) = \frac{\partial}{\partial x_\beta}\left\{E_\alpha D_\beta - \frac{1}{2}\delta_{\alpha\beta} E_\gamma D_\gamma\right\}$$
$$= \nabla \cdot \left\{\mathbf{E}(\mathbf{r},t)\mathbf{D}(\mathbf{r},t) - \frac{1}{2}\mathbf{I}[\mathbf{E}(\mathbf{r},t) \cdot \mathbf{D}(\mathbf{r},t)]\right\}$$

Please note that in this expression the nabla operator ∇ operates on $\mathbf{D}(\mathbf{r},t)$ and $\mathbf{I}$ is the idendity matrix. The same holds for the expression that contains the magnetic fields.

$$\mathbf{H}(\mathbf{r},t)[\nabla \cdot \mathbf{B}(\mathbf{r},t)] - \mathbf{B}(\mathbf{r},t) \times \nabla \times \mathbf{H}(\mathbf{r},t) = \nabla \cdot \left\{\mathbf{H}(\mathbf{r},t)\mathbf{B}(\mathbf{r},t) - \frac{1}{2}\mathbf{I}[\mathbf{H}(\mathbf{r},t) \cdot \mathbf{B}(\mathbf{r},t)]\right\}$$

This simplifies the expression for the force to be

$$\mathbf{f}(\mathbf{r},t) = \nabla \cdot \left\{\mathbf{E}(\mathbf{r},t)\mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t)\mathbf{B}(\mathbf{r},t) - \frac{1}{2}\mathbf{I}[\mathbf{E}(\mathbf{r},t) \cdot \mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t) \cdot \mathbf{B}(\mathbf{r},t)]\right\} - \frac{\partial}{\partial t}[\mathbf{D}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)]$$

Now, by being reminded that $\mathbf{f}(\mathbf{r},t) = \frac{\partial \mathbf{p}(\mathbf{r},t)}{\partial t}$, where $\mathbf{p}(\mathbf{r},t)$ is here the momentum density of the free charges, we can write the expression above as

$$\frac{\partial \mathbf{p}(\mathbf{r},t)}{\partial t} + \frac{\partial}{\partial t}[\mathbf{D}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)]$$
$$= -\nabla \cdot \left\{ -\left(\mathbf{E}(\mathbf{r},t)\mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t)\mathbf{B}(\mathbf{r},t) - \frac{1}{2}\mathbf{I}[\mathbf{E}(\mathbf{r},t) \cdot \mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t) \cdot \mathbf{B}(\mathbf{r},t)] \right) \right\}$$

Now, there is a clear physical interpretation of the individual terms. The first term on the left-hand-side is the rate of change of momentum of the free charges. The second term on the left-hand-side is the rate of change of momentum density of the electromagnetic field. The momentum of the electro-magnetic field here is introduced as $\mathbf{D}(\mathbf{r},t) \times \mathbf{B}(\mathbf{r},t)$. The total momentum of a propagating electromagnetic wave therefore increases upon entering a dielectric medium proportional to the refractive index n of the medium. Please note, this is discussed in a simplified situation where we neglect dispersion and consider only an isotropic non-absorbing medium. This definition for the momentum of the electro-magnetic field is not obvious and was questioned in the course of time. Apparently, it drew criticism since it seems to be incompatible with the conservation of angular momentum. Therefore, other expressions for the electromagnetic wave momentum where suggested, e.g. by Abraham it was suggested that the momentum should be defined as $\frac{1}{c^2}\left(\mathbf{E}(\mathbf{r},t) \times \mathbf{H}(\mathbf{r},t)\right)$. The momentum of an electromagnetic wave entering a dielectric medium then falls proportional to the refractive index n. Under the Abraham tensor, a photon therefore carries less momentum within a medium than in the Minkowski case. This led to quite a controversial discussion, which is, interesting enough, not yet settled. There are regular suggestions for experiments that shall be able to decide what is the change of the momentum in a medium, but so far there is no conclusion. No observable could be thus far perceived for which different predictions exist in the context of the Minkowski or the Abraham picture. For more details you may take a look into review article "***Colloquium***: Momentum of an electromagnetic wave in dielectric media" from Pfeifer *et al.* in Rev. Mod. Phys. **Vol 79**, *1197* (2007),

The right-hand-side of the equation above represents the negative of the divergence of the momentum current density. The momentum current density tensor therefore is

$$\mathbf{J}(\mathbf{r},t) = -\left(\mathbf{E}(\mathbf{r},t)\mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t)\mathbf{B}(\mathbf{r},t) - \frac{1}{2}\mathbf{I}[\mathbf{E}(\mathbf{r},t) \cdot \mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t) \cdot \mathbf{B}(\mathbf{r},t)] \right)$$

and the Minkowksi stress tensor $\mathbf{T}$ is defined as the negative of this

$$\mathbf{T}(\mathbf{r},t) = \mathbf{E}(\mathbf{r},t)\mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t)\mathbf{B}(\mathbf{r},t) - \frac{1}{2}\mathbf{I}[\mathbf{E}(\mathbf{r},t) \cdot \mathbf{D}(\mathbf{r},t) + \mathbf{H}(\mathbf{r},t) \cdot \mathbf{B}(\mathbf{r},t)]$$

The force per area is then defined as

$$\mathbf{T}(\mathbf{r},t) \cdot d\mathbf{A}$$

In the case of vacuum this reduces to

$$\mathbf{T}(\mathbf{r},t) = \varepsilon_0 \mathbf{E}(\mathbf{r},t)\mathbf{E}(\mathbf{r},t) + \frac{1}{\mu_0}\mathbf{B}(\mathbf{r},t)\mathbf{B}(\mathbf{r},t) - \frac{1}{2}\mathbf{I}\left[\varepsilon_0 \mathbf{E}(\mathbf{r},t)\cdot\mathbf{E}(\mathbf{r},t) + \frac{1}{\mu_0}\mathbf{B}(\mathbf{r},t)\cdot\mathbf{B}(\mathbf{r},t)\right]$$

The components T_{ij} of the stress tensor have the following meaning: It is the force per unit area in direction $\hat{\mathbf{e}}_i$ acting on the surface being normal in direction $\hat{\mathbf{e}}_j$. Thus, T_{ii} are pressures (forces normal to surfaces), whereas T_{ij} with $i \neq j$ are shears (forces parallel to surfaces). Note that only the total fields enter in the equations, which contain all information about the charge distributions, on which the forces act.

A more convenient analysis is usually done in frequency domain. Considering the time average of the electromagnetic force and calculating the total force by integration across the entire surface of a particle, we obtain:

$$\begin{aligned}\langle \bar{\mathbf{F}}(\boldsymbol{\omega})\rangle &= \langle \oint \left(\bar{\mathbf{T}}(\mathbf{r},\omega)\cdot\mathbf{n}\right)\mathrm{d}A\rangle \\ &= \int_S \left\{\frac{\varepsilon_0\varepsilon(\omega)}{2}\Re[(\bar{\mathbf{E}}(\mathbf{r},\omega)\cdot\mathbf{n})\bar{\mathbf{E}}^*(\mathbf{r},\omega)] - \frac{\varepsilon_0\varepsilon(\omega)}{4}\left(\bar{\mathbf{E}}(\mathbf{r},\omega)\cdot\bar{\mathbf{E}}^*(\mathbf{r},\omega)\right)\mathbf{n}\right. \\ &\left. + \frac{\mu_0\mu(\omega)}{2}\Re[(\bar{\mathbf{H}}(\mathbf{r},\omega)\cdot\mathbf{n})\bar{\mathbf{H}}^*(\mathbf{r},\omega)] - \frac{\mu_0\mu(\omega)}{4}\left(\mathbf{H}(\mathbf{r},\omega)\cdot\bar{\mathbf{H}}^*(\mathbf{r},\omega)\right)\mathbf{n}\right\} dl'\end{aligned}$$

where dl' is the length of a line segment of the surface.

The net radiation torque on the particle is calculated by

$$\langle \boldsymbol{\tau}(\boldsymbol{\omega})\rangle = \langle \oint \mathbf{r} \times \left(\bar{\mathbf{T}}(\mathbf{r},\omega)\cdot\mathbf{n}\right)\mathrm{d}A\rangle.$$

1.4 Poynting vector and energy balance

1.4.1 Time averaged Poynting vector

One of the last things we wish to discuss here on the base of Maxwell's equations concerns the Poynting vector and its conservation/dissipation. Generally, the flow of energy is expressed using the Poynting vector $\mathbf{S}(\mathbf{r},t)$. Particularly, in the context of discussing the measurement of an optical signal, the energy flux through a plane in normal direction $\mathbf{S}(\mathbf{r},t)\cdot\mathbf{n}$ is important. $\mathbf{n}$ is here a vector pointing in normal direction to the interface. This corresponds to the situation of a detector. The instantaneous energy flux is expressed as

$$\mathbf{S}(\mathbf{r},t) = \mathbf{E}_\mathrm{r}(\mathbf{r},t) \times \mathbf{H}_\mathrm{r}(\mathbf{r},t)$$

Please note, contrary to many other situations where the analysis with the electric and magnetic field as a complex quantity is fully sufficient, here we explicitly have to

consider the real part of the fields only. They are abbreviated with the r as subscript. In the measurement process we have to distinguish different time scales:

- the fast oscillation period of the electromagnetic field → $T_0 = \frac{2\pi}{\omega_0} < 10^{-14}s$
- the duration of a possibly slowly varying field, i.e. the temporal duration of a pulse (see next section) → T_P where is usually holds $T_P \gg T_0$
- and the duration of the measurement → T_m where is usually holds $T_m \gg T_0$ but $T_m \geq T_p$.

Generally, a detector cannot measure the fast oscillation but it only measures the temporal average. Under the assumption that the field can be written as a product of a slowly varying envelope $\tilde{\mathbf{E}}(\mathbf{r},t)$ and a fast oscillation $e^{-i\omega_0 t}$

$$\mathbf{E}_{\mathrm{r}}(\mathbf{r},t) = \frac{1}{2}\left[\tilde{\mathbf{E}}(\mathbf{r},t)e^{-i\omega_0 t} + c.c.\right]$$

we can express the instantaneous Poynting vector as

$$\begin{aligned}\mathbf{S}(\mathbf{r},t) &= \mathbf{E}_{\mathrm{r}}(\mathbf{r},t) \times \mathbf{H}_{\mathrm{r}}(\mathbf{r},t)\\ &= \tfrac{1}{4}\left[\tilde{\mathbf{E}}(\mathbf{r},t) \times \tilde{\mathbf{H}}^*(\mathbf{r},t) + \tilde{\mathbf{E}}^*(\mathbf{r},t) \times \tilde{\mathbf{H}}(\mathbf{r},t)\right]\\ &\quad + \tfrac{1}{4}\left[\tilde{\mathbf{E}}(\mathbf{r},t) \times \tilde{\mathbf{H}}(\mathbf{r},t)e^{-2i\omega_0 t} + \tilde{\mathbf{E}}^*(\mathbf{r},t) \times \tilde{\mathbf{H}}^*(\mathbf{r},t)e^{2i\omega_0 t}\right]\\ &= \tfrac{1}{2}\Re\left[\tilde{\mathbf{E}}(\mathbf{r},t) \times \tilde{\mathbf{H}}^*(\mathbf{r},t)\right] + \tfrac{1}{2}\Re\left[\tilde{\mathbf{E}}(\mathbf{r},t) \times \tilde{\mathbf{H}}(\mathbf{r},t)\right]\cos 2\omega_0 t\\ &\quad + \tfrac{1}{2}\Im\left[\tilde{\mathbf{E}}^*(\mathbf{r},t) \times \tilde{\mathbf{H}}^*(\mathbf{r},t)\right]\sin 2\omega_0 t\end{aligned}$$

The actual measured signal corresponds to the time average of $\mathbf{S}(\mathbf{r},t)$

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

This integration causes a vanishing of the fastly oscillating terms and the only remaining expression corresponsd to

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

For the special case of a stationary, i.e. monochromatic field where the electric field can be expressed as $\tilde{\mathbf{E}}(\mathbf{r},t') = \bar{\mathbf{E}}(\mathbf{r},\omega_0)$ and the magnetic field can be expressed as $\tilde{\mathbf{H}}(\mathbf{r},t') = \bar{\mathbf{H}}(\mathbf{r},\omega_0)$, the time averaged Poynting vector reads as

$$\langle \mathbf{S}(\mathbf{r},t)\rangle = \frac{1}{2}\Re[\bar{\mathbf{E}}(\mathbf{r},\omega_0) \times \bar{\mathbf{H}}^*(\mathbf{r},\omega_0)]$$

This corresponds to the intensity $I = \langle \mathbf{S}(\mathbf{r},t)\rangle$.

1.4.2 Energy balance

What we are looking for at the moment is an expression for the dissipation (consider here only passive optical media) of electromagnetic energy density as a function of the absorption (imaginary part of the permittivity) and the magnitude of the field. For practical purposes this is important, e.g. if you wish to calculate the absorption in a solar cell. Then, you need to know how much energy is dissipated, i.e. absorbed, in the spatial domain occupied by the active material of a solar cell, e.g. the silicon material. This Poynting's theorem as it is also called, is usually derived in lectures on electrodynamics but it is contained here both because of its importance and because it's a nice demonstration on how to derive an expression for an observable quantity directly from Maxwell's equations. Starting from Maxwell's equation in space and time domain and multiplying to the curl equation for the electric field the magnetic field and to the curl equation for the magnetic field the electric field provides

$$\mathbf{H}(\mathbf{r},t)\cdot\mathbf{rot}\,\mathbf{E}(\mathbf{r},t)+\mu_0\mathbf{H}(\mathbf{r},t)\cdot\frac{\partial}{\partial t}\mathbf{H}(\mathbf{r},t)=0$$

$$-\varepsilon_0\mathbf{E}(\mathbf{r},t)\cdot\frac{\partial}{\partial t}\mathbf{E}(\mathbf{r},t)+\mathbf{E}(\mathbf{r},t)\cdot\mathbf{rot}\,\mathbf{H}(\mathbf{r},t)=\mathbf{E}(\mathbf{r},t)\left(\mathbf{j}(\mathbf{r},t)+\frac{\partial}{\partial t}\mathbf{P}(\mathbf{r},t)\right)$$

Since we can write $\mathbf{div}\big(\mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\big)=\mathbf{H}(\mathbf{r},t)\cdot\mathbf{rot}\,\mathbf{E}(\mathbf{r},t)-\mathbf{E}(\mathbf{r},t)\cdot\mathbf{rot}\,\mathbf{H}(\mathbf{r},t)$ we can subtract both above expressions to obtain (dropping from now on the space and time arguments)

$$\frac{1}{2}\varepsilon_0\frac{\partial}{\partial t}\mathbf{E}^2+\frac{1}{2}\mu_0\frac{\partial}{\partial t}\mathbf{H}^2+\mathbf{div}(\mathbf{E}\times\mathbf{H})=-\mathbf{E}\left(\mathbf{j}+\frac{\partial}{\partial t}\mathbf{P}\right)$$

Eventually, we wish to emphasize that this expression holds for the real valued fields, so it needs to be written as

$$\frac{1}{2}\varepsilon_0\frac{\partial}{\partial t}\mathbf{E}_r^2+\frac{1}{2}\mu_0\frac{\partial}{\partial t}\mathbf{H}_r^2+\mathbf{div}(\mathbf{E}_r\times\mathbf{H}_r)=-\mathbf{E}_r\left(\mathbf{j}_r+\frac{\partial}{\partial t}\mathbf{P}_r\right)$$

This expression is a balance equation. The temporal change of the electrical energy density $\frac{1}{2}\varepsilon_0\mathbf{E}_r^2$, the magnetic energy density $\frac{1}{2}\mu_0\mathbf{H}_r^2$ and the source/sink of the Poynting vector are linked to the source terms on the right hand side of this equation.

For a stationary field $\mathbf{E}_r(\mathbf{r},t)=\frac{1}{2}\left[\bar{\mathbf{E}}(\mathbf{r},\omega_0)e^{-i\omega_0 t}+c.c.\right]$ the time average of the left hand side of the equation gives

$$\left\langle\frac{1}{2}\varepsilon_0\frac{\partial}{\partial t}\mathbf{E}_r^2+\frac{1}{2}\mu_0\frac{\partial}{\partial t}\mathbf{H}_r^2+\mathbf{div}(\mathbf{E}_r\times\mathbf{H}_r)\right\rangle=\frac{1}{2}\mathbf{div}(\Re[\bar{\mathbf{E}}(\mathbf{r},\omega_0)\times\bar{\mathbf{H}}^*(\mathbf{r},\omega_0)])$$
$$=\mathbf{div}\,\langle\mathbf{S}(\mathbf{r},t)\rangle$$

This corresponds to the source/sink of the time-averaged energy density. For the right hand side of the equation we obtain

$$-\langle\left(\mathbf{j}_r+\frac{\partial}{\partial t}\mathbf{P}_r\right)\mathbf{E}_r\rangle$$

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

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

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

$$=-\frac{1}{2}\omega_0\varepsilon_0\varepsilon''(\omega_0)\bar{\mathbf{E}}\bar{\mathbf{E}}^*$$

Taking both sides of the equation together we obtain

$$\mathbf{div}\,\langle\mathbf{S}(\mathbf{r},t)\rangle=-\frac{1}{2}\omega_0\varepsilon_0\varepsilon''(\omega_0)\bar{\mathbf{E}}\bar{\mathbf{E}}^*$$

There are sinks in the energy flow whenever the imaginary part of the permittivity is nonzero (with the sign convention chosen here the imaginary part is always positive). This occurs to a notable extent whenever there is a resonance in the permittivity. Then, the resonances are intimately linked to absorption.

For non-stationary fields, i.e. the slowly varying fields as assumed before $\mathbf{E}_\mathrm{r}(\mathbf{r},t)=\frac{1}{2}\left[\tilde{\mathbf{E}}(\mathbf{r},t)e^{-i\omega_0 t}+c.c.\right]$, the material dispersion has to be taken explicitly into account and the Poynting theorem reads as (without derivation)

$$\frac{1}{4}\frac{\partial}{\partial t}\left\{\varepsilon_0\frac{\partial[\omega_0\varepsilon'(\mathbf{r},\omega_0)]}{\partial\omega_0}\left|\tilde{\mathbf{E}}(\mathbf{r},t)\right|^2+\mu_0\left|\tilde{\mathbf{H}}(\mathbf{r},t)\right|^2\right\}+\mathbf{div}\,\langle\mathbf{S}(\mathbf{r},t)\rangle=-\frac{1}{2}\omega_0\varepsilon_0\varepsilon''(\omega_0)\left|\tilde{\mathbf{E}}(\mathbf{r},t)\right|^2$$

1.5 Wave propagation (here for pulses)

1.5.1 Fourier expansion of an arbitrary field

Starting point shall be the wave equation in time domain.

$$\mathbf{rot}\,\mathbf{rot}\,\mathbf{E}(\mathbf{r},t)=-\mu_0\mathbf{rot}\frac{\partial\mathbf{H}(\mathbf{r},t)}{\partial t}=-\mu_0\frac{\partial}{\partial t}\left[\mathbf{j}(\mathbf{r},t)+\frac{\partial\mathbf{P}(\mathbf{r},t)}{\partial t}+\varepsilon_0\frac{\partial\mathbf{E}(\mathbf{r},t)}{\partial t}\right]$$

Using a harmonic Ansatz function in both time and spatial, i.e. looking for solutions where all fields oscillate according to $e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$, assuming a homogenous, isotropic, local, dispersive material that can be described with the generalized dielectric function $\varepsilon(\omega)$ as introduced before, and taking advantage that divergence vanishes such that $\mathbf{rot}\,\mathbf{rot}\,\mathbf{A}=\mathbf{grad}\,(\mathbf{div}\,\mathbf{A})-\Delta\mathbf{A}$ we obtain the following set of algebraic equations using the Fourier transform pairs:

$$\frac{\partial}{\partial t} \overset{FT}{\rightarrow} -i\omega$$

$$\Delta\bar{\mathbf{E}}(\mathbf{r},\omega) + \frac{\omega^2}{c_0^2}\varepsilon(\omega)\bar{\mathbf{E}}(\mathbf{r},\omega) = 0$$

Further Fourier transformation (corresponding to an expansion into plane waves) with respect to the spatial coordinates provides

$$\frac{\partial}{\partial \alpha} \overset{FT}{\rightarrow} ik_\alpha$$

$$\bar{\mathbf{E}}(\mathbf{k},\omega)\left(-\mathbf{k}^2 + \frac{\omega^2}{c_0^2}\varepsilon(\omega)\right) = 0$$

This leads to a very important equation, known as the dispersion relation for the eigensolutions propagating in the medium mentioned above. This dispersion relation reads as

$$\mathbf{k}^2 = \frac{\omega^2}{c_0^2}\varepsilon(\omega)$$

A large class of problems in the context of theoretical optics can be described as finding eigensolutions (normalmodes) and the associated eigenvalues (dispersion relations) for more complicated distributions of materials in space. We solve this problem in a later section in the context of anisotropic materials, but similar extensions can be made towards materials with a magnetic dispersions (as relevant in the context of metamaterials) or for material distributions in space that are inhomogeneous. This holds for example for photonic crystals or waveguides. Whereas for photonic crystals the eigensolutions are Bloch periodic modes, for waveguides the eigensolutions are guided modes. So the procedure above looks simple but is quite canonical. Appreciating its implications puts you in the position to discuss optics and photonics in many situations.

The propagation of a pulsed beam with a finite transverse width and of finite duration can be analytically described using the inverse Fourier transform of its spectrum.

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

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

1.5.2 Taylor expansion of the dispersion relation

Restricting for the moment the considerations to a problem where the field propagates only in one-dimension and considering a fixed polarization, a scalar expression can be used. The scalar quantity $u(x,t)$ represents here the respective coefficient of the electric field.

$$u(x,t) = \int_{-\infty}^{\infty} \bar{u}(k,\omega) e^{i(k(\omega)x-\omega t)} dk d\omega$$

This is an exact and rigorous solution. However, in many situations such integral solution is not necessary since the Fourier spectrum will be narrow. Look into a typical application (spectroscopy, nonlinear optics, telecommunication, material processing) for which we need to consider propagation of pulses. In this situation, we have typically pulse envelopes of 10^{-13} s (100 fs) $\leq T_0 \leq 10^{-10}$ s (100 ps). Let us compute the spectrum of the (Gaussian) pulse:

$$f(t) = e^{-i\omega_0 t} e^{-\frac{t^2}{T_0^2}}$$

$$F(\omega) \sim e^{-\frac{(\omega-\omega_0)^2}{4/T_0^2}}$$

Therefore, the spectral width is in the order of $\omega_S^2 = 4/T_0^2$.

For the example above this gives $4 \cdot 10^{10}$Hz $\leq \omega_S \leq 4 \cdot 10^{13}$Hz. This is small when compared to a (visible) center frequency ω_0 that is in the order of $4 \cdot 10^{15}$ Hz. In this situation, it can be helpful to replace the complicated dispersion relation by a Taylor

$$k(\omega) \approx k(\omega_0) + \left.\frac{\partial k}{\partial \omega}\right|_{\omega_0} (\omega - \omega_0) + \frac{1}{2}\left.\frac{\partial^2 k}{\partial \omega^2}\right|_{\omega_0} (\omega - \omega_0)^2 + \cdots$$

expansion at $\omega = \omega_0$. In general, a parabolic (or cubic) approximation will be sufficient.

The individual terms are commonly used in literature and have the following physical meaning:

(A) $k(\omega_0) = k_0 \rightarrow \frac{1}{v_{ph}} = \frac{k_0}{\omega_0} = \frac{n(\omega_0)}{c_0}$ is called the phase velocity and describes the propagation speed of wave fronts in space with a constant phase

(B) $\frac{1}{v_g} = \frac{\partial k}{\partial \omega}\Big|_{\omega_0}$ is called the group velocity. It describes the velocity of the center of the pulse. With the functional dependency of

$$k(\omega) = \frac{\omega}{c_0} n(\omega) \rightarrow \frac{1}{v_g} = \frac{1}{c_0}\left[n(\omega_0) + \omega_0 \frac{\partial n}{\partial \omega}\Big|_{\omega_0}\right]$$

The group velocity therefore is defined as

$$v_g = \frac{c_0}{\left[n(\omega_0) + \omega_0 \frac{\partial n}{\partial \omega}\Big|_{\omega_0}\right]} = \frac{c_0}{n_g(\omega_0)} = v_{ph} \frac{n(\omega_0)}{n_g(\omega_0)}$$

The group index is defined here as

$$n_g(\omega_0) = n(\omega_0) + \omega_0 \frac{\partial n}{\partial \omega}\Big|_{\omega_0}$$

One usually distinguishes between:

normal dispersion: $\frac{\partial n}{\partial \omega}\Big|_{\omega_0} > 0 \rightarrow n_g > n \rightarrow v_g < v_{ph}$

anomalous dispersion: $\frac{\partial n}{\partial \omega}\Big|_{\omega_0} < 0 \rightarrow n_g < n \rightarrow v_g > v_{ph}$

(C) $D_\omega = \frac{\partial^2 k}{\partial \omega^2}\Big|_{\omega_0}$ is called the group velocity dispersion (GVD or simply dispersion).

Alternatively it can be expressed (especially in the context of telecommunications) as

$$D_\lambda = \frac{\partial}{\partial \lambda}\left(\frac{1}{v_g}\right) = -\frac{2\pi}{\lambda^2} c_0 D_\omega$$

$$D_\omega = \frac{\partial}{\partial \omega}\left(\frac{1}{v_g}\right) = -\frac{1}{v_g^2}\frac{\partial v_g}{\partial \omega}$$

The group velocity dispersion can be:

$$D_\omega > 0 \rightarrow \frac{\partial v_g}{\partial \omega} < 0$$

$$D_\omega < 0 \rightarrow \frac{\partial v_g}{\partial \omega} > 0$$

The entire complicated system is reduced to three parameters only that characterize the material and/or the geometry along which light propagates. The latter is obvious since the dispersion relation can be calculated for an arbitrary system, e.g. an optical fiber, a plasmonic waveguide or even a metamaterial. It is only important that the functional dependency of $k(\omega)$ is known.

2. Diffraction theory

2.1 Scalar approximation

Starting point will be the Helmholtz equation as used in the previous chapter

$$\Delta\bar{\mathbf{E}}(\mathbf{r},\omega)+\frac{\omega^2}{c_0^2}\varepsilon(\omega)\bar{\mathbf{E}}(\mathbf{r},\omega)=0$$

The scalar approximation now consists in approximating the vector with a single scalar quantity that corresponds to a particular component of the electric field

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

Please note this scalar treatment is:

- exact for one-dimensional bundles with a linear polarization, so the field structure is invariant into a second dimension and propagates into the third dimension
- an approximation in the two-dimensional case (exact conditions can be derived later)

In scalar approximation with fixed frequency we have

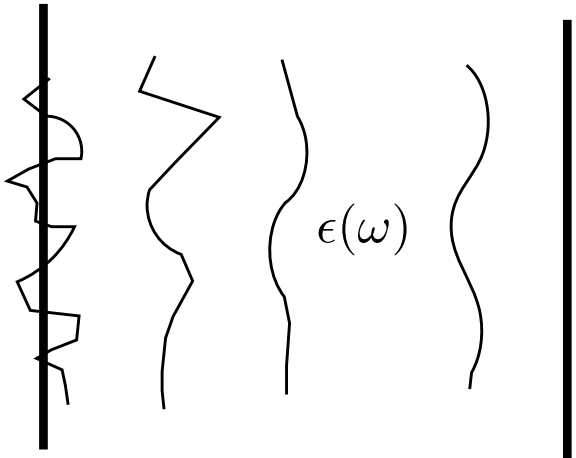
$$\Delta u(\mathbf{r},\omega)+\frac{\omega^2}{c_0^2}\varepsilon(\omega)u(\mathbf{r},\omega)=0$$
$$\Delta u(\mathbf{r},\omega)+k^2(\omega)u(\mathbf{r},\omega)=0$$

The equation is known as the scalar Helmholtz equation. Please note, for convenience we often drop the frequency dependency in the argument since it is implicitly assumed. For some quantities for which it is important to be reminded on that dependency we will maintain it.

2.2 Angular spectrum

It is an exact solution and valid in the realm of the scalar approximation. Since it can be very easily and highly efficient formulated on a computer, it often constitutes the method of choice while propagating fields numerically; but it also serves as an excellent starting point for further considerations and simplifications.

The method can be applied if the task is to find the field distribution in the half-space $z>0$ on the base of the field in an initial plane at, e.g., $z=0$.



In general, a Fourier transform would provide

$$u(\mathbf{r},\omega) = \int_{-\infty}^{\infty} U(\mathbf{k},\omega) e^{i\mathbf{k}(\omega)\cdot\mathbf{r}}\, d^3k$$

The field is written as a superposition of normal modes (elementary plane wave solutions) with different propagation direction and length of the wave vector. However, such description is too simplistic since it does not reflect the fact that only specific wave vectors are allowed with a specific length. The pairs of allowed values for a given frequency are given by the dispersion relation:

$$\mathbf{k^2}(\omega) = k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c_0^2}\varepsilon(\omega)$$

Choosing two components in this equation necessarily fixes the third. For a principal propagation direction of z the natural choice is to fix k_x and k_y. The third component then is usually understood as the propagation constant.

Now, in the context of diffraction theory it is more convenient to introduce Greek letters for the components of the wave vector. No new physics is introduced

$$k_x = \alpha, k_y = \beta, \text{and } k_z = \gamma$$

A 2D Fourier transform eventually is written as

$$u(\mathbf{r}) = \int_{-\infty}^{\infty} U(\alpha,\beta;z) e^{i(\alpha x+\beta y)}\, d\alpha d\beta$$

In analogy to temporal frequencies α and β are called spatial frequencies. The semi-colon here denotes $U(\alpha,\beta;z)$ depends parametrically on z. Plugging this Ansatz into the scalar Helmholtz equation leads to

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

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

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

The solution to this differential equation reads as (dropping the frequency argument now)

$$U(\alpha,\beta;z)=U_1(\alpha,\beta)e^{i\gamma(\alpha,\beta)z}+U_2(\alpha,\beta)e^{-i\gamma(\alpha,\beta)z}$$

There are two different types of solution:

a) $\gamma^2(\omega)>0 \rightarrow \alpha^2+\beta^2\leq k^2(\omega)$ This suggests that the wave vector is real valued which leads to homogenous propagating waves.

b) $\gamma^2(\omega)<0 \rightarrow \alpha^2+\beta^2>k^2(\omega)$ This suggests that the wave vector is complex valued and in particular the propagation constant is imaginary valued, which leads to evanescent waves.

The second term in the solution above ($\sim e^{-i\gamma(\omega)z}$) causes obviously unphysical solutions since it may exponentially grow, therefore we suppress this solution by requiring that $U_2(\alpha,\beta)=0$ independent on the angular frequencies. Indeed, eventually these solutions correspond to backward propagating plane waves ($\rightarrow -z$) emerging from a plane at $z=\infty$. Therefore

$$U(\alpha,\beta;z)=U_1(\alpha,\beta)e^{i\gamma(\alpha,\beta)z}$$

$$U(\alpha,\beta;z)=U_1(\alpha,\beta;0)e^{i\gamma(\alpha,\beta)z}$$

$$U(\alpha,\beta;z)=U_0(\alpha,\beta)e^{i\gamma(\alpha,\beta)z}$$

with the boundary condition $U(\alpha,\beta;0)=U_0(\alpha,\beta)$. The total field in the half space of relevance would be

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

From that equation it is most notably to see that the cause for diffraction (the broadening of the bundle in space) is the different phase accumulation in the propagation direction of the different spatial frequency components characterized by α and β.

The initial angular spectrum that serves as the boundary condition in the equation is taken from a Fourier transform of the field in the referential plane $u_0(x,y) = u(x,y,0)$

$$U_0(\alpha,\beta) = \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} u_0(x,y) e^{-i(\alpha x+\beta y)} dxdy$$

An algorithmic description for the propagation would require at first information on the field in the initial plane $u_0(x,y)$. For the moment it is not known where does this field comes from but we wish to assume it as given. Fourier transform provides the angular spectrum of the initial field $U_0(\alpha,\beta)$. The propagation itself is described as a multiplication with $e^{i\gamma(\alpha,\beta)z}$ to provide a new spectra that is exact in the plane of interest. Inverse Fourier transforming the spectra allows to obtain the field in the plane / point of interest.

It is worth to mention that once the eigenmodes and the associated dispersion relations of a particular space are known, the procedure can be applied to arbitrarily more complicated systems, e.g. photonic crystals.

2.3 Integral Theorem of Helmholtz and Kirchhoff

In the previous section it was shown how any component of the electromagnetic field (or a scalar field) in a source-free half-space $z > 0$ can be expressed in its values in a referential plane at $z = 0$. This at the end requires two double integrations but the general approach of course is very versatile and exact. At first, the initial field needs to be Fourier transformed (first double integration). Then the amplitude of each Fourier component is evolved analytically into the plane of interest. And eventually an inverse Fourier transform (second double integration) is made to express the field in real space. One can, however, also express the field components in their values at $z = 0$ by means of only one surface integral. For this purpose, one can use the so-called integral theorem of Helmholtz and Kirchhoff. This expresses the field at an arbitrary point in space in terms of an integral across a surface

surrounding the space of interest. For a specific choice of the Green's function that emerges this will further reduce to the Rayleigh-Sommerfeld diffraction formula. The Green's function here is an important concept and provides the field everywhere in space depending on the point-wise excitation somewhere else in space. The Rayleigh-Sommerfeld diffraction formula can also be derived from the angular spectrum approach, providing a unique link between the two approaches.

Please note: details on Green's function and Green's second identify that serves as the starting point will be given in the tutorials.

We require that $u(\mathbf{r})$ and $G(\mathbf{r})$ are complex valued functions with single-valued and continuous first and second derivatives. V denotes a volume that is bound by the closed surface S. Green's second identity says that (please note that it is written here on purpose for a Green's function but of course it holds for arbitrary functions)

$$\iiint_V \left(u(\mathbf{r}')\Delta G(\mathbf{r},\mathbf{r}') - G(\mathbf{r},\mathbf{r}')\Delta u(\mathbf{r}')\right) d^3r'$$

$$= \iint_S \left(u(\mathbf{r}')\frac{\partial G(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r},\mathbf{r}')\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}}\right) d^2r'$$

where the derivative is done with respect to the outward normal direction. Please note, this is a convention, which often causes trouble and inconsistencies in the equations. The choice of the outward normal is done, e.g. by Goodman in the book "Introduction to Fourier Optics"; the inward normal is chosen in the book by Born & Wolf "Principles of Optics". If consistently done of course no deviation should occur.

We require that the field $u(\mathbf{r})$ we are eventually looking for is a solution to the Helmholtz equation for a homogenous isotropic medium characterized (here for convenience) by a specific refractive index.

$$(\Delta + k_0^2 n^2)u(\mathbf{r}) = 0$$

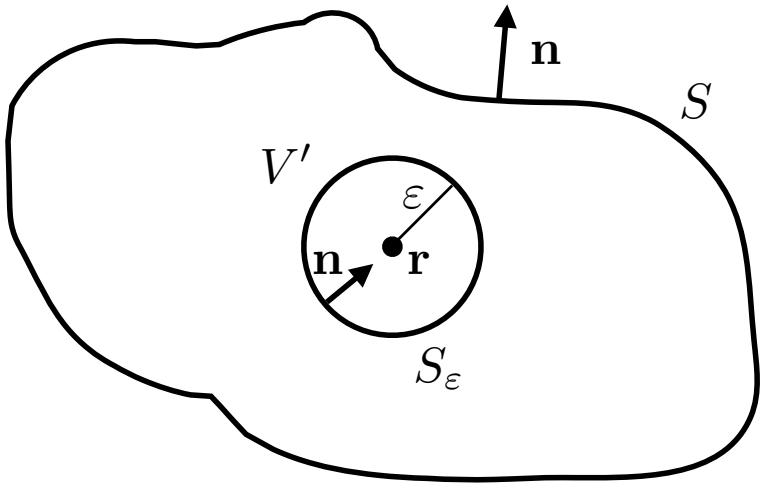
The associated Green's function that solves the problem

$$(\Delta + k_0^2 n^2)G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}' - \mathbf{r})$$

is given by

$$G(\mathbf{r},\mathbf{r}') = \frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}$$

Now apply this to the specific situation



where we consider instead of the original volume a slightly modified volume that excludes a small sphere of radius ϵ around the point $\mathbf{r}$ that is the point of our observation. We want to calculate the field in this point. Please note, it is necessary to exclude a tiny volume around the point of singularity $\mathbf{r}$ (where G has a pole) since otherwise we could not apply Green's second identity. At the end we will consider the limit of this vanishing volume going to zero and arrive at an expression for the field valid everywhere in the volume.

This allows to write $(\Delta + k_0^2 n^2) G(\mathbf{r}, \mathbf{r}') = 0$ which eventually suggests that the left hand side of the Green's second identify vanishes and only the left hand side holds

$$\iint_S \left(u(\mathbf{r}') \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}, \mathbf{r}') \frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} \right) d^2 r' = 0$$

Separating the integral into the two surfaces for the outer boundary and the inner boundary leads to

$$-\iint_{S_\epsilon} \left(u(\mathbf{r}') \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}, \mathbf{r}') \frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} \right) d^2 r'$$

$$= \iint_S \left(u(\mathbf{r}') \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}, \mathbf{r}') \frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} \right) d^2 r'$$

Now we can plug in the Green's function we know $G(\mathbf{r}, \mathbf{r}') = \frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}$ into these expressions. On the outer surface we have

$$\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} = \cos\left(\angle(\mathbf{n}, \mathbf{r}' - \mathbf{r})\right) \left(ik_0 n - \frac{1}{|\mathbf{r}' - \mathbf{r}|} \right) \frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}' - \mathbf{r}|}$$

The $\cos\left(\angle(\mathbf{n}, \mathbf{r}' - \mathbf{r})\right)$ here represents the cosine of the angle between the outward normal and the $\mathbf{r}' - \mathbf{r}$

For the inner surface we can take advantage of the spherical shape of S_ϵ that has a constant radius $\epsilon = |\mathbf{r}' - \mathbf{r}|$.

$$G(\mathbf{r}, \mathbf{r}') = \frac{e^{ik_0 n\epsilon}}{\epsilon}$$

and for the derivative we get with $\cos\left(\angle(\mathbf{n}, \mathbf{r}' - \mathbf{r})\right) = -1$

$$\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} = \left(\frac{1}{\epsilon} - ik_0 n\right)\frac{e^{ik_0 n\epsilon}}{\epsilon}$$

Together we obtain

$$\iint_{S_\epsilon} \left(u(\mathbf{r}')\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}, \mathbf{r}')\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}}\right) d^2 r'$$

$$= \iint_{S_\epsilon} \left(u(\mathbf{r}')\left(\frac{1}{\epsilon} - ik_0 n\right)\frac{e^{ik_0 n\epsilon}}{\epsilon} - \frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}}\frac{e^{ik_0 n\epsilon}}{\epsilon}\right) d^2 r'$$

Now, by choosing ϵ increasingly small, the integrals over these functions essentially become the area of a sphere × $u(\mathbf{r})$ and the area of sphere × $\frac{\partial u(\mathbf{r})}{\partial \mathbf{n}}$ by the first mean value theorem for integration. Therefore, we obtain

$$\iint_{S_\epsilon} \left(u(\mathbf{r}')\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}, \mathbf{r}')\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}}\right) d^2 r'$$

$$\simeq 4\pi\epsilon^2 \left(u(\mathbf{r})\left(\frac{1}{\epsilon} - ik_0 n\right)\frac{e^{ik_0 n\epsilon}}{\epsilon} - \frac{\partial u(\mathbf{r})}{\partial \mathbf{n}}\frac{e^{ik_0 n\epsilon}}{\epsilon}\right)$$

Here, only the first term survives in the limit of $\epsilon \to 0$. Hence:

$$\iint_{S_\epsilon} \left(u(\mathbf{r}')\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} - G(\mathbf{r}, \mathbf{r}')\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}}\right) d^2 r' = 4\pi u(\mathbf{r})$$

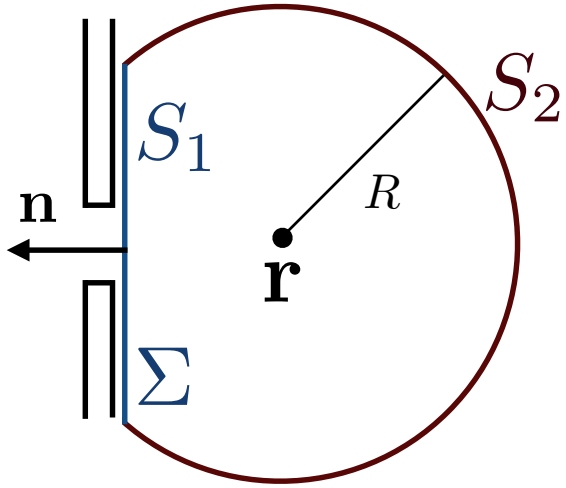
All together this finally gives the theorem of Helmholtz and Kirchhoff

$$u(\mathbf{r}) = \frac{1}{4\pi}\iint_S \left(\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}' - \mathbf{r}|}\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - u(\mathbf{r}')\frac{\partial}{\partial \mathbf{n}}\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}' - \mathbf{r}|}\right) d^2 r'$$

which is just the desired result.

2.4 Kirchhoff formulation

Now what we would like to do in the following is to approach towards a concise connection between the different approaches presented before; the angular spectrum method that requires information on the field in an initial plane and the Helmholtz and Kirchhoff formulation, which requires information on the field and its derivative on a surface enclosing the point of interest.



This is done in the following by separating the solution above to a sum of two integrals reflecting the following geometrical situation. The first is an integration across the screen, the second is an integration across a half sphere for which essentially we consider the limit of $R \to \infty$. The basic idea now consists in considering the sphere with an infinite large radius for which its surface area is in the order of R^2. Now, in the integrand we can then neglect those contributions that decay faster than R^{-2} since their contribution will be vanishing. For the modulus of the Green's function we have

$$|G(\mathbf{r},\mathbf{r}')| = \left|\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}\right| = \left|\frac{1}{|\mathbf{r}'-\mathbf{r}|}\right| = \frac{1}{R}$$

From a physical point of view (and it can be proven on a more solid base as well) it is reasonable to assume that the field $u(\mathbf{r})$ equally decays comparable to a spherical wave. Therefore, we will make a Taylor expansion of the outward normal derivatives that appear in the expression and only retain those terms that decay as R^{-1}. Other terms will not contribute to the integral. This suggests that

$$\frac{\partial}{\partial \mathbf{n}}\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|} = \left(ik_0 n - \frac{1}{R}\right)\frac{e^{ik_0 nR}}{R} = ik_0 nG(\mathbf{r},\mathbf{r}') + \mathcal{O}(R^{-2}) \overset{R\to\infty}{\cong} ik_0 nG(\mathbf{r},\mathbf{r}')$$

Inserting this expression into the integral for only the semi-spherical surface we obtain

$$u(\mathbf{r}) = \frac{1}{4\pi}\iint_{S_2}\left(G(\mathbf{r},\mathbf{r}')\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - u(\mathbf{r}')ik_0 nG(\mathbf{r},\mathbf{r}')\right)d^2r'$$

by expressing the integral in terms of a solid angle

$$u(\mathbf{r}) = \frac{1}{4\pi}\iint_{S_2} d\Omega G(\mathbf{r},\mathbf{r}')\left(\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - ik_0 nu(\mathbf{r}')\right)R^2$$

$$\xrightarrow{R\to\infty} u(\mathbf{r}) = \frac{1}{4\pi}\iint_{S_2} d\Omega\, R\left(\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - ik_0 nu(\mathbf{r}')\right)$$

The integral vanishes if we require that

$$\lim_{R\to\infty} R\left(\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - ik_0 nu(\mathbf{r}')\right) = 0$$

which is known as the Sommerfeld radiation condition. It is also known as the outgoing wave condition since outward propagating fields have to obey this conditions. Please note, this is a condition which not every field satisfies but which we would like to enforce while discussing possible solutions to the wave equations. With that assumption the field eventually reads as

$$u(\mathbf{r}) = \frac{1}{4\pi}\iint_{S_1}\left(\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - u(\mathbf{r}')\frac{\partial}{\partial \mathbf{n}}\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}\right)d^2r'$$

Now Kirchhoff made two further assumptions. While considering the field behind finite apertures inside otherwise opaque screens, he assumes at first that the field behind the aperture (indicated by the area Σ) is the same as the incident field (propagating from $z = -\infty$ into the positive direction) in the absence of the screen. Moreover, outside the geometrical area of the aperture the field and its derivative are exactly equal to zero.

Please note, this corresponds to what people would call a thin element approximation. There, the field behind an object is given by the field before the object just multiplied with a transmission function. The transmission function can be complex valued (e.g. for objects as lenses only the phase is affected) but might also real valued (for gray filters affecting the amplitude only). It might be also be binary

with values being either zero or one, while considering amplitude masks or; apertures in screens.

$$u(\mathbf{r}) = \frac{1}{4\pi}\iint_{\Sigma}\left(\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - u(\mathbf{r}')\frac{\partial}{\partial \mathbf{n}}\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}\right)d^2r'$$

Or in a more general term leaving the choice of the Green's function open

$$u(\mathbf{r}) = \frac{1}{4\pi}\iint_{\Sigma}\left(G(\mathbf{r},\mathbf{r}')\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - u(\mathbf{r}')\frac{\partial G(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}}\right)d^2r'$$

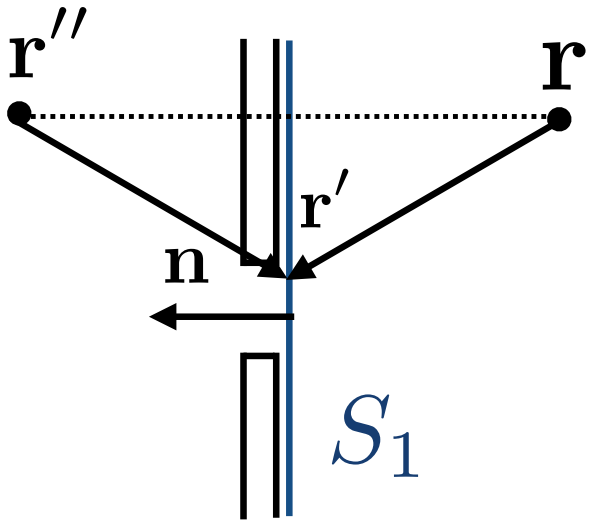
The only requirements for this equation to hold are the applicability of scalar theory, the requirement that $u(\mathbf{r})$ and $G(\mathbf{r},\mathbf{r}')$ satisfy the homogeneous wave scalar wave equation, the satisfaction of the Sommerfeld radiation condition.

2.5 Rayleigh-Sommerfeld

The Kirchhoff theory provides remarkably accurate results and is widely used. However, it also has mathematical inconsistencies. The difficulty is linked to the fact that it is required to know both the field and its normal derivative on the boundary of the screen. Particularly they are set to be zero outside the aperture of interest. Now, in the context of potential theory in mathematics it is known that if a two-dimensional potential function and its normal derivative vanish together along any finite curve segment, then that potential function must vanish over the entire plane. Similarly, if a solution of the three-dimensional wave equation vanishes on any finite surface element, it must vanish everywhere in space. Starting from the equation above we can ask ourselves how to mitigate this problem?

$$u(\mathbf{r}) = \frac{1}{4\pi}\iint_{\Sigma}\left(G(\mathbf{r},\mathbf{r}')\frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} - u(\mathbf{r}')\frac{\partial G(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}}\right)d^2r'$$

The obvious way is to modify the Green's function such that the development leading to the above equation remains valid, but in addition either $G(\mathbf{r},\mathbf{r}')$ or $\frac{\partial G(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}}$ shall vanish across the entire surface S_1. This would remove the requirement on both the field and its normal derivative. This was done by Sommerfeld. Depending on whether the first or the second possible condition mentioned above is fulfilled, we will approach two different Rayleigh-Sommerfeld formulations.



We would like to assume that the Green's function is not just generated from a point in $\mathbf{r}$ but we may add a second Green's function that is located in the same x and y coordinate but which emerges from a spatial point at $-z$. This approach is motivated by similar approaches in the context of mirror charges.

$$G_{\pm}(\mathbf{r},\mathbf{r}') = \frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|} \pm \frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}''|}}{|\mathbf{r}'-\mathbf{r}''|}$$

with $\mathbf{r}'' = (x, y, -z)$. These Green's functions are solutions to same differential equations in the volume of interest; hence they are valid. The response on the left side of the screen is of no importance since this is not relevant in the present context. The Green's functions have the useful properties that $G_{-}(\mathbf{r},\mathbf{r}') = 0$ for $\mathbf{r}' \in \mathbf{S}_1$ and $\frac{\partial G_{+}(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}} = 0$ for $\mathbf{r}' \in \mathbf{S}_1$. Using these two Green's functions we obtain the 1st Rayleigh-Sommerfeld integral as

$$u_1(\mathbf{r}) = -\frac{1}{4\pi}\iint_{\Sigma} u(\mathbf{r}')\frac{\partial G_{-}(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}} d^2 r'$$

We can specify the solution by calculating the normal derivative of $G_{-}(\mathbf{r},\mathbf{r}')$.

$$\frac{\partial G_{-}(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}} = \cos\big(\angle(\mathbf{n},\mathbf{r}'-\mathbf{r})\big)\left(ik_0 n - \frac{1}{|\mathbf{r}'-\mathbf{r}|}\right)\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}$$

$$-\cos\big(\angle(\mathbf{n},\mathbf{r}'-\mathbf{r}'')\big)\left(ik_0 n - \frac{1}{|\mathbf{r}'-\mathbf{r}''|}\right)\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}''|}}{|\mathbf{r}'-\mathbf{r}''|}$$

Now we have $|\mathbf{r}'-\mathbf{r}| = |\mathbf{r}'-\mathbf{r}''|$ and $\cos\big(\angle(\mathbf{n},\mathbf{r}'-\mathbf{r})\big) = -\cos\big(\angle(\mathbf{n},\mathbf{r}'-\mathbf{r}'')\big)$. Therefore we can combine those expressions to

$$\frac{\partial G_{-}(\mathbf{r},\mathbf{r}')}{\partial \mathbf{n}} = 2\cos\big(\angle(\mathbf{n},\mathbf{r}'-\mathbf{r})\big)\left(ik_0 n - \frac{1}{|\mathbf{r}'-\mathbf{r}|}\right)\frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}'-\mathbf{r}|}$$

Now, for the important approximation that the distance of the point of interest relative to the screen is much larger than the wavelength ($|\mathbf{r}' - \mathbf{r}| \gg \lambda$), we can drop the second term above and obtain

$$\frac{\partial G_{-}(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} = 2ik_0 n \cos\big(\angle(\mathbf{n}, \mathbf{r}' - \mathbf{r})\big) \frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}' - \mathbf{r}|}$$

This is just twice the normal derivative of the Green's function $G(\mathbf{r}, \mathbf{r}')$ used in the Kirchhoff formula.

$$u_1(\mathbf{r}) = -\frac{1}{2\pi} \iint_{\Sigma} u(\mathbf{r}') \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial \mathbf{n}} d^2 r'$$

$$u_1(\mathbf{r}) = \frac{n}{i\lambda_0} \iint_{\Sigma} u(\mathbf{r}') \frac{e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}}{|\mathbf{r}' - \mathbf{r}|} \cos\big(\angle(\mathbf{n}, \mathbf{r}' - \mathbf{r})\big)\, d^2 r'$$

This expression is known as the first Rayleigh-Sommerfeld diffraction formula. The same procedure can be performed for the different choice of the Green's function, leading to the 2nd Rayleigh-Sommerfeld solution as

$$u_2(\mathbf{r}) = \frac{1}{4\pi} \iint_{\Sigma} \left(G_{+}(\mathbf{r}, \mathbf{r}') \frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} \right) d^2 r'$$

$$u_2(\mathbf{r}) = \frac{1}{2\pi} \iint_{\Sigma} \left(G(\mathbf{r}, \mathbf{r}') \frac{\partial u(\mathbf{r}')}{\partial \mathbf{n}} \right) d^2 r'$$

2.6 Fresnel approximation

The equations above are useful but can be often simplified under some assumption. Please note, the contents of the following two chapters has been already to some extent discussed in the lecture "Fundamentals of Optics and Photonics" and the following shall only serve as a reminder.

At first we would like to assume that $x', y' \ll z$ and $x, y \ll z$. This suggests at first the cosine in the equation above can be approximated as $\cos\big(\angle(\mathbf{n}, \mathbf{r}' - \mathbf{r})\big) \approx 1$. A back of the envelope approximation says that this is correct to within 5% for angles smaller than 20°, which is optically pretty large.

Moreover, $\frac{1}{|\mathbf{r}'-\mathbf{r}|}$ can be safely approximated as $\frac{1}{|\mathbf{r}'-\mathbf{r}|} \approx \frac{1}{z}$ and it is a term that will largely affect the amplitude. However, how to approximate the phase term

$e^{ik_0 n|\mathbf{r}'-\mathbf{r}|}$? We will assume in the following that $z' = 0$, hence putting the screen into the center of the coordinate system. With

$$|\mathbf{r}' - \mathbf{r}|^2 = z^2 + (x' - x)^2 + (y' - y)^2$$

we can expand the square-root into a Taylor series of second order

$$k_0 n|\mathbf{r}' - \mathbf{r}| = k_0 nz\sqrt{1 + \left(\frac{x' - x}{z}\right)^2 + \left(\frac{y' - y}{z}\right)^2}$$

$$\cong k_0 nz\left[1 + \frac{1}{2}\left(\frac{x' - x}{z}\right)^2 + \frac{1}{2}\left(\frac{y' - y}{z}\right)^2\right]$$

Plugging this into the first Rayleigh-Sommerfeld diffraction formula simplifies the expression to

$$u_{\text{Fresnel}}(\mathbf{r}) = \frac{ne^{ik_0 nz}}{i\lambda_0 z}\iint_{\Sigma} u(\mathbf{r}')e^{i\frac{k_0 n}{2z}[(x\prime - x)^2 + (y\prime - y)^2]}d^2r'$$

This is known as the Fresnel formula. Here, the expansion of the square root still depend on $x', x'^2, y', \text{and } y'^2$.

The advantage of the formula only pays of if we slightly reformulate the integral such that it will appear as a Fourier transform of the initial field multiplied with a suitable pre-factor. In particular we have

$$u_{\text{Fresnel}}(\mathbf{r}) = \frac{ne^{ik_0 nz}e^{i\frac{k_0 n(x^2+y^2)}{2\quad z}}}{i\lambda_0 z}\iint_{\Sigma}\left[u(\mathbf{r}')e^{i\frac{k_0 n(x\prime^2+y\prime^2)}{2\quad z}}\right]e^{-2\pi i\left(\frac{nx}{\lambda_0 z}x' + \frac{ny}{\lambda_0 z}y'\right)}d^2r'$$

where $\frac{nx}{\lambda_0 z}$ and $\frac{ny}{\lambda_0 z}$ are the spatial frequencies.

Please note in the Fourier space this approximation corresponds to assuming that $\alpha^2 + \beta^2 \ll k^2$ which allows to expand the square root for the propagation constant into a Taylor series and retaining only the two lowest orders in the expansion, i.e.

$$e^{i\gamma(\alpha,\beta)z} \approx e^{ik_0 nz}e^{-i\frac{z}{2k_0 n}(\alpha^2+\beta^2)}$$

This equation describes a spherical wave in paraxial approximation. Starting from the original expression obtained in the section on the angular spectrum method for the field in terms of its plane wave spectrum

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

we can express the field in Fresnel approximation as

$$u_{\text{Fresnel}}(x,y,z_B) = \iint_{-\infty}^{\infty} U_+(\alpha,\beta;z_A) e^{ik_0 n z_B} e^{-i\frac{z_B}{2k_0 n}(\alpha^2+\beta^2)} e^{i(\alpha x+\beta y)} d\alpha d\beta$$

where the Fourier spectrum behind the aperture at z_A is given by $U_+(\alpha,\beta;z_A)$ and we propagate the field towards z_B.

2.7 Fraunhofer approximation

The last possible approximation is the far-field approximation. This imposes a further approximation to the square root in the equation for the Fresnel approximation developed above by neglecting the terms that depend on x'^2 and y'^2.

$$k_0 n|\mathbf{r}'-\mathbf{r}| \cong k_0 nz + \frac{k_0 n}{2}\frac{(x^2+y^2)}{z} - k_0 n\frac{(xx'+yy')}{z} + \frac{k_0 n}{2}\frac{(x'^2+y'^2)}{z}$$

$$\rightarrow \qquad k_0 n|\mathbf{r}'-\mathbf{r}| \cong k_0 nz + \frac{k_0 n}{2}\frac{(x^2+y^2)}{z} - k_0 n\frac{(xx'+yy')}{z}$$

$$u_{\text{Fraunhofer}}(\mathbf{r}) = \frac{ne^{ik_0 nz + i\frac{k_0 n}{2}\frac{(x^2+y^2)}{z}}}{i\lambda_0 z}\iint_{\Sigma} u(\mathbf{r}')e^{-2\pi i\left(\frac{nx}{\lambda_0 z}x'+\frac{ny}{\lambda_0 z}y'\right)} d^2r'$$

Now this is just the Fourier transform using

$$U\left(k\frac{x}{z}, k\frac{y}{z}\right) = \left(\frac{1}{2\pi}\right)^2 \iint_{\Sigma} u(\mathbf{r}')e^{-i\left(k\frac{x}{z}x'+k\frac{y}{z}y'\right)} d^2r'$$

$$u_{\text{Fraunhofer}}(\mathbf{r}) = \frac{n(2\pi)^2}{i\lambda_0 z} e^{ik_0 nz + i\frac{k_0 n}{2}\frac{(x^2+y^2)}{z}} U\left(k\frac{x}{z}, k\frac{y}{z}\right)$$

Here, it is just important to memory that the field in the far-field corresponds to the Fourier transform of the transmission function. Applications will be discussed in the seminar.

2.8 Method of stationary phase

The method of stationary phases is in general a method to calculate integrals that read as

$$I = \iint g(p,q)e^{i\kappa f(p,q)} dpdq$$

in an approximate manner whenever $\kappa \gg 1$ and if $g(p,q)$ varies slowly. As we will argue below, these approximations hold in specific regimes for the Fresnel integral

which lead to the Fraunhofer expressions if the method of stationary phases is applied.

The idea for the application of the method of stationary phase is the following. If $\kappa \gg 1$ the exponent in the integral oscillates rapidly even if there is only a small change in the variation of $f(p,q)$. It is simply the product of the two that counts. If on the scales where the exponent varies rapidly the function $g(p,q)$ varies only slowly, the different contribution in the integral compensate each other. There is however one exception. At the stationary points characterized by the coordinates p_m and q_m for which it holds that

$$\left.\frac{\partial f}{\partial p}\right|_{p_m,q_m} = f_p\big|_{p_m,q_m} = 0$$

and simultaneously

$$\left.\frac{\partial f}{\partial q}\right|_{p_m,q_m} = f_q\big|_{p_m,q_m} = 0$$

the integral will not vanish. Instead, the integral can be approximated as a sum across the stationary points

$$I = \frac{2\pi}{i\kappa}\sum_{m=1}^{M}\frac{1}{\sqrt{f_{pp(m)}f_{qq(m)} - \frac{1}{4}f_{pq(m)}^2}}g(p_m,q_m)e^{i\kappa f(p_m,q_m)}$$

The exact derivation of this expression shall be of no concern here and can be done in the exercises. Now we wish to apply the method of stationary phases to the Fresnel integral using the angular spectrum representation. The integral is slightly reformulated to simplify the identification of the link to

$$u_{\mathrm{Fresnel}}(x,y,z_B) = e^{ik_0nz_B}\iint_{-\infty}^{\infty}U_+(\alpha,\beta;z_A=0)e^{ik_0nz_B\left[\left(\frac{\alpha}{k_0nz_B}\frac{x}{}+\frac{\beta}{k_0nz_B}\frac{y}{}\right)-\frac{1}{2}\left(\frac{\alpha^2}{k_0^2n^2}+\frac{\beta^2}{k_0^2n^2}\right)\right]}d\alpha d\beta$$

Now we make the substitutions $p = \frac{\alpha}{k_0n}$, $q = \frac{\beta}{k_0n}$ and $\kappa = k_0nz_B$ from which we require that $\kappa \gg 1$. We can identify the function $f(p,q)$ as

$$f(p,q) = p\frac{x}{z_B} + q\frac{y}{z_B} - \frac{1}{2}(p^2+q^2)$$

This allows to write the Fresnel integral as

$$u_{\mathrm{Fresnel}}(x,y,z_B) = k_0^2n^2e^{i\kappa}\iint_{-\infty}^{\infty}U_+(k_0np,k_0nq;z_A=0)e^{i\kappa f(p,q)}dpdq$$

The derivatives we need to calculate in the expressions above and to calculate the stationary points are

$$\frac{\partial f}{\partial p} = \frac{x}{z_B} - p, \frac{\partial f}{\partial q} = \frac{y}{z_B} - q, \frac{\partial^2 f}{\partial p^2} = \frac{\partial^2 f}{\partial q^2} = -1, \ \frac{\partial^2 f}{\partial p \partial q} = 0$$

The only stationary point is then at

$$p_1 = \frac{x}{z_B} \text{ and } q_1 = \frac{y}{z_B}$$

and the function value $f(p_1, q_1)$ itself is then

$$f(p_1, q_1) = \frac{x^2 + y^2}{2z_B}$$

Inserting this into the approximate expression for the integral

$$I = \frac{2\pi}{i\kappa} \sum_{m=1}^{M} \frac{1}{\sqrt{f_{pp(m)} f_{qq(m)} - \frac{1}{4} f_{pq(m)}^2}} g(p_m, q_m) e^{i\kappa f(p_m, q_m)}$$

gives exactly the expression for the field in Fraunhofer approximation

$$u_{\text{Fraunhofer}}(x, y, z_B) = \frac{n(2\pi)^2}{i\lambda_0 z_B} e^{ik_0 n z_B} U_+ \left(k\frac{x}{z_B}, k\frac{y}{z_B}; z_A = 0 \right) e^{i\frac{k_0 n}{2z_B}(x^2 + y^2)}$$

Again, we see that the amplitude distribution of the diffraction pattern corresponds to the Fourier transform of the field behind the screen (inside the aperture). The interpretation is such that at any spatial position (x, y) only the spatial frequency with $\left(\alpha = k\frac{x}{z_B}, \beta = k\frac{y}{z_B}\right)$ and its respective spectral amplitude $U_+ \left(k\frac{x}{z_B}, k\frac{y}{z_B}; z_A = 0 \right)$ contribute to the diffraction pattern. The contribution of all other plane waves cancels out by interference.

To make this approximation applicable, we require that the ratio of aperture size with respect to wavelength is small (the angular spectrum varies then smoothly with the angular frequency) and the ratio of aperture to propagation distance should be small as well (then $\kappa \gg 1$). This condition is expressed in terms of the Fresnel number that should be smaller than one tenths

$$N_F = \frac{a}{\lambda} \frac{a}{z_B} \lesssim 0.1$$

2.9 Basics of holography

As we have seen in the previous chapters, the full visual information concerning an object is encoded in the field distribution in a given plane (assuming that the wave satisfies suitable radiation boundary condition). This requires information on the field concerns both amplitude and the phase. This complex field propagates in space and evolves and eventually can be perceived by an external observer, e.g. by forming an image with an eye. The idea of holography is to record the field emanating from an object and to reconstruct at a later stage for other purposes. In modern times, we also use computer-generated holograms to express wave fields from objects that do not exist in reality. The method has been used to cause many interesting artistic effects but there are also many industrial applications. Examples are the metrology of complicated lenses whose generated wave fronts are interferometrically compared to wave fronts from perfect lenses that have been either generated in a computer or that were recorded once from a perfect device. The challenge in holography is the reconstruction of the phase of the field emanated from such objects. In 1948 Dennis Gabor developed a method that allows recording the information on the wave field using a coherent reference wave in a medium that is in turn only responsive to the intensity. The method was such ground breaking that Dennis Gabor was awarded the Nobel prize in physics in 1971. Holography in general is a two-step process. First, a recording takes place. Second, a reconstruction takes place.

2.9.1 Reconstruction of a hologram

To start with, we consider a monochromatic scalar field (also called the primary field) $u(\mathbf{r})$, which is emanated by an object. Please don't think too mysterious here. You simply use a coherent light source, illuminate the object of interest, and the scattered field from the object will contain all the information from it. Eventually, this is just the way we perceive our surrounding. In addition we consider a reference wave that is for simplicity here a plane wave $u_0(\mathbf{r}) = a_0 e^{i\mathbf{k}_0\cdot\mathbf{r}}$. Now we define as ξ the coordinates along some surface (for simplicity a plane) in which we wish to place later our responsive recording material. The intensity in this plane results from the superposition of the object and the reference plane

$$I(\xi) = |u(\xi) + u_0(\xi)|^2 = |a_0|^2 + a_0^* u(\xi) e^{-i\mathbf{k}_0 \cdot \xi} + a_0 u^*(\xi) e^{i\mathbf{k}_0 \cdot \xi} + |u(\xi)|^2$$

Under the assumption that the intensity of the reference wave is much larger than the wave emanated from the object, the last term in the expression above can be neglected for simplicity. By placing a traditional recording media made of a thin film that blackens proportional to the local intensity into the considered plane and developing afterwards the film, we obtain a filter that has a local transmission function corresponding to the recorded intensity

$$t(\xi) = 1 - \gamma I(\xi)$$

where γ is here some sort of contrast function that is obviously depends on the used material and on the recording conditions. Illuminating this filter with a reconstruction wave $u_1(\mathbf{r}) = a_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}}$ that corresponds to the reference wave used for the recording (albeit with an arbitrary amplitude and propagation direction at this moment), the field behind the filter is given by

$$w(\xi) = t(\xi) u_1(\xi)$$

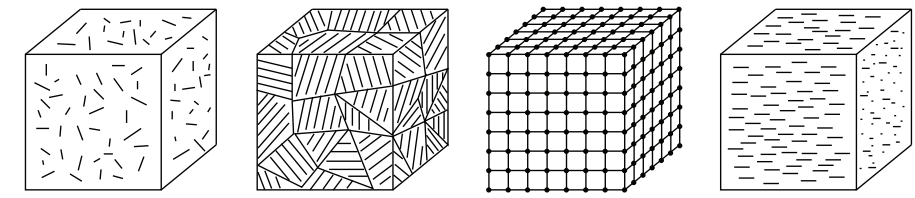
$$= a_1(1 - \gamma |a_0|^2) e^{i\mathbf{k}_1 \cdot \xi} - \gamma a_1 a_0^* u(\xi) e^{i(\mathbf{k}_1 - \mathbf{k}_0) \cdot \xi} - \gamma a_1 a_0 u^*(\xi) e^{i(\mathbf{k}_1 + \mathbf{k}_0) \cdot \xi}$$

These are obviously three different terms that can be best understood, while setting $\mathbf{k}_1 = \mathbf{k}_0$.

- The first terms corresponds to a plane wave that propagates into the direction of the reconstruction wave. This wave is a consequence of the possible imperfect modulation of the recorded hologram, i.e. γ is smaller than unity. This contribution can be engineered to be sufficiently small, e.g. by a suitable engineering of the hologram or actually by Fourier filter techniques.
- The second term corresponds up to a constant factor to the field emanated by the considered object $u(\xi)$. This is the field one is usually interested.
- The third term is the complex conjugate of the object field $u(\xi)$ and it propagates into the direction $e^{i2\mathbf{k}_0 \cdot \xi}$

3. Optics in anisotropic materials

Thus far, we were only concerned with homogenous local, dispersive and isotropic material. We saw eventually that the normal modes of a space filled with such material are elliptically polarized plane waves. Understanding the subtleties of the dispersion relation is the key to describe on a sound theoretical base the propagation of light in such materials. Although many materials obey these requirements of isotropy (glasses, air, water, metals), many materials are not isotropic but anisotropic. They offer unique opportunities in the context of many applications. To make use of them, it is therefore important to understand the light propagation in such materials. In the following we discuss the basics of the light propagation in such materials. Again, we are primarily concerned with the discussion of the normal modes, i.e., their field distribution as well as the governing dispersion relation.



3.1. Susceptibility and permittivity tensor

Thus far we considered isotropic materials. Their optical properties are independent on the direction of light propagation and the direction of the electric field polarization that is considered as illumination. In many optical materials, especially crystals, the induced polarization, however, depends on the direction of the electric field. Understanding the polarization for bound charges in terms of a harmonic oscillator, it suggests that the amplitude of the displacement shall then depend on the direction of the electric field. This eventually reflects the symmetry of the crystals.

Examples for anisotropic materials and possible applications are

Lithiumniobate ($LiNbO_3$) $\rightarrow$ electro-optical material

Quartz	→ polarizers
Liquid crystals	→ display elements, nonlinear optics
Multiple quantum wells	→ optoelectronics

We will stick to a few assumptions in the following that shall simplify our lives:

- We will consider only a single frequency in time (monochromatic field) and space (plane wave)
- We will consider materials to be free of absorption (otherwise expansion into normal modes wouldn't make sense in each case)

For an isotropic material the normal modes were elliptically polarized, plane, monochromatic waves. The functional dependency of the polarization and the electric displacement have been obeying the following equations:

$$\bar{\mathbf{P}}(\mathbf{r},\omega) = \varepsilon_0 \chi(\omega) \bar{\mathbf{E}}(\mathbf{r},\omega)$$
$$\bar{\mathbf{D}}(\mathbf{r},\omega) = \varepsilon_0 \varepsilon(\omega) \bar{\mathbf{E}}(\mathbf{r},\omega)$$

In the following, we will drop the bar above the field vectors since we consider in each case only monochromatic fields: $\bar{\mathbf{E}}(\mathbf{r},\omega) \rightarrow \mathbf{E}(\mathbf{r},\omega)$. Now, in the case of anisotropic materials where the induced polarization will depend on the magnitude of the electric field in the respective direction we have

$$P_i(\mathbf{r},\omega) = \varepsilon_0 \sum_{j=1}^{3} \chi_{ij}(\omega)\, E_j(\mathbf{r},\omega) \overset{\text{def}}{=} \varepsilon_0 \chi_{ij}(\omega) E_j(\mathbf{r},\omega)$$

Here, we implicitly assume that the sum convention according to Einstein applies (summation over double indices). Please note, the susceptibility here is a 2nd rank tensor $\hat{\boldsymbol{\chi}}(\omega)$ and $\chi_{ij}(\omega)$ are the components of the tensor in a specific coordinate system.

Important in this equation is:

- $\mathbf{P} \nparallel \mathbf{E}$, the induced polarization is not anymore parallel to the electric field
- The susceptibility tensor reflects the crystal symmetry, but the light will not probe for the actual arrangement of the atom due to largely disparate length scales. A typical wavelength is 500 nm whereas the lattice spacing of the crystal is rather in the order of 0.5 nm, the field therefore doesn't probe the properties of the periodic arrangement at the level of individual atoms; but the field remains to be sensitive against the symmetry.

Analog equations and considerations hold for the electric displacement field

$$D_i(\mathbf{r},\omega) = \varepsilon_0 \varepsilon_{ij}(\omega) E_j(\mathbf{r},\omega)$$

$$\mathbf{D}(\mathbf{r},\omega) = \varepsilon_0 \hat{\boldsymbol{\varepsilon}}(\omega)\mathbf{E}(\mathbf{r},\omega)$$

- $\mathbf{D} \nparallel \mathbf{E}$
- $\hat{\boldsymbol{\varepsilon}} = (\varepsilon_{ij})$ is the permittivity tensor

A last important quantity, which we will use in the following occasionally, is the inverse permittivity tensor $\hat{\boldsymbol{\sigma}} = (\hat{\boldsymbol{\varepsilon}})^{-1} = (\sigma_{ij})$, leading to $\sigma_{ij}(\omega) D_j(\mathbf{r},\omega) = \varepsilon_0 E_i(\mathbf{r},\omega)$.

What are the properties of the tensors? For simplicity, we drop in the following the frequency dependency in the arguments

$\sigma_{ij}, \varepsilon_{ij}$ $\rightarrow$ real valued in the transparency region (which we are interested in here)

$\rightarrow$ tensors are symmetric, leaving only six independent components $\sigma_{ij} = \sigma_{ji}$ and $\varepsilon_{ij} = \varepsilon_{ji}$, moreover the matrices are Hermitian, suggesting that the element in the *i*-th row and *j*-th column is equal to the complex conjugate of the element in the *j*-th row and *i*-th column, for all indices *i* and *j*

Proof:

Considering source-free Maxwell's equations in the Fourier space:

$$\mathbf{rot}\,\mathbf{E} = i\omega\mu_0 \mathbf{H}$$

$$\mathbf{rot}\,\mathbf{H} = -i\omega\varepsilon_0 \hat{\boldsymbol{\varepsilon}}\mathbf{E}$$

The magnetic field can be eliminated from that equation to obtain the vector Helmholtz equation for the electric field:

$$\mathbf{rot}\,\mathbf{rot}\,\mathbf{E} - \omega^2 \varepsilon_0 \mu_0 \hat{\boldsymbol{\varepsilon}}\mathbf{E} = \mathbf{0}$$

It should be noted that, in general, the vector Helmholtz equation in anisotropic media is not equivalent to the component-by-component scalar Helmholtz equation even for a homogenous medium. The problem is that in the anisotropic case it does not follow from $\mathbf{div}\,\mathbf{D} = \mathbf{div}\,(\hat{\boldsymbol{\varepsilon}}\mathbf{E}) = \mathbf{0}$ that $\mathbf{div}\,\mathbf{E} = 0$.

Now, for the average over long time intervals for the divergence of the complex Poynting vector (the dissipated energy) it follows that

$$\langle \mathbf{div}\,\mathbf{S} \rangle = \frac{1}{2}\,\mathbf{div}\,(\mathbf{E} \times \mathbf{H}^*)$$

$$= \frac{1}{2}\mathbf{H}^* \cdot \mathbf{rot}\,\mathbf{E} - \frac{1}{2}\mathbf{E} \cdot \mathbf{rot}\,\mathbf{H}^*$$

$$= \frac{i\omega\mu_0}{2}|\mathbf{H}|^2 - \frac{i\omega\varepsilon_0}{2}\mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*)$$

$$= 2i\omega\langle w_m\rangle - \frac{i\omega\varepsilon_0}{2}\mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*)$$

where we introduced $\langle w_m\rangle$ as the time average of the magnetic energy density. In analogy to the case of an isotropic material, the second term on the right hand side of the equation should be equal to $-2i\omega\langle w_e\rangle$, eventually corresponding to the electric energy density for an anisotropic material

$$\langle w_e\rangle = \frac{\varepsilon_0}{4}\mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*)$$

On the other hand we have an expression for the time averaged electric energy density that reads as

$$\langle w_e\rangle = \frac{1}{4}\Re(\mathbf{E} \cdot \mathbf{D}^*)$$

$$\langle w_e\rangle = \frac{\varepsilon_0}{4}\Re\big(\mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*)\big)$$

$$\langle w_e\rangle = \frac{\varepsilon_0}{8}[\mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*) + \hat{\boldsymbol{\varepsilon}}\mathbf{E} \cdot \mathbf{E}^*]$$

Both expressions for the electrical energy density are only equal if and only if

$$\hat{\boldsymbol{\varepsilon}}\mathbf{E} \cdot \mathbf{E}^* = \mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*)$$

Since for two complex vectors $\mathbf{V}$ and $\mathbf{W}$ it always holds that

$$\hat{\boldsymbol{\varepsilon}}\mathbf{V} \cdot \mathbf{W}^* = \mathbf{V} \cdot \hat{\boldsymbol{\varepsilon}}^{\boldsymbol{H}}\mathbf{W}^*$$

where $\varepsilon_{ij}^H = \varepsilon_{ji}^*$ it follows that the left hand side can be written as

$$\mathbf{E} \cdot \hat{\boldsymbol{\varepsilon}}^{\boldsymbol{H}}\mathbf{E}^* = \mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*)$$

So it can be concluded that the permittivity tensor at least is equal to its transpose. Furthermore it can be shown that $\Im\big(\mathbf{E} \cdot (\hat{\boldsymbol{\varepsilon}}^*\mathbf{E}^*)\big) = 0$

$\rightarrow$ this should also hold for an arbitrary electric field

a) this holds for example also for the case that all components are zero except the x-component of the electric field $E_x \neq 0$

$$\Im\big(E_x \cdot (\varepsilon_{xx}^* E_x^*)\big) = |E_x|^2\Im(\varepsilon_{xx}^*) = 0$$

The same holds for ε_{yy} and ε_{zz}, $\rightarrow$ this suggests that all diagonal components of the tensor are real valued

b) For the other components it holds

$$\Im\left(\varepsilon_{xy}^{*}E_xE_y^{*}+\varepsilon_{yx}^{*}E_yE_x^{*}\right)=\Im\left(\varepsilon_{xy}^{*}E_xE_y^{*}-\varepsilon_{yx}E_y^{*}E_x\right)$$
$$=\Im\left(\left[\varepsilon_{xy}^{*}-\varepsilon_{yx}\right]E_xE_y^{*}\right)$$
$$=0$$

From that expression we can conclude that $\varepsilon_{xy}^{*}=\varepsilon_{yx}$. The derivation can be performed for all possible of-diagonal entries of the permittivity tensor, which shows that it is a hermitian tensor. Similar deviations can be made for σ_{ij} and χ_{ij}.

From now we assume that all tensors are real valued. However, this suggests by no means that of-diagonal elements are always real-valued. They are important, e.g., in the context of magneto-optics and the materials are called gyrotropic materials.

A further important concept is the idea of a principle coordinate system in which the tensors are diagonal. This is generally possible for all Hermitian matrices. This will be shown in the following at the example of the inverse permittivity tensor. Only in this coordinate system it will hold that $\mathbf{D} \parallel \mathbf{E}$. Mathematically, this can be expressed in terms of an eigenvalue equation. It is required that

$$\varepsilon_0 E_i=\sigma_{ij}D_j \doteq \lambda D_i$$

This requires a non-trivial solution to the equation, which asks for a vanishing determinant

$$\det\left[\sigma_{ij}-\lambda I_{ij}\right]=0$$

where the identity matrix has been introduced as $I_{ij}=\delta_{ij}$.

This is an equation of 3rd order that has three roots as solutions, which are called in the following $\lambda^{(\alpha)}$. The associated eigenvectors read as

$$\sigma_{ij}D_j^{(\alpha)}=\lambda^{(\alpha)}D_i^{(\alpha)}$$

The eigenvectors are orthogonal because

$$\sigma_{ij}D_j^{(\alpha)}=\lambda^{(\alpha)}D_i^{(\alpha)} \text{ and } \sigma_{ij}D_j^{(\beta)}=\lambda^{(\beta)}D_i^{(\beta)}$$

If we multiply both equations with the respective other eigenvector and subtract both we obtain

$$D_i^{(\beta)}\sigma_{ij}D_j^{(\alpha)}-D_i^{(\alpha)}\sigma_{ij}D_j^{(\beta)}=\left(\lambda^{(\alpha)}-\lambda^{(\beta)}\right)D_i^{(\alpha)}D_i^{(\beta)}$$
$$D_j^{(\beta)}\sigma_{ji}D_i^{(\alpha)}-D_i^{(\alpha)}\sigma_{ij}D_j^{(\beta)}=0$$

because $\sigma_{ji}=\sigma_{ij}$. Here it can be concluded that $D_i^{(\alpha)}D_i^{(\beta)}=0$ for all $\lambda^{(\alpha)}\neq\lambda^{(\beta)}$.

The principle axes now reflect the crystal symmetry.

$$\varepsilon_{ij} = \varepsilon_i \delta_{ij},\ \sigma_{ij} = \sigma_i \delta_{ij} = \frac{1}{\varepsilon_i}\delta_{ij}.$$

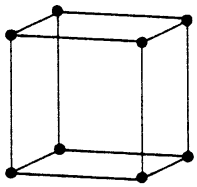
$$(\varepsilon_{ij}) = \begin{bmatrix} \varepsilon_1(\omega) & 0 & 0 \\ 0 & \varepsilon_2(\omega) & 0 \\ 0 & 0 & \varepsilon_3(\omega) \end{bmatrix}$$

An anisotropic material can be in most cases fully described using three different dielectric functions (in the principle coordinate system). In most cases the analysis is done in the principle coordinate system; although in the end a transformation back in the coordinate system of the laboratory is made; but the principle coordinate system makes everything much easier.

3.2 Optical classification of crystals

a) Isotropic

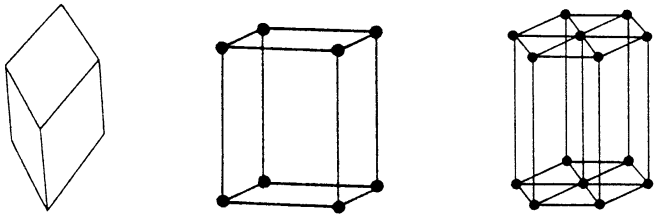
- Three crystallographic identical and perpendicular axes
- Cubic crystals (diamond, silicon)



- $\varepsilon_1(\omega) = \varepsilon_2(\omega) = \varepsilon_3(\omega)$
- It follows in that particular case $D_i = \varepsilon_0 \varepsilon E_i$
- The same type of descriptions applies for gases amorphous solid states, liquids and polycrystalline materials with grain sizes much smaller than the wavelength such that light on a mesoscopic scale will not resolve the fine structure

b) Uniaxial

- Two crystallographic equivalent directions
- Trigonal, tetragonal, hexagonal crystals



- $\varepsilon_1(\omega) = \varepsilon_2(\omega) \neq \varepsilon_3(\omega)$

c) biaxial

- there are no two identical crystallographic directions
- orthorhombic, monocline, tricline

- $\varepsilon_1(\omega) \neq \varepsilon_2(\omega) \neq \varepsilon_3(\omega)$

3.3. The index ellipsoid

The index ellipsoid is a geometrical representation of the inverse dielectric tensor. The defining equation for the index ellipsoid is a surface of second order. The advantage of the representation is its invariance against coordinate transformations. Its is defined as

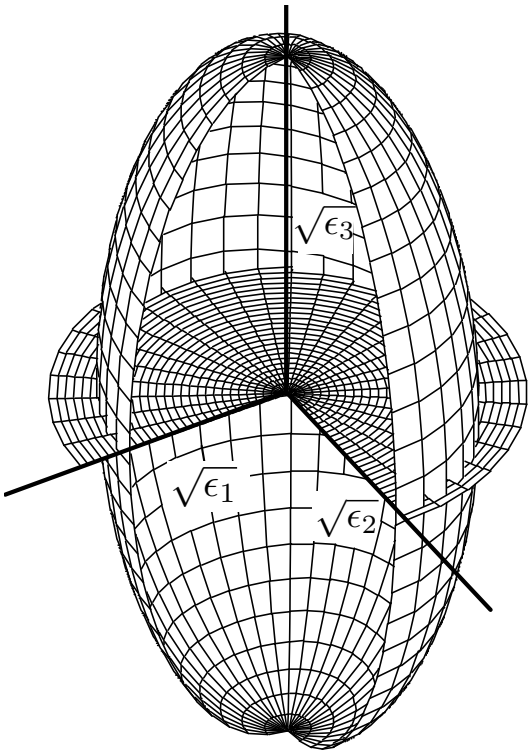
$$\sum_{i,j=1}^{3} \sigma_{ij}\, x_i x_j = 1$$

The index ellipsoid defines a surface of constant electric energy density since it can be understood as

$\sum_{i,j=1}^{3} \sigma_{ij}\, D_i D_j = \varepsilon_0 \sum_{i=1}^{3} E_i D_i = 2w_{el}$

In the principle coordinate system this index ellipsoid can be defined as

$$\sigma_1 x_1^2 + \sigma_2 x_2^2 + \sigma_3 x_3^2 = \frac{x_1^2}{\varepsilon_1} + \frac{x_2^2}{\varepsilon_2} + \frac{x_3^2}{\varepsilon_3} = 1$$



Please note, very often the radii of the index ellipsoid are associated to the indices in the principle coordinate system. On purpose we have not put this onto the axes of the coordinate system since this is misleading and would contradict the philosophy of the course. The index is a wave optical property used to describe the evolution of a plane wave; it is not a material property as discussed up to that point.

In the degenerate case of a cubic symmetry it remains to mention that this is a sphere; and for a uniaxial crystal this will be a rotational symmetric structure around the z-axis with $n_1 = n_2$.

3.4. Normalmodes in an anisotropic material

As a reminder I would like to stress that a normal mode is a solution to the source free wave equation which experiences upon propagation only a spatial-temporal phase variation, the amplitude and the polarization are preserved and constant. The spatial-temporal phase variation is connected via the dispersion relation, which expresses the functional dependency of the frequency on the wave vector $\omega = \omega(\mathbf{k})$ or the functional dependency of the wave vector on the frequency $\mathbf{k} = \mathbf{k}(\omega)$.

For an isotropic medium the normal modes are monochromatic plane waves

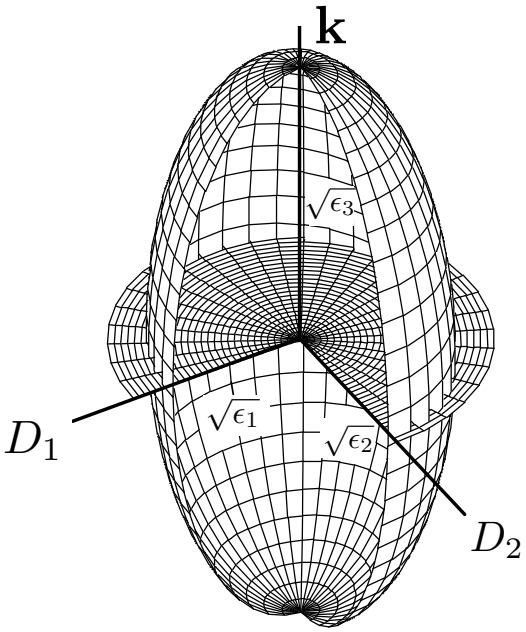
$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_\omega e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}$$

$$\mathbf{k^2}(\omega) = k^2(\omega) = \frac{\omega^2}{c_0^2}\varepsilon(\omega)$$

with $\varepsilon(\omega) > 0$ and $\mathbf{k} \cdot \mathbf{E}_\omega = \mathbf{k} \cdot \mathbf{D}_\omega = 0$. The wave is usually elliptically polarized and the state of polarization is preserved upon propagation. Now we are looking for the normal modes in an anisotropic medium. The solution while propagation in the principal coordinate system is quite easy and its consideration is insightful, afterwards we want to generalize it to an arbitrary propagation direction.

3.4.1. Normal modes for a propagation in the principal coordinate system

The assumption is that the principal axes are in the x, y, z directions. We furthermore require that $\mathbf{k} \cdot \mathbf{E}_\omega = \mathbf{k} \cdot \mathbf{D}_\omega = 0$. We assume for the moment that propagation is in z direction ($\mathbf{k} \rightarrow k_z$). This suggests that $D_x, D_y \neq 0$. The field can be arbitrary in the $x - y -$ plane. Finally, the electric displacement is linked to the electric field via $D_i = \varepsilon_0 \varepsilon_i E_i$ (please note, no summation here).



Only in this particular case we have the unique situation that the two possible polarization directions are decoupled.

$$D_1, \varepsilon_1 \rightarrow D_1 e^{i(\mathbf{k}_1 \cdot \mathbf{r} - \omega t)} = D_1 e^{i\varphi_1} e^{-i\omega t} \quad \text{with} \quad \mathbf{k}_1^2 = \frac{\omega^2}{c_0^2}\varepsilon_1(\omega)$$

$$D_2, \varepsilon_2 \rightarrow D_2 e^{i(\mathbf{k}_2 \cdot \mathbf{r} - \omega t)} = D_2 e^{i\varphi_2} e^{-i\omega t} \quad \text{with} \quad \mathbf{k}_2^2 = \frac{\omega^2}{c_0^2}\varepsilon_2(\omega)$$

From that it follows that $\mathbf{D} \nparallel \mathbf{E}$. This requires to conclude that the elliptical polarization is not a normal mode sustained by the material, since the state of polarization would change upon propagation. However, this doesn't hold if the electric field is polarized along one of the principal axes. Therefore, these plane waves linearly polarized along one of the principal axes are the normal modes of the system.

$$\mathbf{D}^{(a)} = \left\{D_1 e^{i(\mathbf{k}_a \cdot \mathbf{r} - \omega t)}\right\}\mathbf{e}_1 \quad \rightarrow \quad \mathbf{k}_a^2 = \frac{\omega^2}{c_0^2} n_a^2 \quad \rightarrow \quad \text{normal mode a}$$

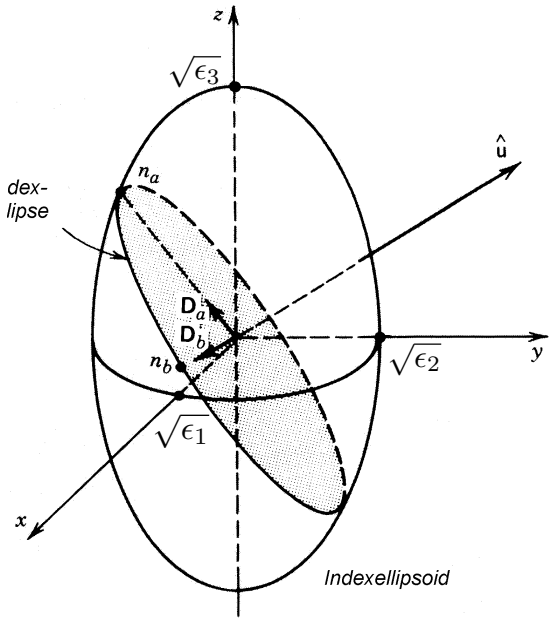
$$\mathbf{D}^{(b)} = \left\{D_2 e^{i(\mathbf{k}_b \cdot \mathbf{r} - \omega t)}\right\}\mathbf{e}_2 \quad \rightarrow \quad \mathbf{k}_b^2 = \frac{\omega^2}{c_0^2} n_b^2 \quad \rightarrow \quad \text{normal mode b}$$

There exist two perpendicular linearly polarized eigenmodes for which $\mathbf{D} \parallel \mathbf{E}$ holds. This is a very simplified and fortunate situation. The question is rather how do the normal modes look like for an arbitrary propagation direction?

3.4.2. Normal modes for an arbitrary propagation direction

3.4.2.1 Geometrical construction I

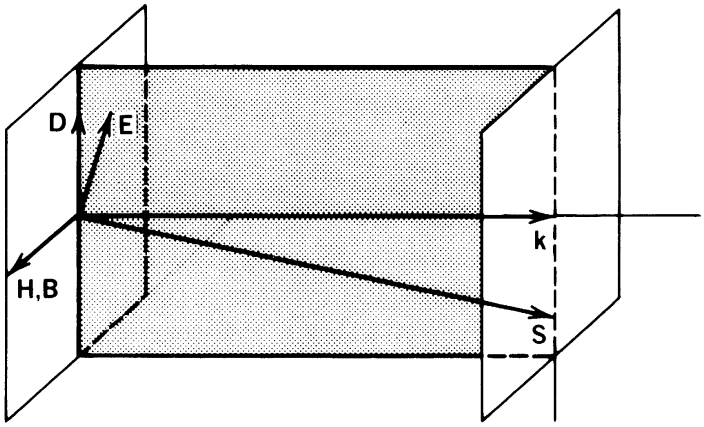
For a given frequency and a given crystal the permittivity tensor is known and with that the index ellipsoid. We furthermore define a direction of the propagation direction we are interested in $\rightarrow \mathbf{k}/k$. Now, we need to sketch the plane that is perpendicular to $\mathbf{k}$ and which goes through the origin of the index ellipsoid.



Now the cross-section is an ellipse, which is just the index ellipse. The principal axes correspond to the refractive indices n_a and n_b of the normal modes in this propagation direction. They obey the same equation as above, $\mathbf{k}_{a,b}^2 = \frac{\omega^2}{c_0^2} n_{a,b}^2$. The directions of the two principal axes correspond to the polarization direction of $\mathbf{D}^{(a)}$ and $\mathbf{D}^{(b)}$. $\mathbf{E}^{(a)}$ and $\mathbf{E}^{(b)}$ follow from that according to $E_i^{(a,b)} = \frac{D_i^{(a,b)}}{\varepsilon_0 \varepsilon_i}$. With that it follows that $\mathbf{D}^{(a,b)} \nparallel \mathbf{E}^{(a,b)}$ and, moreover, $\mathbf{E}^{(a,b)}$ is not perpendicular to $\mathbf{k}$.

This has a direct consequence on the Poynting vector, which is no longer parallel to the wave vector!

$$\langle \mathbf{S} \rangle = \frac{1}{2} \Re(\mathbf{E} \times \mathbf{H}^*) \quad \rightarrow \quad \mathbf{k} \nparallel \langle \mathbf{S} \rangle \text{ because } \mathbf{E} \perp \langle \mathbf{S} \rangle$$



In the case that the index ellipse is a circle, the direction of the associated wave vector defines the optical axis of the crystal.

3.4.2.2 Derivation of the dispersion relation

In the isotropic case I just want to remind you that the length of the wave vector was independent on its direction. The dispersion relation was

$$\mathbf{k}^2(\omega) = k^2(\omega) = \frac{\omega^2}{c_0^2} \varepsilon(\omega)$$

and the normal modes where elliptically polarized plane waves according to

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_\omega e^{i(\mathbf{k}(\omega)\cdot\mathbf{r} - \omega t)}$$

$$\mathbf{D}(\mathbf{r}, t) = \mathbf{D}_\omega e^{i(\mathbf{k}(\omega)\cdot\mathbf{r} - \omega t)}$$

In the anisotropic case the normal mode is again a plane monochromatic wave $\sim e^{i(\mathbf{k}(\omega)\cdot\mathbf{r} - \omega t)}$ but the wave number now explicitly depends on the propagation direction, so $k(\omega, \text{direction})$. Furthermore, we make the educated guess that the

polarization of the normal modes is not elliptically anymore. For a proper notation of the directional dependency we introduce a component notation for the wave vector:

$$\mathbf{k} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = k \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \text{ with } u_1^2 + u_2^2 + u_3^2 = 1$$

The goal of the following derivation, just as before, is the identification of the functional dependency of $\omega = \omega(k_1, k_2, k_3)$ or $\omega = \omega(k, u_1, u_2, u_3)$ or the other way around $k = k(\omega, u_1, u_2, u_3)$. Please note, eventually the latter expressions are slightly redundant since for two given component of the vector that defines the direction, the third is fixed by the constraint that the sum of the square of the components is unity. There is only one equation with four unknowns that have to be solved for.

Starting from Maxwell's equations in spatial Fourier domain with the ansatz indicated above we obtain

$$\mathbf{k} \cdot \mathbf{D} = 0 \qquad \mathbf{k} \times \mathbf{E} = \omega \mu_0 \mathbf{H}$$
$$\mathbf{k} \cdot \mathbf{H} = 0 \qquad \mathbf{k} \times \mathbf{H} = -\omega \mathbf{D}$$

Here, we have dropped for simplicity the ω in the subscript. Now we follow the usual derivation for the wave equation

$$-[\mathbf{k} \times (\mathbf{k} \times \mathbf{E})] = \frac{\omega^2}{c_0^2} \frac{1}{\varepsilon_0} \mathbf{D}$$

$$-\mathbf{k}(\mathbf{k} \cdot \mathbf{E}) + \mathbf{k^2 E} = \frac{\omega^2}{c_0^2} \frac{1}{\varepsilon_0} \mathbf{D}$$

In the principle coordinate system where we have

$$D_i = \varepsilon_0 \varepsilon_i E_i$$

it follows in component notation that

$$-k_i \sum_{j=1}^{3} k_j E_j + k^2 E_i = \frac{\omega^2}{c_0^2} \varepsilon_i E_i$$

$$\left(\frac{\omega^2}{c_0^2} \varepsilon_i - k^2 \right) E_i = -k_i \sum_{j=1}^{3} k_j E_j$$

Please note in the special case of an isotropic material the right hand side is equal to zero and we restore the governing equation for the eigenmodes in the isotropic space. This equation above is an eigenvalue equation and to solve it we have to solve the following characteristic equation

$$\begin{bmatrix} \frac{\omega^2}{c_0^2}\varepsilon_1 - k_2^2 - k_3^2 & k_1k_2 & k_1k_3 \\ k_2k_1 & \frac{\omega^2}{c_0^2}\varepsilon_2 - k_1^2 - k_3^2 & k_2k_3 \\ k_3k_1 & k_3k_2 & \frac{\omega^2}{c_0^2}\varepsilon_3 - k_1^2 - k_2^2 \end{bmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The only meaningful solution we have if the determinant of this linear system is zero. Therefore, from the analysis of det[…]=0 we can obtain the dispersion relation of $\omega = \omega(k)$ for a given ratio of k_i/k.

However, there is a simpler way to obtain this where we start from

$$\left(\frac{\omega^2}{c_0^2}\varepsilon_i - \mathbf{k^2}\right)E_i = -k_i\sum_{j=1}^{3} k_jE_j$$

$$E_i = -\frac{k_i}{\left(\frac{\omega^2}{c_0^2}\varepsilon_i - \mathbf{k^2}\right)}\sum_{j=1}^{3} k_jE_j$$

Now, multiplication with k_i and summation over i, and eventually a substitution between i and j on the left hand side and we obtain

$$\sum_{j=1}^{3} k_jE_j = -\sum_{i=1}^{3}\frac{k_i^2}{\left(\frac{\omega^2}{c_0^2}\varepsilon_i - \mathbf{k^2}\right)}\sum_{j=1}^{3} k_jE_j$$

By exploiting the fact that the divergence of the electric field is not vanishing, i.e. $\mathbf{div\ E} = \sum_{j=1}^{3} k_jE_j \neq 0$ it follows that

$$\sum_{i=1}^{3}\frac{k_i^2}{\left(\mathbf{k^2} - \frac{\omega^2}{c_0^2}\varepsilon_i\right)} = 1$$

is the preliminary dispersion relation. In combination with

$$\begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = k(\omega)\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \frac{\omega}{c_0}n(\omega)\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

$$\sum_{i=1}^{3}\frac{k_i^2}{\left(\mathbf{k^2} - \frac{\omega^2}{c_0^2}\varepsilon_i(\omega)\right)} = 1 \rightarrow \sum_{i=1}^{3}\frac{u_i^2}{\left(1 - \frac{\varepsilon_i}{n^2(\omega)}\right)} = 1$$

$$\sum_{i=1}^{3}\frac{u_i^2}{(n^2(\omega) - \varepsilon_i(\omega))} = \frac{1}{n^2(\omega)}$$

This is the final form of the dispersion relation. Now the actual action is such that for a given crystal characterized by its permittivity tensor and an indication on the directions in terms of u_1 and u_2 (u_3 would be redundant since the sum of the square must be unity), we can calculate with the equation above the effective index the normal mode will experience to be $n(\omega, u_1, u_2)$.

However, please take care that as soon as pathological cases emerge, e.g. those for which $\mathbf{div}\,\mathbf{E} = 0$ in e.g. cubical crystals, it is required to return to the explicit expression and work with it

$$u_1^2(n^2-\varepsilon_2)(n^2-\varepsilon_3)n^2 + u_2^2(n^2-\varepsilon_1)(n^2-\varepsilon_3)n^2 + u_3^2(n^2-\varepsilon_1)(n^2-\varepsilon_2)n^2 = (n^2-\varepsilon_1)(n^2-\varepsilon_2)(n^2-\varepsilon_3)$$

This is after all a quadratic equation in n^2 since the terms with n^6 cancel out. This equation has two solutions n_a and n_b and with this also , $k_{a,b} = \frac{\omega}{c_0}\, n_{a,b}$.

The normal modes have a polarization in the electric displacement for which they are perpendicular on each other.

In the special case of a propagation direction along one of the principal axes, as previously discussed, for which $u_3 = 1$ it follows that

$$(n^2-\varepsilon_1)(n^2-\varepsilon_2)n^2 = (n^2-\varepsilon_1)(n^2-\varepsilon_2)(n^2-\varepsilon_3)$$

Multiplying this at first out and removing all the terms that depend on the six power of the index and recollecting the terms leads to

$$(n^2-\varepsilon_1)(n^2-\varepsilon_2)\varepsilon_3 = 0$$

which has as solutions

$$n_a^2 = \varepsilon_1 \quad \text{and} \quad n_b^2 = \varepsilon_2$$

This corresponds just to the solutions presented above while considering the propagation along one of the principal axes.

Eventually we can also compute in a very last step the fields of the normal modes. From above we know that

$$\left(\frac{\omega^2}{c_0^2}\varepsilon_i - k^2\right)E_i = -k_i \sum_{j=1}^{3} k_j E_j$$

$$E_i = -\frac{k_i}{\left(\frac{\omega^2}{c_0^2}\varepsilon_i - k^2\right)}\sum_{j=1}^{3} k_j E_j$$

where the sum is independent on the index i. Therefore it follows that $\sum_{j=1}^{3} k_j E_j = const$ and we can write down the ratio of the amplitudes as

$$E_1 : E_2 : E_3 = \frac{k_1}{\left(\frac{\omega^2}{c_0^2}\varepsilon_1 - k^2\right)} : \frac{k_2}{\left(\frac{\omega^2}{c_0^2}\varepsilon_2 - k^2\right)} : \frac{k_3}{\left(\frac{\omega^2}{c_0^2}\varepsilon_3 - k^2\right)}$$

In combination with $D_i = \varepsilon_0 \varepsilon_i E_i$ we obtain

$$D_1 : D_2 : D_3 = \frac{\varepsilon_1 k_1}{\left(\frac{\omega^2}{c_0^2}\varepsilon_1 - k^2\right)} : \frac{\varepsilon_2 k_2}{\left(\frac{\omega^2}{c_0^2}\varepsilon_2 - k^2\right)} : \frac{\varepsilon_3 k_3}{\left(\frac{\omega^2}{c_0^2}\varepsilon_3 - k^2\right)}$$

As we can see that the field components are all real values, there is no phase difference between the different components, which necessarily leads to a linear polarization for the normal modes.

The last thing we can prove is the orthogonality of the modes $\mathbf{D}^{(a)}$ and $\mathbf{D}^{(b)}$ using the scalar product between the electric displacement field. Please note, the electric field won't be perpendicular on each other.

$$\mathbf{D}^{(a)} \cdot \mathbf{D}^{(b)} \sim \sum_{i=1}^{3} \frac{\varepsilon_i^2 k_a k_b u_i^2}{\left(k_a^2 - \frac{\omega^2}{c_0^2}\varepsilon_i\right)\left(k_b^2 - \frac{\omega^2}{c_0^2}\varepsilon_i\right)}$$

$$= \frac{c_0^2}{\omega^2}\frac{k_a k_b}{(k_b^2 - k_a^2)}\left[k_a^2 \sum_{i=1}^{3} \frac{\varepsilon_i u_i^2}{\left(k_a^2 - \frac{\omega^2}{c_0^2}\varepsilon_i\right)} - k_b^2 \sum_{i=1}^{3} \frac{\varepsilon_i u_i^2}{\left(k_b^2 - \frac{\omega^2}{c_0^2}\varepsilon_i\right)}\right]$$

Now the terms in the sum are vanishing as can be seen while writing the dispersion relation as

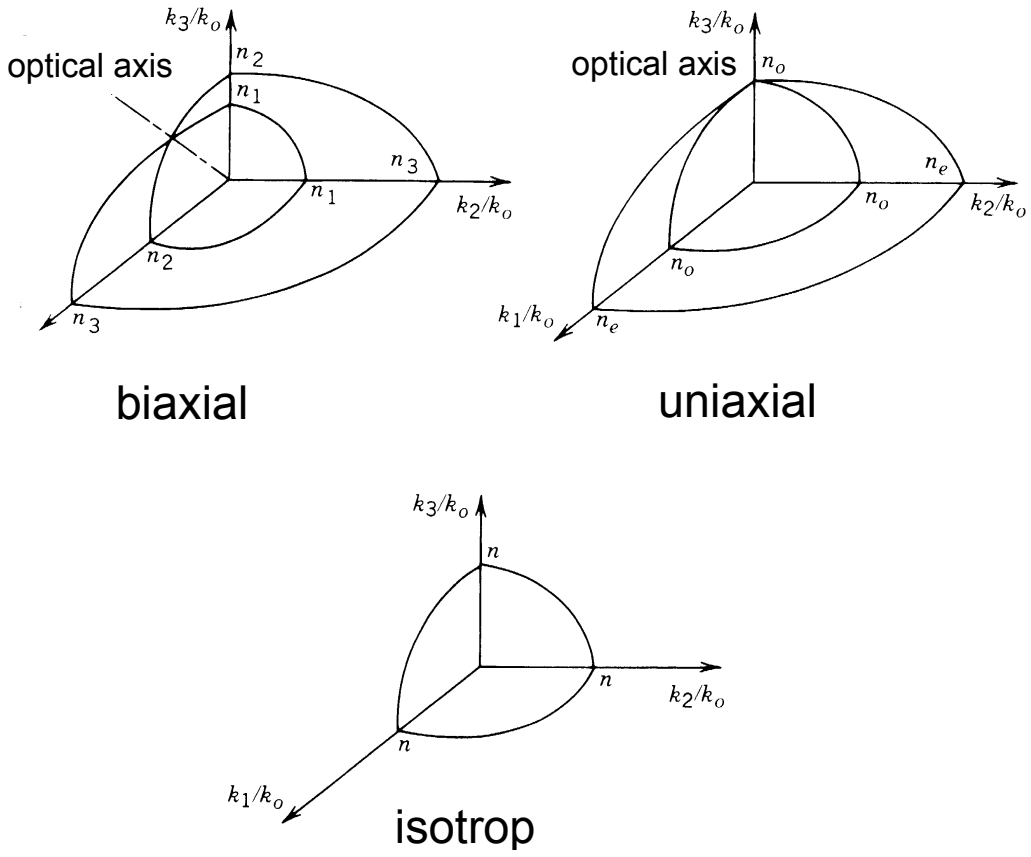
$$1 = \sum_{i=1}^{3} \frac{k_{a,b}^2 u_i^2}{\left(k_{a,b}^2 - \frac{\omega^2}{c_0^2}\varepsilon_i(\omega)\right)} = \sum_{i=1}^{3} \frac{\left(k_{a,b}^2 - \frac{\omega^2}{c_0^2}\varepsilon_i(\omega) + \frac{\omega^2}{c_0^2}\varepsilon_i(\omega)\right)u_i^2}{\left(k_{a,b}^2 - \frac{\omega^2}{c_0^2}\varepsilon_i(\omega)\right)}$$

$$= 1 + \frac{\omega^2}{c_0^2}\sum_{i=1}^{3} \frac{\varepsilon_i u_i^2}{\left(k_{a,b}^2 - \frac{\omega^2}{c_0^2}\varepsilon_i\right)}$$

Therefore, the eigenmodes are orthogonal.

3.4.2.3 Geometrical interpretation II: normal surfaces

In addition to the considerations above we can also plot the index of the two modes as surfaces in the space spanned by the components of the wave vector k_i. This is providing us a centro-symmetric two-layer surface. Please note, normal surfaces are often called iso-frequency surfaces or just iso-surfaces. Eventually they display the functional dependency of the admissible directions (or elements of the k-vectors) for a given frequency.



The cross sections with the principal axes are either circles or ellipses.

Biaxial: the two surfaces intersect in four different points. The connecting lines between the two points are the two optical axes. Please note, the optical axis of a crystal is the direction in which the wave does not suffer from any birefringence.

Uniaxial: It's a body of revolution made from an ellipse and a sphere. There are two intersection points at the poles. The connecting line equally provides information on the optical axes. If $\varepsilon_1 = \varepsilon_2 = \varepsilon_{\mathrm{or}}$ and $\varepsilon_3 = \varepsilon_{\mathrm{e}}$.

The subscripts or and e stands for ordinary and extraordinary optical axes.

cubic: The structure is isotropic and the two interfaces are identical

How the figures need to be understood?

At first you need to fix the directions u_1 and u_2. Then you have to identify the intersections with the surfaces. The distance between the center of the coordinate system and the intersections provides you the refractive indices of the normal modes. Only in the special case that you are in the optical axes, the two possible indices are identical $n_a = n_b$.

Taking all together you may have access now to two different geometrical interpretations:

(a) Index ellipsoid

-Direction fixed $\rightarrow$ identifying index ellipse $\rightarrow$ semi-axes provide n_a and n_b, being the indices, which are experienced by the normal modes

(b) Normal surfaces

-Direction fixed $\rightarrow$ cross section to the normal surfaces $\rightarrow$ distance to the center provide n_a and n_b, optical axis is the connecting line between the center and the cross section of the two branches

In a nutshell, in an anisotropic material there exist two normal modes. These normal modes are linearly polarized plane monochromatic waves. They have two different phase velocities, given by $\frac{c_0}{n_{a,b}}$ and two perpendicular polarization direction. All you need to provide is information on the material and the direction of propagation.

3.4.3 Uniaxial crystals

Uniaxial crystals have as a crystal structure to be either trigonal, tetragonal, or hexagonal. Their index ellipsoid is rotational symmetric around the z-axes and they are characterized by an ordinary and an extraordinary index that are derived from

$$\varepsilon_1 = \varepsilon_2 = \varepsilon_{\mathrm{or}} \quad \text{and} \quad \varepsilon_3 = \varepsilon_{\mathrm{e}}$$

The normal modes, in general, do not see the associated indices directly.

There are two different normal modes

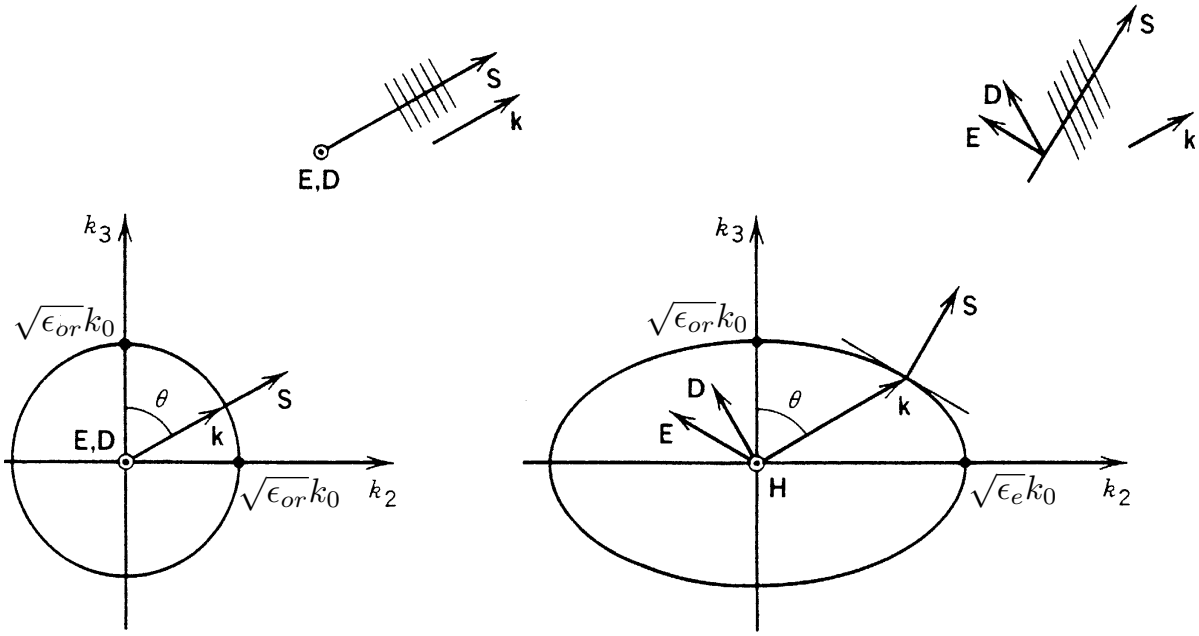
a) Ordinary waves → index is independent on the direction

→ $D^{(\mathrm{or})}$ polarized perpendicular to the z-axis and to the wave vector

b) Extraordinary waves → index is dependent on the direction

→ polarized perpendicular to the wave vector and within the plane spanned by the wave vector and the rotational axis, $D^{(e)}$ is perpendicular to $D^{(or)}$

According to our definitions the z-axes is the optical axis with $n_a = n_b$.



(a) Ordinary *(b)* Extraordinary

Now how does the dispersion relation looks like for such material? The best starting point will be the explicit solution, which we had before.

$$u_1^2(n^2-\varepsilon_{\mathrm{or}})(n^2-\varepsilon_{\mathrm{e}})n^2+u_2^2(n^2-\varepsilon_{\mathrm{or}})(n^2-\varepsilon_{\mathrm{e}})n^2+u_3^2(n^2-\varepsilon_{\mathrm{or}})^2n^2$$
$$=(n^2-\varepsilon_{\mathrm{or}})^2(n^2-\varepsilon_{\mathrm{e}})$$

$$(n^2-\varepsilon_{\mathrm{or}})\left[(u_1^2+u_2^2)(n^2-\varepsilon_{\mathrm{e}})+u_3^2(n^2-\varepsilon_{\mathrm{or}})-\frac{(n^2-\varepsilon_{\mathrm{e}})(n^2-\varepsilon_{\mathrm{or}})}{n^2}\right]=0$$

Obviously there are two solutions.

1. Ordinary wave

$$n_a^2=\varepsilon_{\mathrm{or}} \quad \rightarrow \quad k_a^2=\frac{\omega^2}{c_0^2}\varepsilon_{\mathrm{or}}=\frac{\omega^2}{c_0^2}\,n_a^2$$

2. Extraordinary wave

$$(u_1^2 + u_2^2)(n_b^2 - \varepsilon_\mathrm{e}) + u_3^2(n_b^2 - \varepsilon_\mathrm{or}) = \frac{(n_b^2 - \varepsilon_\mathrm{e})(n_b^2 - \varepsilon_\mathrm{or})}{n^2}$$

$$n_b^2(u_1^2 + u_2^2 + u_3^2) - n_b^2 - (u_1^2 + u_2^2)\varepsilon_\mathrm{e} - u_3^2\varepsilon_\mathrm{or} + \varepsilon_\mathrm{e} + \varepsilon_\mathrm{or} - \frac{\varepsilon_\mathrm{e}\varepsilon_\mathrm{or}}{n_b^2} = 0$$

$$\varepsilon_\mathrm{e}(1 - u_1^2 - u_2^2) + \varepsilon_\mathrm{or}(1 - u_3^2) - \frac{\varepsilon_\mathrm{e}\varepsilon_\mathrm{or}}{n_b^2} = 0$$

$$\frac{(u_1^2 + u_2^2)}{\varepsilon_\mathrm{e}(\omega)} + \frac{u_3^2}{\varepsilon_\mathrm{or}(\omega)} - \frac{1}{n_b^2} = 0$$

Now on the base of that equation we can calculate n_a and n_b for a given illumination direction.

Geometrical representation of the normal surfaces is straightforward

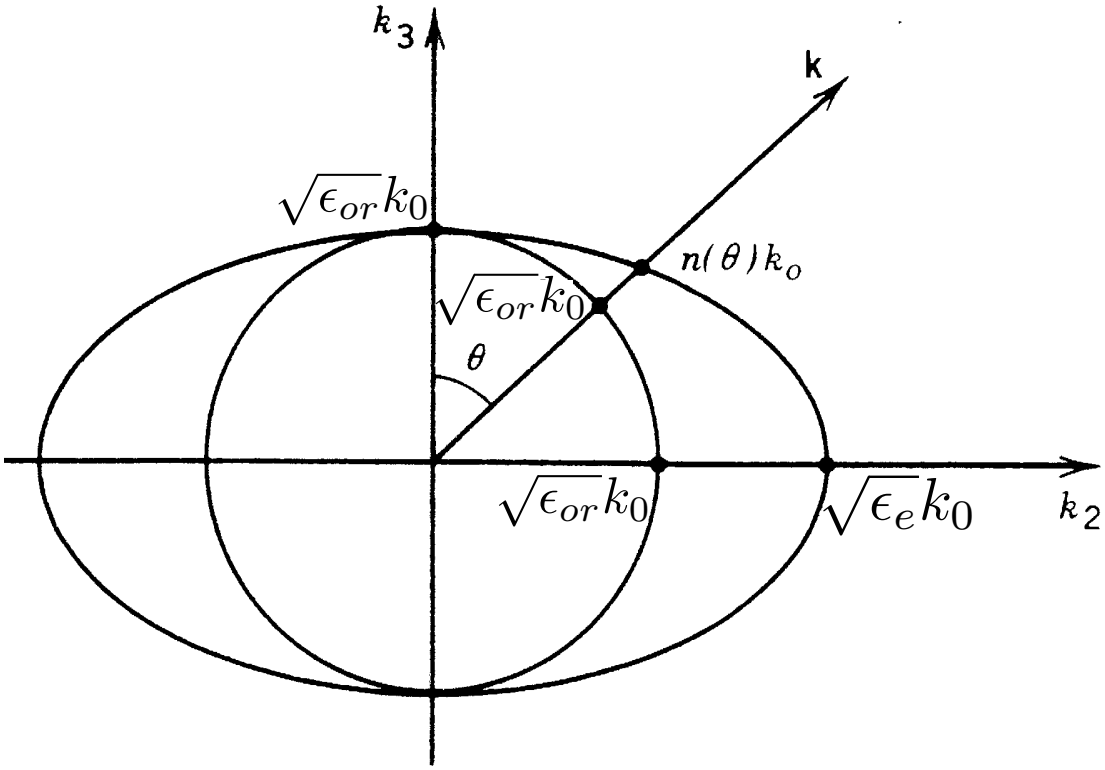
1. Ordinary wave

$$k_a^2 = k_1^2 + k_2^2 + k_3^2 = k_0^2\varepsilon_\mathrm{or}$$

2. Extraordinary wave

$$\frac{1}{\varepsilon_\mathrm{e}}\frac{(k_1^2 + k_2^2)}{k_0^2} + \frac{1}{\varepsilon_\mathrm{or}}\frac{k_3^2}{k_0^2} = 1$$

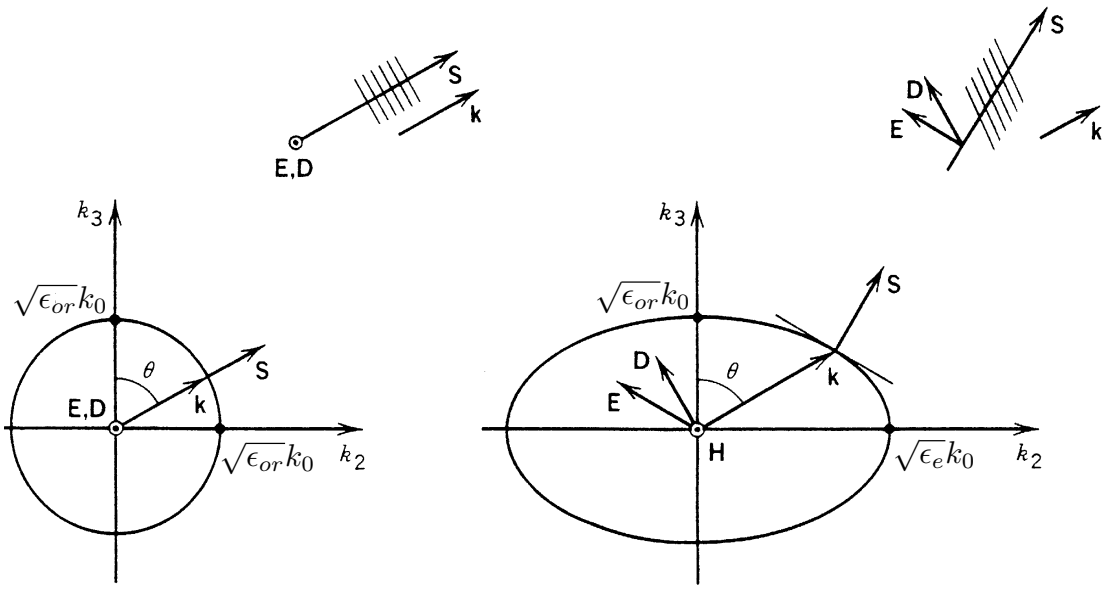
This is the equation for a rotational symmetric ellipsoid. For a simplification of the treatment but by no means as a restriction, we would like to assume in the following that the propagation takes place in the $y - z$-plane, suggesting that $u_1 = 0$. Please note that the size of the ellipsoid changes with frequency. The isosurfaces then look like



What eventually can be said about the fields? From the general discussion as outlined above we know that

$$D_1 : D_2 : D_3 = \frac{\varepsilon_{\mathrm{or}} k_1}{\left(\frac{\omega^2}{c_0^2}\varepsilon_{\mathrm{or}} - k^2\right)} : \frac{\varepsilon_{\mathrm{or}} k_2}{\left(\frac{\omega^2}{c_0^2}\varepsilon_{\mathrm{or}} - k^2\right)} : \frac{\varepsilon_{\mathrm{e}} k_3}{\left(\frac{\omega^2}{c_0^2}\varepsilon_{\mathrm{e}} - k^2\right)}$$

For the extraordinary wave the denominators are finite and since we have $k_1 = 0$, we also have $D_1^{(e)} = 0$. Therefore, the field is polarized in the y-z-plane. Since moreover the ordinary wave needs to be polarized perpendicular to the extraordinary wave, the ordinary wave must be polarized in the x-direction. Moreover, independent from that it is also known that the ordinary field is polarized perpendicular to the plane that is given by the optical axis and the wave vector. In the example shown below the optical axis is the z-direction. Hence, the ordinary wave must be polarized in the x-direction.



(a) Ordinary *(b)* Extraordinary

In this case we can also perform a simple calculation of the effective index $n^2(\Theta)$ of the extraordinary wave depending on the angle of propagation in the y-z-plane. In general it holds that

$$\frac{(u_1^2 + u_2^2)}{\varepsilon_{\mathrm{e}}(\omega)} + \frac{u_3^2}{\varepsilon_{\mathrm{or}}(\omega)} - \frac{1}{n_b^2(\omega)} = 0$$

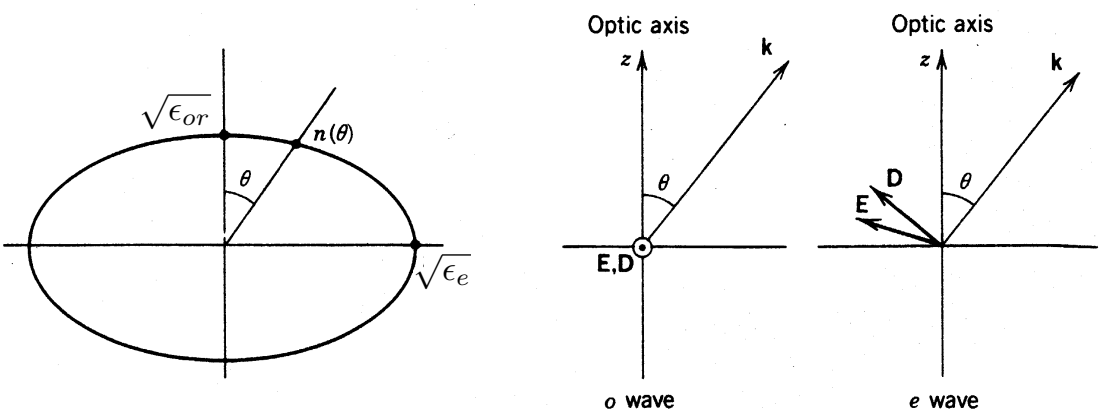
For the special situation mentioned above that $u_1 = 0$ it simplifies to

$$\frac{u_2^2}{\varepsilon_{\mathrm{e}}(\omega)} + \frac{u_3^2}{\varepsilon_{\mathrm{or}}(\omega)} - \frac{1}{n_b^2(\omega)} = 0$$

$$u_2 = \sin\Theta \quad \text{and} \quad u_3 = \cos\Theta$$

$$\frac{\sin^2\Theta}{\varepsilon_{\mathrm{e}}(\omega)} + \frac{\cos^2\Theta}{\varepsilon_{\mathrm{or}}(\omega)} - \frac{1}{n_b^2(\omega,\Theta)} = 0$$

$$n_b^2(\omega,\Theta) = \frac{\varepsilon_{\mathrm{e}}(\omega)\varepsilon_{\mathrm{or}}(\omega)}{\varepsilon_{\mathrm{or}}(\omega)\sin^2\Theta + \varepsilon_{\mathrm{e}}(\omega)\cos^2\Theta}$$



Finally we would like to mention that a classification is usually used for uniaxial optical materials. For

$$\varepsilon_{\mathrm{or}}(\omega) > \varepsilon_{\mathrm{e}}(\omega) \quad \rightarrow \quad \text{negatively uniaxal}$$
$$\varepsilon_{\mathrm{or}}(\omega) < \varepsilon_{\mathrm{e}}(\omega) \quad \rightarrow \quad \text{positively uniaxal}$$

3.5 Derived quantities from the dispersion relation

In a final note we would like to outline a few important parameters that can be derived from the dispersion relation and which describe the evolution of Gaussian beams in materials. It is specified here for a uniaxial material but the general concepts are applicable to any other material as well. This makes it such powerful. The concept can be also applied to metamaterials, photonic crystals, coupled waveguides etc.; all systems for which a dispersion relation needs to be known.

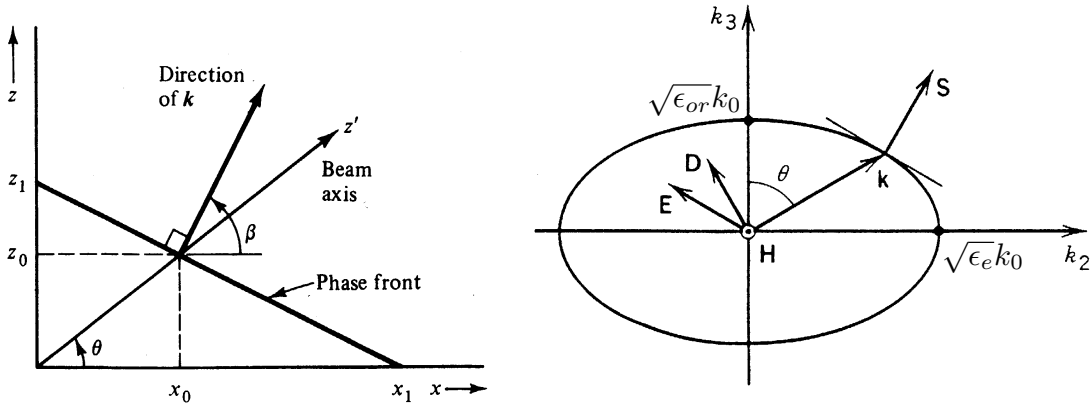
The first important quantity is the inclination coefficient. This eventually is given by the tangent of the iso-frequency surface; being the normal surface for a particular wave

$$V(k_2^0) = \left.\frac{\partial k_3(\omega, k_2)}{\partial k_2}\right|_{k_2^0}$$

Here we assume that the principal propagation direction is along the z-axis and the propagation takes place in the y-z-plane. The inclination coefficient is related to the transverse shift a Gaussian beam experiences during propagation. Since there is a difference between the direction the energy transport takes place (so basically direction of the time averaged Poynting vector) and the normal direction of the fronts of constant phase, a Gaussian beam will not propagate in the direction of the

wave vector. Instead, it will propagate to the direction normal to the negative tangent.

$$-\left.\frac{\partial k_2}{\partial k_3(\omega,k_2)}\right|_{k_2^0}=\frac{k_3}{k_2^0}\frac{\varepsilon_{\mathrm{e}}(\omega)}{\varepsilon_{\mathrm{or}}(\omega)}=\tan\beta\,\frac{\varepsilon_{\mathrm{e}}(\omega)}{\varepsilon_{\mathrm{or}}(\omega)}=\tan\Theta$$

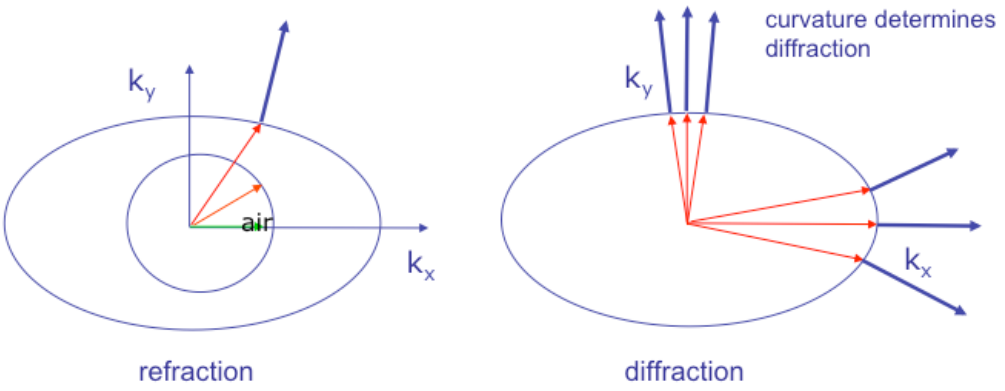


The direction of a Gaussian beam is perpendicular to the surface and agrees with the direction of the Poynting vector. For an extraordinary wave this no longer coincides with the direction of the wave vector. This also suggests that the direction of the propagation of ordinary and extraordinary wave no longer agrees; a phenomena that is called as spatial walk off. This walk off needs to be carefully considered in the applications of birefringent effects, e.g., in the context of nonlinear optics.

From the inclination coefficient the refraction coefficient can be derived that expresses the way refraction takes place at the interface between two media characterized by a given inclination coefficient

$$R(k_2^0)=\frac{V_1(k_2^0)}{V_2(k_2^0)}$$

The sign of *R* distinguishes between normal (*R* > 0) and anomalous (negative) refraction (*R*<0).



A further important parameter is the diffraction coefficient. It expresses how strong the diffractive spread of a Gaussian beam will be while propagating through a material.

$$D(k_2^0) = \left. \frac{\frac{\partial^2 k_3(\omega, k_2)}{\partial k_2^2}}{\left\{1 + \left(\frac{\partial k_3(\omega, k_2)}{\partial k_2}\Big|_{k_2^0}\right)^2\right\}^{3/2}} \right|_{k_2^0}$$

The diffraction coefficient evaluated at k_2^0 is a measure for the spreading or focusing of the beam along its mean propagation direction. If $D < 0$ diffraction is normal, similar to an ordinary homogeneous medium, and if $D > 0$ it is anomalous, observable, for instance, in waveguide arrays and photonic crystals.

Please note, there is a strong analogy to the propagation of pulses in temporal domain as earlier discussed. There, the dispersion relation was expanded in temporal frequency space to evaluate the speed of propagation of the wave package (first derivative) and the dispersive spread (second derivative). This is full analogy to the parameters discussed here; just by replacing the temporal frequency with the spatial frequency.

The most important thing to remember is that whenever the dispersion relation is known, the light propagation in the medium can be theoretically understood.

4. Coherence optics

4.1. Basics

The theory of optical coherence is a part of the field of statistical optics. These fields of science have been pioneered by contributions from Max Born and Emil Wolf in the fifties of the last century. It is since then of major importance and has been further developed. Current research interest, e.g. deal with the proper description partially coherent light in opto-electronic devices such as a solar cell.

Subject of concern: Properties of "*random*" light, light that has been experiencing or was exposed to fluctuations of the properties of the source or to fluctuations of the optical properties of the medium through which the light has been propagating.

Referential examples:

- Natural light as emitted from thermal sources such as an incandescent lamp or the sun, sources where many independent atoms contribute to the emission. They all emit light at slight different frequencies and phases. Light emitted from such sources cannot be described by an infinite wave train that is characterized by a singular frequency.
- Light that is reflected at randomly rough surfaces. Here, the frequency is preserved but the phase of an incident planar wave suffers from fluctuations.

Properties of these light sources are discussed in the context of coherence theory.

Assumptions: scalar approximation $u(\mathbf{r}, t)$

Thus far we have only considered fields that an be written as:

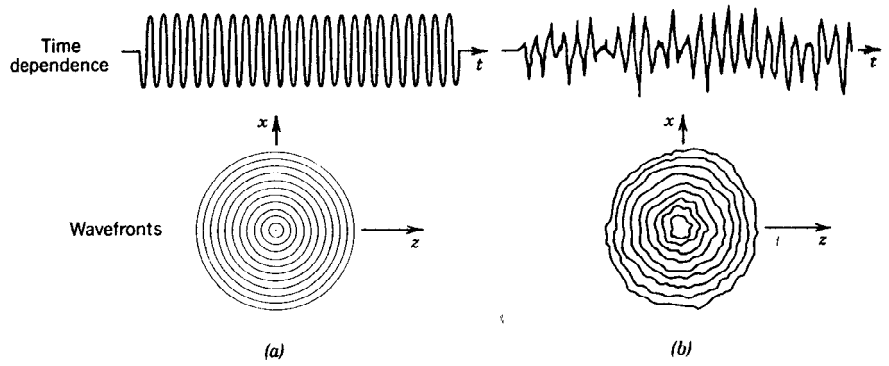
$$u(\mathbf{r}, t) = U(\mathbf{r}, \omega)e^{-i\omega t} + c.c.$$

with, for example,

$$U(\mathbf{r}, \omega) = e^{i\mathbf{k}(\omega)\cdot\mathbf{r}}$$

Such light field is entirely determined in all its properties; it is what is called coherent light. Please note, such ideal time-harmonic field is required to be existent forever; it cannot therefore exist in reality. Only if the frequencies are confined to a narrow interval the light can be considered as quasi-monochromatic. How narrow it needs to be has to be answered in the context of a specific optical device.

What will be considered now is a field that is a random variable; but it shall still obey the wave equation. This type of light can be visualized either in time or in a spatial domain (please note in time or space we plot here only the real part of the complex variables or actually the wave front, the field itself is complex and has a real and imaginary part):



Light in general can be categorized according to their degree of phase correlation:

- Coherent light
- Partially coherent light
- Incoherent light

Please note, the problem we will tackle is independent from the problem of treating a polychromatic source. There, due to superposition principle, the response of an optical system to each frequency needs to be calculated and afterwards summed up. Unfortunately, in most cases we do not know the precise frequencies, the amplitudes and the phases of the harmonic components of the field emitted by a real source. The only information we have available is a statistical model for the distribution of frequencies, amplitudes and phases. These models carry over their statistical features into the field emitted by the source elsewhere in space, e.g. in the image plane of an objective lens of a microscope. Fortunately, the statistical models turn out to be sufficient to predict the outcome of most optical experiments. These statistical models do not need to be complete in a sense that all properties are perfectly reflected, in most cases we will rely here on second-order averages called coherence functions that are entirely adequate.

The theory of optical coherence eventually shall allow us to describe the deviations of partially coherent waves from perfectly coherent waves. In the following we will

always distinguish between time coherence, spatial coherence, and mutual coherence.

4.2. Statistical properties of light

4.2.1 Definitions

We assume in the following that $u(\mathbf{r},t)$ is a complex analytical signal

$$u(\mathbf{r},t)=\int_0^\infty U(\mathbf{r},\omega)e^{-i\omega t}d\omega$$

$$u_R(\mathbf{r},t)=\Re\left[\int_0^\infty U(\mathbf{r},\omega)e^{-i\omega t}d\omega\right]$$

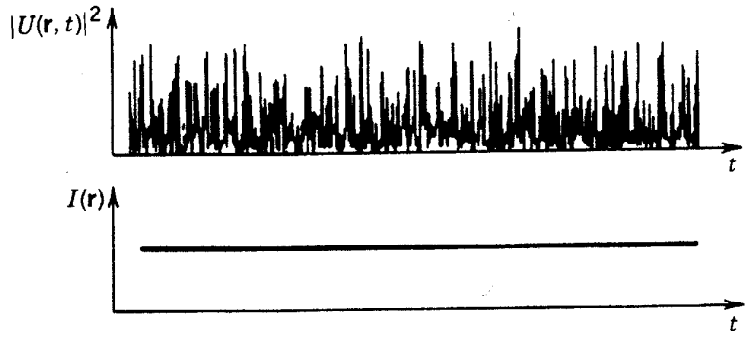
It corresponds to an arbitrary physical component of a non-monochromatic field at position $\mathbf{r}$ and time t.

a) Intensity

$$I(\mathbf{r},t)=\langle|u(\mathbf{r},t)|^2\rangle_e$$

$|u(\mathbf{r},t)|$ is a randomly fluctuating intensity and $\langle\cdots\rangle_e$ describes here the ensemble average. This requires a sufficient large number of measurements on a nominally identically prepared system.

Example I: stationary statistical light source: light valve driven by a constant current



$$I(\mathbf{r},t)=I(\mathbf{r})$$

Now an important statement is given that will be valid for the systems we consider: The ergodic hypothesis that applies says that, broadly speaking, the system has the same behavior averaged over time as averaged over the different individual implementations of the same system. Time and ensemble average are the same.

$$\langle \cdots \rangle_e = \langle \cdots \rangle_t = \langle \cdots \rangle$$

From the ergodic hypothesis it follows that the intensity will be

$$I(\mathbf{r}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |u(\mathbf{r},t)|^2 dt$$

In detail, we can derive the constant intensity as described using the idea that from a source the light is emitted by individual atoms that emit light of a specific frequency and with an arbitrary phase. We can technically write the field as a superposition of N monochromatic waves with equal amplitude, random phases and frequencies randomly chosen within an interval $\Delta\omega$ centered around some central frequency.

$$u(t) = A \sum_{n=1}^{N} e^{i(-\omega_n t + \varphi_n)}$$

The average over a time interval of length $2T$ of the intensity is

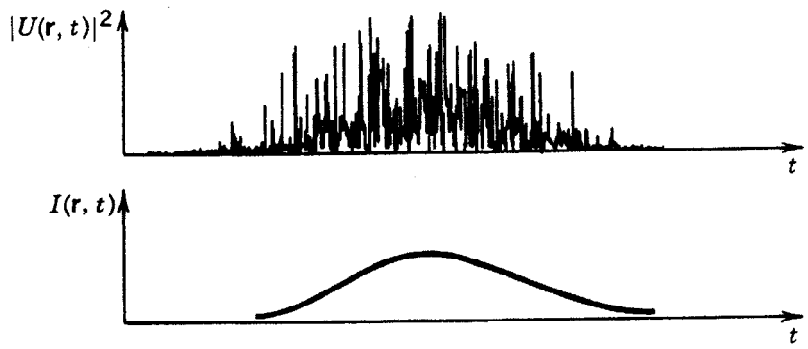
$$I = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |u(t)|^2 dt$$

$$I = \lim_{T\to\infty} \frac{A^2}{2T} \int_{-T}^{T} \left| \sum_{n=1}^{N} e^{i(-\omega_n t + \varphi_n)} \right|^2 dt$$

$$I = \lim_{T\to\infty} \frac{A^2}{2T} \int_{-T}^{T} \sum_{n=1}^{N} \sum_{m=1}^{N} e^{i((\omega_m - \omega_n)t + \varphi_n - \varphi_m)}\, dt$$

If the integration time is long enough when compared to the period of a beat associated with the frequency difference of $\omega_m - \omega_n$, the oscillating parts of the integral tend to cancel each other and the inter terms that persists are those with $\omega_m = \omega_n$. The time average intensity is then constant and proportional to A^2 and to N.

Example II: non-stationary statistical light source: light valve driven by a time dependent current $\rightarrow I(\mathbf{r},t)$



b) Temporal coherence and spectra

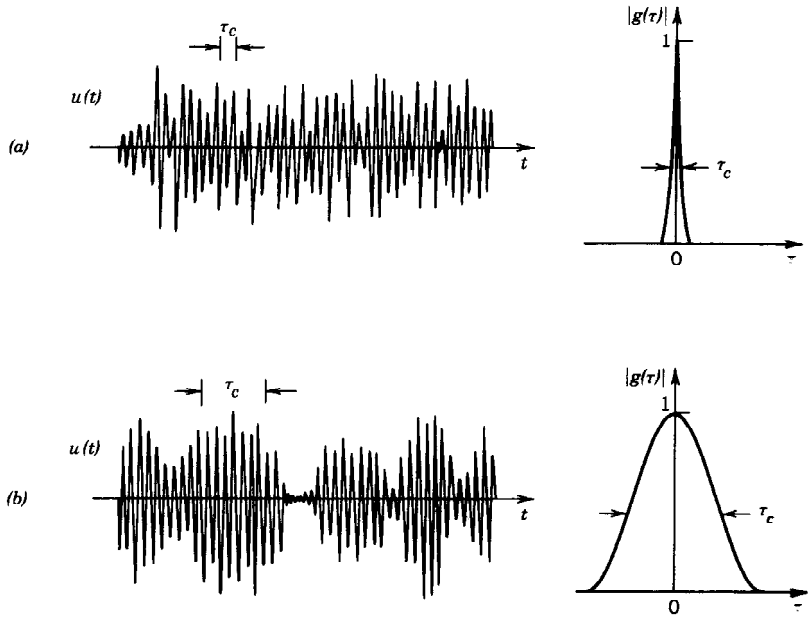
We will assume in the following stationary statistical process $\rightarrow I(\mathbf{r})$ and we only consider the intensity at a fixed point in space $\rightarrow I(\mathbf{r}) = I$.

The measure for the correlation of the field at $u(t)$ and $u(t+\tau)$ is the auto-correlation function $G(\tau)$

$$G(\tau) = \langle u^*(t)u(t+\tau)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} u^*(t)u(t+\tau)dt$$

The function $G(\tau)$ is called the temporal coherence function. Since it is defined with respect to it-self, it is also called a self-coherence function.

The temporal width of the function $G(\tau)$ tells us something how quickly the envelope of the function $u(t)$ changes as a function of time. The function $u(t)$ only changes smoothly within a time τ_C called the coherence time. Beyond this time the function is rough, i.e. $G(\tau) = 0$ for $\tau > \tau_C$. Of course, this is not meant in a strict sense but shall only qualitatively express the functionality.



In particular it holds that

$$G(0) = I$$

Now the function $G(\tau)$ unfortunately contains two different information: information on the coherence but also on the intensity. Both effects should be separated which asks to introduce a normalized quantity.

For that purpose we introduce the complex degree of temporal coherence $g(\tau)$. Please note, we will introduce later higher order correlation functions (related to the intensity). Therefore this degree of coherence is often called a first-order correlation function abbreviated with '(1)' in the superscript. This is omitted here as long as we face no possible confusion. The function is defined as

$$g(\tau) = \frac{G(\tau)}{G(0)} = \frac{\langle u^*(t)u(t+\tau)\rangle}{\langle u^*(t)u(t)\rangle}$$

The function is bound between zero and one, i.e. $0 \leq |g(\tau)| \leq 1$.

Example I: perfect deterministic light, i.e. coherent light

$$u(t) = Ae^{-i\omega_0 t}$$

$$g(\tau) = \frac{\frac{1}{2T}\int_{-T}^{T}|A|^2 e^{-i\omega_0\tau}dt}{\frac{1}{2T}\int_{-T}^{T}|A|^2 dt} = e^{-i\omega_0\tau} \quad \rightarrow \quad |g(\tau)| = 1$$

Example II: general case

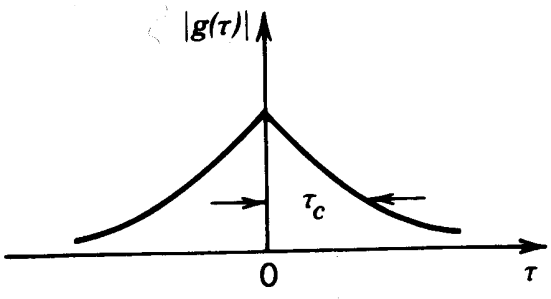
Allows a more rigid definition of the coherence time:

$|g(\tau)|$ is a monotonously decaying function. This allows to define the coherence time as the width at which the function $|g(\tau)|$ has been decaying to $|g(\tau_c)| = {}^1/_e$. Please note, other definitions also consider a decay to one half.

$$\tau_c = \int_{-\infty}^{\infty}|g(\tau)|^2 d\tau$$

$$|g(\tau)| = 1 \quad \curvearrowright \quad \tau_c \rightarrow \infty$$

Example: $g(\tau) = e^{-\frac{|\tau|}{A}} \curvearrowright \tau_c = 2\int_0^{\infty} e^{-2\frac{|\tau|}{A}} d\tau = A(1-0) = A$



The coherence length can be calculated from the coherence time by $l_c = c\tau_c$.

If all the relevant length scales of the optical system are smaller (or ideally much smaller) than the coherence length, we may consider the system as coherent.

c) The theorem of Wiener-Khinchin

The Wiener-Khinchin-theorem is important since it expresses the dependency between the spectrum and the temporal coherence of a source, a wave field.

Theorem: The Fourier transform of the autocorrelation function of a signal is identical with the magnitude square of the Fourier transform of the signal. The latter being the spectrum

Here, the theorem establishes a relation between the temporal coherence function and the spectrum of the signal

$$G(\tau) = \int_{-\infty}^{\infty} S(\omega) e^{-i\omega\tau} d\omega$$

where the signal $S(\omega)$ is defined here as

$$S(\omega) = |U(\omega)|^2$$

Proof: (please consider that in the following we will deal with finite times; since for stationary process an infinite time will lead to infinite energies which are difficult to handle)

$$G(\tau) = \langle u^*(t)u(t+\tau)\rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u^*(t)u(t+\tau)dt$$

Now for a truncated Fourier transform this reads as

$$U_T(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u_T(t) e^{i\omega t} dt$$

$$u_T(t) = \begin{cases} u(t) & |t| \le T \\ 0 & \text{otherwise} \end{cases}$$

$$u_T(t) = \int_{-\infty}^{\infty} U_T(\omega) e^{-i\omega t} d\omega$$

$$G(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_T^*(t) u_T(t+\tau) dt$$

$$G(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_T^*(t) \left[\int_{-\infty}^{\infty} U_T(\mathbf{r},\omega) e^{-i\omega(t+\tau)} d\omega \right] dt$$

$$G(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{-\infty}^{\infty} \left[u_T^*(t) e^{-i\omega t} dt\right] U_T(\mathbf{r}, \omega) e^{-i\omega\tau} d\omega$$

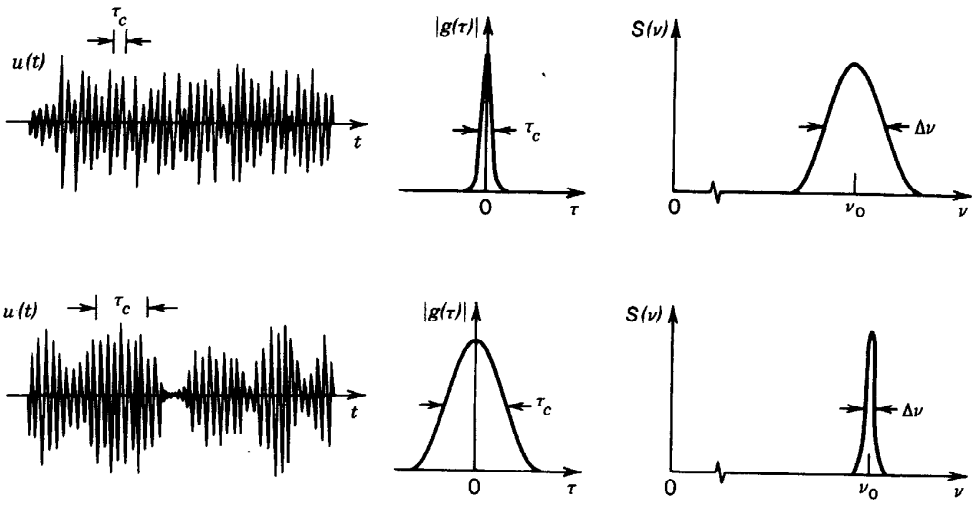
$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |U_T(\omega)|^2 e^{-i\omega\tau} d\omega$$

$$\sim S(\omega) = \lim_{T\to\infty} \frac{1}{2T} |U_T(\omega)|^2$$

$S(\omega)d\omega$ is the spectral intensity between ω and $\omega + d\omega$. Since the Fourier transform of the field does not contain negative frequencies, i.e. $U(-\omega) = 0$, it follows that

$$I = \int_0^{\infty} S(\omega) d\omega$$

Moreover, it holds that the coherence of a light source is given by the spectral intensity.



It holds in general that the product between coherence time and the bandwidth is constant. The longer the coherence time the shorter the bandwidth and vice versa.

$$\tau_c \Delta\omega = \text{const.}$$

where the spectral width is defined as

$$\Delta\omega = \frac{\left[\int_0^{\infty} S(\omega) d\omega\right]^2}{\int_0^{\infty} S^2(\omega) d\omega}$$

Therefore, spectral filters can be used to improve the coherence, but this is always accompanied by a loss of intensity.

A limiting example to be considered is a monochromatic wave. There, $\Delta\omega \to 0$ and we have maximized the temporal coherence, i.e. $\tau_c \to \infty$. If $\Delta\omega \ll \omega_0$ or $\tau_c \gg t_{system}$, we would speak of a quasi-monochromatic field that behaves as coherent light.

d) Mutual coherence function (spatial and temporal coherence)

The mutual coherence function $G(\mathbf{r}_1, \mathbf{r}_2, \tau)$ is given as the cross correlation function between the scalar field at two different spatial positions and at two different moments in time, i.e. between $u(\mathbf{r}_1, t_1)$ and $u(\mathbf{r}_2, t_2)$ with $\tau = t_2 - t_1$.

$$G(\mathbf{r}_1, \mathbf{r}_2, \tau) = \langle u^*(\mathbf{r}_1, t) u(\mathbf{r}_2, t+\tau) \rangle$$

Then, the complex degree of the degree of mutual coherence is defined accordingly as

$$g(\mathbf{r}_1, \mathbf{r}_2, \tau) = \frac{G(\mathbf{r}_1, \mathbf{r}_2, \tau)}{\sqrt{I(\mathbf{r}_1) I(\mathbf{r}_2)}}$$

The function is equally bound between

$$0 \leq |g(\mathbf{r}_1, \mathbf{r}_2, \tau)| \leq 1$$

The complex degree of the degree of mutual coherence is a measure for the correlation of the field between $(\mathbf{r}_1, t)$ and $(\mathbf{r}_2, t+\tau)$.

Example: plane, monochromatic wave

$$|g(\mathbf{r}_1, \mathbf{r}_2, \tau)| = \left| e^{i\mathbf{k}(\mathbf{r}_1 - \mathbf{r}_2) - i\omega\tau} \right| = 1$$

The function collapses to an expression

(a) for temporal coherence if $\mathbf{r}_1 = \mathbf{r}_2$

(b) for spatial coherence for $\tau = 0$

e) Mutual intensity – spatial coherence function

This concerns the explicit details to the special case mentioned above with $\tau = 0$.

$$G(\mathbf{r}_1, \mathbf{r}_2, 0) = \langle u^*(\mathbf{r}_1, t) u(\mathbf{r}_2, t) \rangle = G(\mathbf{r}_1, \mathbf{r}_2)$$

$G(\mathbf{r}_1, \mathbf{r}_2)$ is called the mutual intensity. The normalized mutual intensity $g(\mathbf{r}_1, \mathbf{r}_2)$ reads as

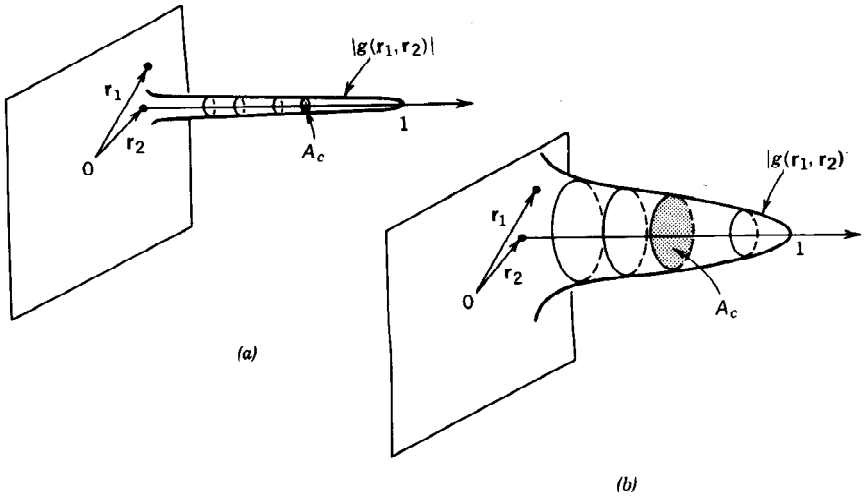
$$g(\mathbf{r}_1, \mathbf{r}_2) = \frac{G(\mathbf{r}_1, \mathbf{r}_2)}{\sqrt{I(\mathbf{r}_1) I(\mathbf{r}_2)}}$$

Comment:

If the optical path difference in an optical system is smaller than the coherence length, spatial and temporal coherence can be separated from each other, and the light wave is temporally coherent; just as mentioned above, this is called quasi-monochromatic light

$$G(\mathbf{r}_1, \mathbf{r}_2, \tau) = G(\mathbf{r}_1, \mathbf{r}_2) e^{-i\omega_0 \tau} = G(\mathbf{r}_1, \mathbf{r}_2)\, G(\tau)$$

Impact of an aperture on the area of coherence (spatial domain where the spatial degree of coherence dropped by 1/e)



If the aperture images a spot that is smaller than the area of coherence, $|g| = 1$ and effectively the light is coherent. On the other hand, if the spatial resolution only allows to glimpse into an area that is larger than the are of coherence, $|g| \ll 1$ and the light is effectively incoherent. A thermal source of radiation, for example, has a area of coherence that is in the order of λ^2, so except in high resolution systems this light source will be perceived as incoherent.

4.3. Interference of partially coherent light

4.3.1. Interference at one point of two partially coherent waves

In the notation introduce above we consider here two random fields $u_1(\mathbf{r}, t)$ and $u_2(\mathbf{r}, t)$. They have the intensity

$$I_1 = \langle |u_1(\mathbf{r}, t)|^2 \rangle \quad \text{and} \quad I_2 = \langle |u_2(\mathbf{r}, t)|^2 \rangle$$

at point $\mathbf{r}$. They are characterized by the cross correlation function

$$G_{12} = \langle u_1^* u_2 \rangle$$

and the respective normalized quantity, the mutual complex degree of coherence

$$g_{12} = \frac{\langle u_1^* u_2 \rangle}{\sqrt{I_1 I_2}}$$

Please note here the difference between the quantities with and without subscript as used thus far.

The interference between both fields gives

$$\begin{aligned} I(\mathbf{r}) &= \langle |u_1 + u_2|^2 \rangle = \langle |u_1|^2 \rangle + \langle |u_2|^2 \rangle + \langle u_1^* u_2 \rangle + \langle u_1 u_2^* \rangle \\ &= I_1 + I_2 + G_{12} + G_{12}^* = I_1 + I_2 + 2\Re(G_{12}) \\ &= I_1 + I_2 + 2\sqrt{I_1 I_2}\Re(g_{12}) \end{aligned}$$

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2}|g_{12}| \cos\phi$$

where $\phi = \arg g_{12}$.

Now different cases can be discussed

- $|g_{12}| = 1 \rightarrow$ interference pattern visible just as for coherent light
- $|g_{12}| = 0 \rightarrow I = I_1 + I_2$ no interference just as for incoherent light
- in general, $0 \leq |g_{12}| \leq 1$, interference pattern with some visibility or contrast according to

$$v = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = 2\frac{\sqrt{I_1 I_2}}{I_1 + I_2}|g_{12}|$$

for $I_1 = I_2$ this expression reduces to

$$v = |g_{12}|$$

4.3.2. Interference and temporal coherence

In the following we consider a signal at the same spatial location but at two different times.

$$\begin{aligned} u_1 = u(t) \quad &\rightarrow \quad I_0 = \langle |u(t)|^2 \rangle \\ u_2 = u(t+\tau) \quad &\rightarrow \quad I_0 = \langle |u(t+\tau)|^2 \rangle \end{aligned}$$

It is the same but just at a different time. Therefore, time averaged they have the same intensity. With this we have

$$g_{12} = \frac{\langle u_1^* u_2 \rangle}{I_0} = \frac{\langle u^*(t) u(t+\tau) \rangle}{I_0} = g(\tau)$$

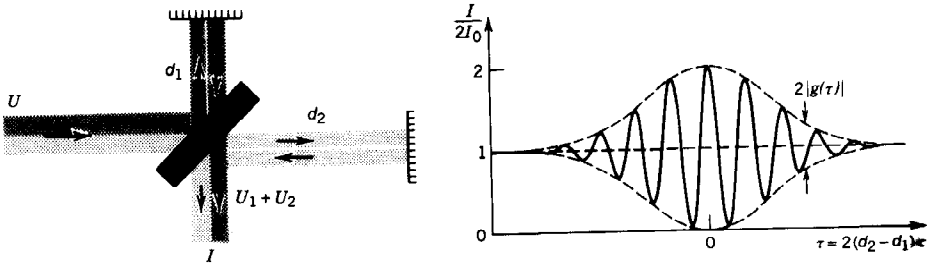
The interference pattern can be calculated according to

$$I(\mathbf{r}) = I_1 + I_2 + 2\sqrt{I_1 I_2}\Re(g_{12})$$

$$I(\tau) = 2I_0\{1 + \Re(g(\tau))\} = 2I_0\{1 + |g(\tau)|\cos\phi(\tau)\}$$

The same discussion as performed above applies. Depending on $|g(\tau)|$ we can define a visibility $v = |g(\tau)|$ and the $\cos\phi(\tau)$ only defines the spatial position of the interference pattern. We have perfect contrast ($v = 1$) for $\tau = 0$ and a vanishing contrast ($v = 0$) for $\tau > \tau_c$.

A measurement can be realized, e.g., with a Michelson or Mach-Zehnder interferometer.



Here, τ is defined as the path difference in the two arms, i.e. $\tau = \frac{2(d_2 - d_1)}{c}$.

For quasi-monochromatic light ($\omega_0 \gg \Delta\omega$)

$$u(t) = a(t)e^{-i\omega_0 t}$$

$$g(\tau) = g_a(\tau)e^{-i\omega_0\tau} = |g_a(\tau)|e^{-i(\omega_0\tau - \varphi_a)}$$

$$g_a(\tau) = \frac{\langle a^*(t)a(t+\tau)\rangle}{|a(t)|^2}$$

$$I(\tau) = 2I_0\{1 + |g_a(\tau)|\cos(\omega_0\tau - \phi_a(\tau))\}$$

Here the term $|g_a(\tau)|$ dictates the contrast, $\omega_0\tau$ the period of the interference fringes and $\phi(\tau)$ eventually describes the absolute position.

We can take advantage of these interference phenomena in specific applications, e.g. here shown at the example of the **Fourier-Transform spectroscopy**.

We know from before that

$$I = \int_0^\infty S(\omega)d\omega$$

which necessarily requires that $S(\omega)$ is a real valued function which we defined before as the spectral intensity between ω and $\omega + d\omega$. Following the Wiener-Kintchin theorem we have

$$G(\tau) = g(\tau)I_0 = \int_0^\infty S(\omega)e^{-i\omega\tau}d\omega$$

$$I(\tau) = 2I_0 + 2I_0\Re\big(g(\tau)\big)$$

$$= 2\int_0^\infty S(\omega)d\omega + 2\Re\left(\int_0^\infty S(\omega)e^{-i\omega\tau}d\omega\right)$$

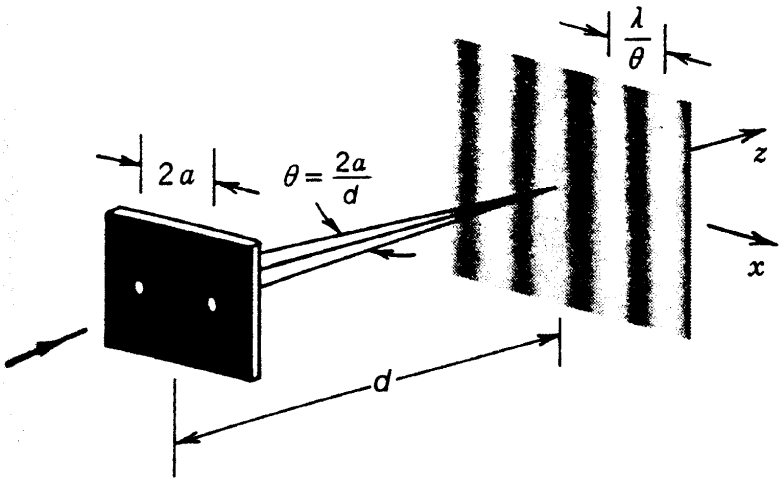
$$= 2\int_0^\infty S(\omega)d\omega + 2\left(\int_0^\infty S(\omega)\cos\omega\tau\, d\omega\right)$$

$$= 2\int_0^\infty S(\omega)[1+\cos\omega\tau]d\omega$$

From these expressions we can see that we can measure $I(\tau)$, making an inverse Fourier transform eventually and we can obtain the spectrum, $\mathrm{FT}^{-1}[I(\tau)] = S(\omega)$. From this quantity we can also extract $G(\tau)$.

4.3.3. Interference and spatial coherence

The classical example for the consideration to understand the concept is a Young interferometer. There, interference at a double slit or circular apertures can be observed.



$u_1 = u_1(\mathbf{r}, t)$ is the light emanating from aperture 1 and $u_2 = u_2(\mathbf{r}, t)$ is the light emanating from aperture 2. The measurement of $I(\mathbf{r}, t)$ done with the following coordinates

$$\mathbf{r}_1 = (-a, 0, 0) \quad \text{and} \quad \mathbf{r}_2 = (a, 0, 0) \quad \text{and} \quad \mathbf{r} = (x, 0, d) \quad \text{and} \quad \Theta \approx \frac{2a}{d}$$

would provide in Fresnel approximation for the light from the first aperture

$$u_1(\mathbf{r},t) = u\left(\mathbf{r}_1, t - \frac{|\mathbf{r}-\mathbf{r}_1|}{c}\right) \approx u\left(\mathbf{r}_1, t - \frac{d + (x+a)^2/2d}{c}\right)$$

and for the second aperture

$$u_2(\mathbf{r},t) = u\left(\mathbf{r}_2, t - \frac{|\mathbf{r}-\mathbf{r}_2|}{c}\right) \approx u\left(\mathbf{r}_2, t - \frac{d + (x-a)^2/2d}{c}\right)$$

Furthermore we assume that $I_1 \approx I_2 \approx I_0$. It was already indicated before that

$$I = 2I_0\{1 + |g_{12}|\cos\phi\}$$

with

$$g_{12} = \frac{\langle u_1^* u_2\rangle}{I_0} = \frac{\langle u^*(\mathbf{r}_1,\bar{t})u(\mathbf{r}_2,\bar{t}+\tau)\rangle}{I_0} = g(\mathbf{r}_1,\mathbf{r}_2,\tau)$$

which is here the mutual complex degree of coherence. In this expression we have

$$\bar{t} = \frac{d + (x+a)^2/2d}{c}$$

$$\tau = \frac{(x+a)^2}{2dc} - \frac{(x-a)^2}{2dc} = \frac{2ax}{dc} = \Theta\frac{x}{c}$$

The degree of coherence can then be determined by measuring the intensity

$$I = 2I_0\{1 + |g(\mathbf{r}_1,\mathbf{r}_2,\tau(x))|\cos\phi(x)\}$$

This is a general expression. If the field is temporally coherent ($\tau < \tau_c$) this can be written using

$$g(\mathbf{r}_1,\mathbf{r}_2,\tau) = g(\mathbf{r}_1,\mathbf{r}_2)e^{-i\omega_0\tau}$$

$$I = 2I_0\left\{1 + |g(\mathbf{r}_1,\mathbf{r}_2)|\cos\left(\omega_0\Theta\frac{x}{c} + \phi(\mathbf{r}_1,\mathbf{r}_2)\right)\right\}$$

The period of the fringe pattern will be $x_P = \frac{\lambda}{\Theta}$, the visibility will be $v = |g(\mathbf{r}_1,\mathbf{r}_2)|$ and the spatial position of the fringes will be given by $\phi(\mathbf{r}_1,\mathbf{r}_2)$.