Propagation of electromagnetic waves
Propagation of electromagnetic waves

Michael Mattes

Spring Term 2016
This document has been typeset using \LaTeX and KOMA-Script.
## Contents

1 Introduction

2 The Maxwell equations
   2.1 Maxwell’s equations in the time domain
   2.2 On the question of current in equation (I)
   2.3 Boundary conditions
   2.4 Energy conservation and electromagnetism
   2.5 Scalar and vector phasors
   2.6 Maxwell’s equations in the frequency domain
   2.7 Invariance of Maxwell’s equations
   2.8 Symmetry properties of Maxwell’s equations
   2.9 The wave equation. Electromagnetic potentials
      2.9.1 Introduction
      2.9.2 Time-varying electromagnetic fields
      2.9.3 Wave equation in the frequency domain
      2.9.4 Scalar and Vector Potentials
      2.9.5 Solution of the wave equation
      2.9.6 Hertz’s electric vector. Polarization potentials
      2.9.7 Duality Principle. Symmetric Maxwell equations
      2.9.8 Hertz’s magnetic vector
   2.10 Summary

3 Velocities of propagation
   3.1 Phase velocity
   3.2 Group velocity
      3.2.1 Wave train of two monochromatic waves
      3.2.2 Wave train with continuous spectrum
   3.3 Velocity of energy flow
   3.4 Signal velocity
   3.5 Concluding remarks

4 Electromagnetic Waveguides
   4.1 Introduction
   4.2 Transmission line theory
      4.2.1 Review
      4.2.2 The Telegraph Equations and their solutions
      4.2.3 Example: coaxial cable and its transmission line parameters
   4.3 Uniform waveguides
      4.3.1 Definition of uniform waveguides
4.3.2 Cylindrical coordinate system .............................................. 63
4.3.3 Electromagnetic waves in cylindrical systems ...................... 64
4.3.4 Method of Hertz Vectors ..................................................... 64
4.3.5 Boundary conditions ......................................................... 66
4.3.6 Separation of variables ...................................................... 67
4.3.7 Types of electromagnetic waves in cylindrical systems ........... 68
4.3.8 Fields related to TEM-, TE- and TM-modes .......................... 73
4.3.9 Discussion of the three types of waveguide modes ............... 77
4.3.10 Eigenfunctions and eigenvalues ......................................... 78
4.3.11 Vector eigenfunctions and mode expansion .......................... 79
4.3.12 Normalized modes ......................................................... 80
4.3.13 Field representation in uniform waveguides: modal expansion .................................................. 84
4.3.14 Eigensolutions of Maxwell’s equations in rectangular waveguides .......................................................... 85
4.3.15 Eigensolutions of Maxwell’s equations in circular waveguides .......................................................... 91
4.3.16 Eigensolutions of Maxwell’s equations in circular coaxial waveguides .......................................................... 92
4.3.17 The equivalent transmission line of a waveguide ............... 92
4.3.18 Frequency characteristics and behaviour ........................... 96

A Exercices ................................................................. 107

B Orthogonal coordinate systems .............................................. 111
  B.1 Cylindrical coordinate systems ......................................... 111
    B.1.1 Rectangular coordinates ................................................. 111
    B.1.2 Circular-cylinder coordinates ........................................ 112
    B.1.3 Elliptic-cylinder coordinates ......................................... 113
    B.1.4 Parabolic-cylinder coordinates ....................................... 114
  B.2 Rotational coordinate systems ......................................... 115
    B.2.1 Spherical coordinates .................................................... 115
    B.2.2 Prolate spheroidal coordinates ....................................... 116
    B.2.3 Oblate spheroidal coordinates ........................................ 117
    B.2.4 Parabolic coordinates ................................................... 118
  B.3 General coordinate systems .............................................. 119
    B.3.1 Conical coordinates ..................................................... 119
    B.3.2 Ellipsoidal coordinates ................................................ 120
    B.4 Paraboloidal coordinates ................................................ 121

C The method of separation of variables .................................. 123

D Sturm-Liouville problem and related theorems .......................... 127
Contents

E Uniform waveguides 131
   E.1 Method of Longitudinal Components . . . . . . . . . . . . . . . . . 131
   E.2 Equivalence between both methods . . . . . . . . . . . . . . . . . 133

F Lorenz or Lorentz? 135
CHAPTER 1

Introduction

The art of the electrical engineering design partly relies on the ability to properly model the physical structure under consideration. A good model enables an efficient and accurate analysis, so that the designer can reach his/her goal with a few iterations on the model and, usually, a few steps of experimental verification.

Most electrical and electronic engineers use circuit-theory models to analyze various passive and active circuits. Such models are simple and straightforward to implement, they do not require bulky theoretical background, and they are easy to visualize. However, they may fail to predict circuit behaviour even at power frequencies, let alone analyse radiation phenomena. Let us not forget that the circuit-theory models need a link to the physical structure they represent to provide meaningful results. For example, we need to know how to evaluate the resistance of a wire to represent it by a resistor.

Electromagnetic field models are predominantly used by antenna and microwave engineers. The analysis starts from the physical structure (i.e., the geometry and electrical properties of materials involved), and it gives a full insight into the properties of devices and circuits (including propagation, radiation, parasitic effects, etc.). Most electromagnetic field problems do not have an analytical solution and a numerical approach is required. However, writing a computer code for the solution of a class of problems is a hard task. Even to properly use codes for the electromagnetic field analysis, a lot of background and experience is required. This software is usually very sophisticated, it covers only a narrow region of applications, and it may sometimes require a long central processor unit (CPU) time to produce results.

At this place, a remark should be made on the dimensionality of the electromagnetic fields and unknowns. An electromagnetic field is always a three-dimensional spatial phenomenon, meaning that it exists within a finite or infinite region (volume). In most cases, the field vectors are functions of three spatial coordinates (e.g., the Cartesian $x$, $y$, and $z$ coordinates), and such problems are referred to as three-dimensional (3D) electromagnetic field problems. In some problems, the fields are functions of only two coordinates. For example, the electrostatic field of an infinitely long two-wire line depends only on the transverse coordinates. The related problems are referred to as two-dimensional (2D) problems. Even simpler cases are when the fields depend only on one spatial coordinate. For example, the electric and magnetic fields of a uniform plane wave depend only on the longitudinal coordinate. In such cases we speak about one-dimensional (1D) problems.

The dimensionality of an electromagnetic field problem should not be confused
with the *mathematical* dimensionality of the unknowns. They may or may not coincide. For example, when the unknowns are fields in a 3D electromagnetic problem, the unknowns are also functions of three spatial coordinates, and we have a 3D *mathematical* problem. However, if we solve for the field sources, the situation may be different. For example, if we analyze scattering from a rectangular metallic plate in a vacuum, the unknowns are currents induced on the plate, which depend on two local coordinates associated with the plate. Hence, the unknowns constitute a 2D mathematical problem. If we consider scattering from a thin wire in a vacuum, the unknown is the current distribution along the wire, and we have an 1D mathematical problem.

For the analysis in the time domain, the temporal variable increases the mathematical dimensionality of the problem by one. In this lecture, however, we deal mainly with the frequency-domain analysis.

Another point is that the electric and magnetic field are three-dimensional vector-fields the components of which can depend on one, two or three variables. When speaking about the dimensionality of an electromagnetic field problem care must be therefore taken what is meant with dimension:

a) the dimension of the vector field,

b) the dependency of the problem on the number of coordinates,

c) the dimension of the underlying mathematical problem, i. e. the dimensionality of the unknowns (e. g. surface current).
CHAPTER 2

The Maxwell equations

2.1 Maxwell’s equations in the time domain

The discussion of Maxwell’s equations here is primarily to develop the formulae to be used in subsequent chapters. This chapter is not intended as a comprehensive review of Maxwell’s equations but rather as basis of common notation.

The complete set of Maxwell’s equations in differential form is written in Table 2.1 in the time domain. Essentially, Maxwell’s equations define the relations between the sources and the electromagnetic phenomena for every point in the space. The electromagnetic phenomena are the four field quantities $E$, $H$, $D$, and $B$, which represent the electric and magnetic field and the electric and magnetic flux (also referred to as the magnetic induction), respectively. These phenomena are generated by the sources $\rho_e$ and $J$ that are the density of the electric charge and the electric current. Although not explicitly mentioned, all quantities depend on time and location, i.e. for example $E = E(\vec{r}, t)$, where $\vec{r}$ is an arbitrary point in the three dimensional space.

The first equation relates the magnetic field $H$ to the electric current $J$ and Maxwell’s displacement current $\partial D/\partial t$. Until Maxwell’s work, the known laws of electricity and magnetism did not include this displacement current. In particular, the equation for the magnetic field of steady currents, i.e. time-independent currents, was known only as

$$\text{rot} \, B = J.$$  \hspace{1cm} (2.1)

Maxwell began by considering these known laws and expressing them as differential equations. He then noticed that there was something strange about (2.1). If one takes the divergence of (2.1), the left-hand side will be zero, because the divergence of a rotation is always zero ($\text{div} \, \text{rot} \, \mathbf{A} = 0$). So, this equation requires that the divergence of the electric current $\mathbf{J}$ vanishes, too. That is, $\text{div} \, \mathbf{J} \equiv 0$. But if the divergence of $\mathbf{J}$ is zero, then the total flux of current out of any closed surface is zero, too.

On the other hand, the flux of current from a closed surface is the decrease of the charge inside the surface. This certainly cannot in general be zero because we simply can move the charges inside the surface from one place to another. But this leads to a current, as shown by the law of charge conservation

$$\text{div} \, \mathbf{J} = -\frac{\partial \rho_e}{\partial t}.$$  \hspace{1cm} (2.2)

This equation expresses the fundamental law that electric charges are con-
served, in other words, they cannot be destroyed. Any flow of charge must come from some supply. Maxwell appreciated this difficulty and proposed that it could be avoided by adding the term $\partial D/\partial t$ to the right-hand side of (2.1). He then got the first equation in Table 2.1

$$\text{rot } \mathcal{H} = \mathcal{J} + \frac{\partial D}{\partial t}.$$  (I)

With this definition the law of charge conservation is implicitly included in Maxwell’s equation. Nevertheless, we have added it to Table 2.1 to emphasize its importance.

The second equation is Faraday’s law and represents the law of induction. The third equation represents for the magnetic field what the last equation is for the electric field. The flux of $\mathcal{E}$ through any closed surface is proportional to the charge inside. This is known as Gauss’ law. Since there are no magnetic charges, the right-hand side of (III) has to be zero.

Besides the Maxwell equations and the law of charge conservation, we have what we called constitutive equations. These equations represent the relation between electromagnetic flux ($D$, $B$) and electromagnetic field ($\mathcal{E}$, $\mathcal{H}$). This relation is a property of the material the electromagnetic field passes through. In general, the relation can be a non-linear one. For example in ferromagnets, the law relating $\mathcal{H}$ and $B$ is given by a hysteresis loop. Another example for a non-linear relationship between flux and field are plasma.

However, in most of the cases and so in our one, too, the relationship is linear. This linear law takes on a very simple form especially in the frequency domain as we will see later.

### 2.2 On the question of current in equation (I)

In equation (I), we denoted by $\mathcal{J}$ the electric current that generates the magnetic field $\mathcal{H}$. This current is, in general, composed of two parts: an impressed current and an induced one. By impressed, we will understand a source that will not be influenced by its environment, especially by fields produced by other sources. The impressed current source can be seen as an ideal current generator in circuit theory as in Figure 2.2.

On the other hand, there is the induced part of $\mathcal{J}$, which we denoted by $\mathcal{J}_f$, since it emerges from free charges. In Figure 2.1 we have an impressed current $\mathcal{J}_i$ that generates a magnetic field $\mathcal{H}_i$. When this magnetic field hits the metallic obstacle, it produces a current $\mathcal{J}_f$ on its surface due to the free electrons in the metal. This current, in turn, is going to generate the magnetic field $\mathcal{H}_s$. The subscript $s$ comes from scattered, since the metallic obstacle acts like a scatterer. By definition, this scattered field $\mathcal{H}_s$ cannot modify the impressed source $\mathcal{J}_i$. On the contrary, a different current $\mathcal{J}_i$ will generate a different magnetic field $\mathcal{H}_s$, which in turn results in a different induced current $\mathcal{J}_f$. Together, $\mathcal{H}_i$ and $\mathcal{H}_s$ form the total magnetic field $\mathcal{H}$ in equation (I) and $\mathcal{J}_i$ and $\mathcal{J}_f$ the total current $\mathcal{J}$. From the view point of circuit theory, the current $\mathcal{J}_f$ corresponds to a current
### Maxwell’s equations

<table>
<thead>
<tr>
<th></th>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$\text{rot } \mathcal{H} = \mathcal{J} + \frac{\partial \mathcal{D}}{\partial t}$</td>
<td><strong>Ampère-Maxwell’s law</strong>&lt;br&gt;Line integral of the magnetic field around a loop = electric current through the loop + the time variation of the electric flux through the loop</td>
</tr>
<tr>
<td>II</td>
<td>$\text{rot } \mathcal{E} = -\frac{\partial \mathcal{B}}{\partial t}$</td>
<td><strong>Faraday’s law</strong>&lt;br&gt;Line integral of the electric field around a loop = time variation of the flux of the magnetic field through the loop</td>
</tr>
<tr>
<td>III</td>
<td>$\text{div } \mathcal{B} = 0$</td>
<td>Flux of the magnetic field through a closed surface = 0 (no magnetic charges)</td>
</tr>
<tr>
<td>IV</td>
<td>$\text{div } \mathcal{D} = \rho_e$</td>
<td><strong>Gauß’ law</strong>&lt;br&gt;Flux of the electric field through a closed surface = charge inside</td>
</tr>
</tbody>
</table>

### Conservation of charge

$$\text{div } \mathcal{J} + \frac{\partial \rho_e}{\partial t} = 0$$

### Constitutive equations

<table>
<thead>
<tr>
<th>General media</th>
<th>$\mathcal{D} = \mathcal{D}(\mathcal{E})$</th>
<th>$\mathcal{B} = \mathcal{B}(\mathcal{H})$</th>
<th>$\mathcal{J}_f = \mathcal{J}_f(\mathcal{E})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current of free charges</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.1** Maxwell’s equations in the *time* domain, the law of charge conservation, and constitutive equations.
The impressed current is not changed by the total magnetic field $\mathcal{H} = \mathcal{H}_i + \mathcal{H}_s$. On the other hand, the induced current $\mathcal{J}_f$, generated by the field $\mathcal{H}_i$, will change when $\mathcal{H}_i$ is modified. $\mathcal{J}_f$ produces the scattered field $\mathcal{H}_s$, so that $\mathcal{H}$ satisfies the boundary conditions on the surface of the metallic obstacle.

produced by a real current generator, that changes if the load $Z$ changes, i.e., the environment.

### 2.3 Boundary conditions

Up to now, we considered Maxwell’s equations only in a homogeneous, infinite medium. In this section, we will recall the boundary conditions of the electromagnetic field at the interface between two different media $\mathcal{E}_1$ and $\mathcal{E}_2$ (see Figure 2.3).

The electric fields $\mathcal{E}_1$ and $\mathcal{E}_2$ are subject to the boundary condition

$$
(\mathcal{E}_1 - \mathcal{E}_2)|_S \times \mathbf{n} = 0,
$$

(2.3)

where $\mathbf{n}$ is a unit vector normal to the interface pointing from medium $\mathcal{E}_2$ to medium $\mathcal{E}_1$. (2.3) states that the tangential (with respect to the interface) component of the electric field has to be continuous.

For the normal component, the electric flux is considered which has to satisfy

$$
\mathbf{n} \cdot (\mathcal{D}_1 - \mathcal{D}_2) = \rho_s|_S,
$$

(2.4)

with $\rho_s$ being the surface charge located in the interface.

The boundary condition for the magnetic field is

$$
\mathbf{n} \times (\mathcal{H}_1 - \mathcal{H}_2) = \mathcal{J}_s|_S,
$$

(2.5)
Figure 2.2 The ideal generator produces the impressed current $J_i$. Due to the internal impedance $Z_i$ of the real generator, the current $J_f$ changes if the load $Z$ is altered. This is comparable to the difference between impressed and induced currents in Maxwell's equations: the induced current $J_f$ is susceptible to the environment, whereas $J_i$ is not.

Figure 2.3 Boundary conditions of the electromagnetic field at the interface $S$ between the two media ① and ②.

where $J_s$ is the surface current in the interface. For the magnetic flux we have

$$n \cdot (B_1 - B_2)|_S = 0,$$  \hspace{1cm} (2.6)

The fact that the right-hand side of (2.6) is zero, is a consequence of (III), since magnetic charges are not known to exist.

2.4 Energy conservation and electromagnetism: Poynting's theorem

In this section, we want to express the conservation of energy for electromagnetism, often called Poynting's theorem. To do that, we have to describe how much energy there is in any volume element of space, and also the rate of energy flow. Suppose we think first only of the electromagnetic field energy. We will let $w$ represent the energy density in the field, that is, the amount of energy per unit volume in space. The energy flux will be represented by $S$, that is, the flow
of energy per unit time across a unit area perpendicular to the flow. Then, in perfect analogy with the conservation of charge, (2.2), the “local” law of energy conservation in the field can be written as

\[
\frac{\partial w}{\partial t} = - \text{div} \mathbf{S}.
\] (2.7)

Of course, this law is not true in general; it is not true that the field energy is conserved. Suppose you are in a dark room and then turn on the light switch. Suddenly, the room is full of light, so there is energy in the field, although there was not any energy there before. (2.7) is not the complete conservation law, because the field energy alone is not conserved, only the total energy in the world. The field energy will change if there is some work being done by matter on the field or by the field on matter.

Let us come back to the field energy. The total field energy in a given volume \( V \) decreases either because field energy flows out of the volume or because the field loses energy to matter (or gains energy, which is just a negative loss). The field energy inside a volume \( V \) is

\[
\int_V w \, dV,
\] (2.8)

and its rate of decrease is minus the time derivative of this integral. The flow of field energy out of the volume \( V \) is the integral of the normal component of \( \mathbf{S} \) over the surface \( \Sigma = \partial V \) that encloses \( V \),

\[
\int_\Sigma \mathbf{n} \cdot \mathbf{S} \, dS,
\] (2.9)

where \( \mathbf{n} \) is the outward normal to \( \Sigma \).

The integral form of (2.7) becomes therefore (Gauß' theorem)

\[
- \frac{\partial}{\partial t} \int_V w \, dV = \int_\Sigma \mathbf{n} \cdot \mathbf{S} \, dS + (\text{work done on matter inside } V) .
\] (2.10)

The work done by the field on each unit volume of matter is \( \mathbf{J} \cdot \mathbf{E} \). So, the quantity \( \mathbf{J} \cdot \mathbf{E} \) must be equal to the loss of energy per unit time and per unit volume by the field. (2.10) then becomes

\[
- \frac{\partial}{\partial t} \int_V w \, dV = \int_\Sigma \mathbf{n} \cdot \mathbf{S} \, dS + \int_V \mathbf{J} \cdot \mathbf{E} \, dV.
\] (2.11)

This is our conservation law of energy in the field. It can be converted into a differential equation like (2.7) if we change the first term on the right-hand side

\[1\)The force on a particle is given by the Lorentz force \( \mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \), and the rate of doing work is \( \mathbf{F} \cdot \mathbf{v} = q \mathbf{E} \cdot \mathbf{v} \). If there are \( N \) particles per unit volume, the rate of doing work per unit volume is \( Nq \mathbf{E} \cdot \mathbf{v} \), but \( Nqv = \mathbf{J} \).
into a volume integral. This is done by means of Gauß’ theorem and we obtain

$$- \frac{\partial w}{\partial t} = \text{div } S + \mathcal{J} \cdot \mathcal{E}. \quad (2.12)$$

To find the expressions for the energy density $w$ and the flux $S$, we reconsider (2.11) and write it slightly differently as

$$- \int_V \mathcal{J} \cdot \mathcal{E} \, dV = \frac{\partial}{\partial t} \int_V w \, dV + \int_{\Sigma} n \cdot S \, dS. \quad (2.13)$$

Since we are interested in the energy density and flow in the field, the left-hand side of (2.13) has to be expressed by means of the electromagnetic field. This we can do by using equation (I) of Maxwell’s equation and replacing $\mathcal{J}$. We obtain

$$- \int_V \left( \mathcal{E} \cdot \text{rot } \mathcal{H} - \mathcal{E} \cdot \frac{\partial \mathcal{D}}{\partial t} \right) \, dV = \frac{\partial}{\partial t} \int_V w \, dV + \int_{\Sigma} n \cdot S \, dS. \quad (2.14)$$

Now we apply the vector identity

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

to the last equation and after using (II) and rearranging the terms on the left-hand side we get

$$\int_V \left[ \mathcal{E} \cdot \frac{\partial \mathcal{D}}{\partial t} + \mathcal{H} \cdot \frac{\partial \mathcal{B}}{\partial t} + \text{div}(\mathcal{E} \times \mathcal{H}) \right] \, dV = \frac{\partial}{\partial t} \int_V w \, dV + \int_{\Sigma} n \cdot S \, dS. \quad (2.15)$$

We nearly arrived now. We can identify by using Gauß’ theorem that the Poynting vector $S$ is given by

$$S = \mathcal{E} \times \mathcal{H}. \quad (2.16)$$

It defines the energy flow.

The energy density can be associated with the left term on the left-hand side, that is

$$\frac{\partial w}{\partial t} = \mathcal{E} \cdot \frac{\partial \mathcal{D}}{\partial t} + \mathcal{H} \cdot \frac{\partial \mathcal{B}}{\partial t}. \quad (2.17)$$

### 2.5 Scalar and vector phasors

In principle, the field vectors can be an arbitrary function of time subject to certain boundary conditions. For engineering applications (e.g. narrowband signals), it is often sufficient to assume a steady-state (sinusoidal) regime.

Before we introduce Maxwell’s equations in the frequency domain, we will recall the concept of phasors (complex vectors) in this section. We shall reveal the canonical form of a sinusoidal scalar quantity on the example of a current that is a sinusoidal function of time.

For the description of linear time-invariant systems in the steady-state, we can use a scalar, complex-valued phasor instead of a scalar, real-valued func-
tion having a harmonic time dependence. Let us consider the time-harmonic function
\[ i(t) = \hat{I} \cos(\omega t + \phi_0), \]  
(2.18)
with \( \omega = 2\pi f \). We call \( \hat{I} \) the amplitude and \( \phi_0 \) the phase of \( i(t) \) at the instant \( t = 0 \). Using Euler’s formula, we can rewrite the equation as
\[ i(t) = Re\{\hat{I}e^{j\phi_0}e^{j\omega t}\}. \]  
(2.19)
The new scalar and complex-valued time-independent quantity
\[ I = \hat{I}e^{j\phi_0} \]  
(2.20)
is called phasor of the current. The current \( i(t) \) is obtained using the relation
\[ i(t) = Re\{Ie^{j\omega t}\}, \]  
(2.21)
thus by multiplying the phasor with the time-harmonic variation \( e^{j\omega t} \) and taking the real part of the product \( ^2 \).

A scalar, complex-valued quantity is given by either its Modulus:
\[ |I| = \hat{I}, \]  
and Phase:
\[ \arg(I) = \phi_0, \]
or by itseal part:
\[ Re\{I\} = \hat{I} \cos \phi_0, \]
and imaginary part:
\[ Im\{I\} = \hat{I} \sin \phi_0. \]

This representation can also be applied to vector quantities, the components of which have a sinusoidal time dependence. As an example, let us consider the magnetic field \( \mathcal{H} \) in Cartesian coordinates
\[ \mathcal{H}_x(\vec{r},t) = H_x(\vec{r}) \cos(\omega t + \phi_x(\vec{r})), \]  
(2.22a)
\[ \mathcal{H}_y(\vec{r},t) = H_y(\vec{r}) \cos(\omega t + \phi_y(\vec{r})), \]  
(2.22b)
\[ \mathcal{H}_z(\vec{r},t) = H_z(\vec{r}) \cos(\omega t + \phi_z(\vec{r})) \]  
(2.22c)
with the amplitudes \( H_x, H_y, H_z \) and the corresponding phases \( \phi_x, \phi_y, \phi_z \). In general, these six quantities depend also on the position vector \( \vec{r} \). Analogously to the scalar case, we can define a vector phasor applying the operation \( Re\{\} \) on each component of the vector. Thus, we obtain
\[ \mathcal{H}(\vec{r},t) = Re\{\mathcal{H}(\vec{r})e^{j\omega t}\}, \]  
(2.23)
with \( \mathcal{H} \) being the complex-valued vector phasor. As in the scalar case, the vector phasor does not depend on the time but only on the position \( \vec{r} \). Contrary to

\(^2\)Instead of taking the real part, we also could have taken the imaginary part. This, however, leads to a phase shift of 90°. The essential point is to obtain a real-valued, time-dependent function.
the scalar case, the vector phasor $\mathbf{H}$ cannot be represented, in general, by a modulus and a phase but only by two vectors $\mathbf{H}_r$ and $\mathbf{H}_i$ corresponding to the real and imaginary part of $\mathbf{H}$, thus

$$\mathbf{H} = \mathbf{H}_r + j\mathbf{H}_i.$$  

(2.24)

### 2.6 Maxwell's equations in the frequency domain

With the concept of phasors that we have recalled in the last section it is now very easy to derive Maxwell's equations in the frequency domain. In fact, all we have to do, is to separate the time dependence from the dependence on the location $\mathbf{r}$. For this, we proceed for the electric field, current, and charge distribution as we did it for the magnetic field in the last section. Thus, we assume that

$$\mathcal{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{j\omega t},$$  

(2.25a)

$$\mathcal{H}(\mathbf{r}, t) = \mathbf{H}(\mathbf{r})e^{j\omega t},$$  

(2.25b)

$$\mathcal{J}(\mathbf{r}, t) = \mathbf{J}(\mathbf{r})e^{j\omega t},$$  

(2.25c)

$$\rho_e(\mathbf{r}, t) = \hat{\rho}_e(\mathbf{r})e^{j\omega t}.$$  

(2.25d)

Inserting these definitions into Maxwell's equations (I)-(IV) as well as the law of charge conservation, we obtain the relations in Table 2.2.

For the constitutive equations, we cannot simply replace the time-domain function by the phasor counterparts, because, in general, the relation between flux and field can be non-linear.

However, if we assume that we deal with linear media, these equations take a simple form. Introducing the permittivity $\varepsilon$ and permeability $\mu$, which are tensors, we can express the relation between flux and field as a matrix product between tensor (two-dimensional matrix in our case) and vector field. The same holds for the electric current and the conductivity $\sigma$. The tensor notation simply means that the factor of proportionality between flux and field is different for different directions. If the constant is independent of the direction, then we speak about an isotropic medium. In this case, the tensors reduce to scalar constants $\varepsilon$, $\mu$, and $\sigma$.

$\varepsilon$, $\mu$, and $\sigma$ can be complex in general. Let us consider as an example the permittivity $\varepsilon = \varepsilon' - j\varepsilon''$. We restrict ourselves to a linear, isotropic medium to show the ideas. If we go back to the time domain, then this means that the electric flux $\mathcal{D}(\mathbf{r}, t)$ and the electric field $\mathcal{E}(\mathbf{r}, t)$ oscillate with the same frequency $\omega$, but there is a phase shift between the two quantities. Moreover, the imaginary part of $\varepsilon$ has to be negative, which implies that $\varepsilon'' > 0$. This means, that the electric flux is retarded with respect to the electric field. This is a consequence of the principle of causality: the reaction ($\mathcal{D}$) cannot come in time before the action ($\mathcal{E}$). This is also known as the Kramers-Kronig relation.

In addition to this, $\varepsilon$ can depend on the frequency $\omega$. In this case, the relation between flux and field in the time domain is given by a convolution integral.
The existence of an imaginary part $\varepsilon''$ is due to losses in the medium. Losses, however, correspond to an ohmic current in the medium. This current on the other hand is an induced current. Thus, in linear, isotropic media we can rewrite equation (1) as

$$\text{rot } \mathbf{H} = \mathbf{J}_i + \sigma \mathbf{E} + j\omega \varepsilon' \mathbf{E}. \quad (2.26)$$

Now, we can factor out $\mathbf{E}$ on the right hand side and obtain

$$\text{rot } \mathbf{H} = \mathbf{J}_i + (j\omega \varepsilon' + \sigma) \mathbf{E}. \quad (2.27)$$

The quantity in parentheses can be replaced by the complex permittivity $\varepsilon = \varepsilon' - j\sigma/\omega$, that is, $\varepsilon'' = \sigma/\omega$.

Similar considerations hold for the permeability $\mu$.

The permittivity and permeability of free space are

$$\varepsilon_0 = 8.854 \cdot 10^{-12} \frac{\text{As}}{\text{Vm}} \quad \text{and} \quad \mu_0 = 4\pi \cdot 10^{-7} \frac{\text{Vs}}{\text{Am}}. \quad (2.28)$$

These values are used to introduce the relative permittivity $\varepsilon_r = \varepsilon/\varepsilon_0$ and permeability $\mu_r = \mu/\mu_0$. For instance, we have for water at 1GHz $\varepsilon_r = 80 - j10$ and $\mu_r = 1$.

A further consequence of linear, isotropic media is that only two vectors are necessary to describe the electromagnetic phenomena. In general, the electric field $\mathbf{E}$ and the magnetic field $\mathbf{H}$ are used and their counterparts $\mathbf{D}$ and $\mathbf{B}$ are eliminated.

A summary of electromagnetic quantities together with their units is provided in Table 2.3.

### 2.7 Invariance of Maxwell’s equations

Since space has no preferred direction there is no unique choice of the $x$-, $y$-, and $z$-axes in Maxwell’s equations. In fact the rotation of the coordinate system (frame) and field quantities to a new position is an operation under which the Maxwell equations should remain unchanged. The fact that they do not change for some symmetry operation (e.g., rotation) is called invariance of Maxwell’s equations, i.e., they take on the same form in different coordinate systems. This is a fundamental principle in physics and does not only apply to Maxwell’s equations.

Although the Maxwell equations are invariant with respect to a symmetry operation, any given solution, satisfying some boundary conditions, need not to be.

For instance a wave moving to the right along a transmission line (waveguide) can be transformed under reflection into a wave moving to the left. Thus, the moving wave does change the propagation direction, i.e. it is not invariant with respect to the reflection. However, considering a standing wave with the symmetry plane at a node or maximum is left unchanged by the reflection.

Let us consider a specific example of symmetry operation that leaves the
Maxwell's equations

I  $\text{rot } \mathbf{H} = \mathbf{J} + j\omega \mathbf{D}$  **Ampère-Maxwell's law**
Line integral of the magnetic field around a loop = electric current through the loop + the time variation of the electric flux through the loop

II  $\text{rot } \mathbf{E} = -j\omega \mathbf{B}$  **Faraday's law**
Line integral of the electric field around a loop = time variation of the flux of the magnetic field through the loop

III  $\text{div } \mathbf{B} = 0$  Flux of the magnetic field through a closed surface = 0 (no magnetic charges)

IV  $\text{div } \mathbf{D} = \hat{\rho}_e$  **Gauß' law**
Flux of the electric field through a closed surface = charge inside

**Conservation of charge**

$\text{div } \mathbf{J} + j\omega \hat{\rho}_e = 0$

**Constitutive equations**

**linear media**

\[ \mathbf{D} = \varepsilon \mathbf{E} \]
\[ \mathbf{B} = \mu \mathbf{H} \]
\[ \mathbf{J}_f = \sigma \mathbf{E} \]

**linear isotropic media**

\[ \mathbf{D} = \varepsilon \mathbf{E} \]
\[ \mathbf{B} = \mu \mathbf{H} \]
\[ \mathbf{J}_f = \sigma \mathbf{E} \]

**Table 2.2** Maxwell’s and related equations in the frequency domain.
<table>
<thead>
<tr>
<th>Quantity</th>
<th>Name(s)</th>
<th>SI Units</th>
<th>Alternative Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{E}$</td>
<td>Electric field</td>
<td>$\frac{\text{m} \cdot \text{kg}}{\text{s}^3 \cdot \text{A}}$</td>
<td>$\frac{\text{V}}{\text{m}}$</td>
</tr>
<tr>
<td></td>
<td>Electric field strength</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathbf{H}$</td>
<td>Magnetic field</td>
<td>$\frac{\text{A}}{\text{m}}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Magnetic field strength</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathbf{D}$</td>
<td>Electric displacement</td>
<td>$\frac{\text{A} \cdot \text{s}}{\text{m}^2}$</td>
<td>$\frac{\text{C}}{\text{m}^2}$</td>
</tr>
<tr>
<td>$\mathbf{B}$</td>
<td>Magnetic flux density</td>
<td>$\frac{\text{kg}}{\text{s}^2 \cdot \text{A}}$</td>
<td>$\text{T} = \frac{\text{Wb}}{\text{m}^2} = \frac{\text{V} \cdot \text{s}}{\text{m}^2}$</td>
</tr>
<tr>
<td></td>
<td>Magnetic induction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_e$</td>
<td>Electric (volume) charge density</td>
<td>$\frac{\text{A} \cdot \text{s}}{\text{m}^3}$</td>
<td>$\frac{\text{C}}{\text{m}^3}$</td>
</tr>
<tr>
<td>$\mathbf{J}$</td>
<td>Electric (volume) current density</td>
<td>$\frac{\text{A}}{\text{m}^3}$</td>
<td>-</td>
</tr>
<tr>
<td>$\rho_s$</td>
<td>Electric surface charge density</td>
<td>$\frac{\text{A} \cdot \text{s}}{\text{m}^2}$</td>
<td>$\frac{\text{C}}{\text{m}^2}$</td>
</tr>
<tr>
<td>$\mathbf{J}_s$</td>
<td>Electric surface current density</td>
<td>$\frac{\text{A}}{\text{m}}$</td>
<td>-</td>
</tr>
<tr>
<td>$\rho_s$</td>
<td>Electromagnetic (volume) energy density</td>
<td>$\frac{\text{kg}}{\text{m} \cdot \text{s}^2}$</td>
<td>$\frac{\text{W} \cdot \text{s}}{\text{m}^3}$</td>
</tr>
<tr>
<td>$\mathbf{S}$</td>
<td>Poynting vector</td>
<td>$\frac{\text{kg}}{\text{s}^2}$</td>
<td>$\frac{\text{W}}{\text{m}^2}$</td>
</tr>
<tr>
<td>$\varepsilon_0$</td>
<td>Permittivity of free space</td>
<td>$\frac{\text{A}^2 \cdot \text{s}^4}{\text{kg} \cdot \text{m}^3}$</td>
<td>$\frac{\text{F}}{\text{m}} = \frac{\text{C}}{\text{V} \cdot \text{m}} = \frac{\text{A} \cdot \text{s}}{\text{V} \cdot \text{m}}$</td>
</tr>
<tr>
<td></td>
<td>8.854187817... \cdot 10^{-12}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>Permeability of free space</td>
<td>$\frac{\text{kg} \cdot \text{m}}{\text{A}^2 \cdot \text{s}^2}$</td>
<td>$\frac{\text{H}}{\text{m}} = \frac{\text{Wb}}{\text{A} \cdot \text{m}} = \frac{\text{V} \cdot \text{s}}{\text{A} \cdot \text{m}}$</td>
</tr>
<tr>
<td></td>
<td>$4\pi \cdot 10^{-7}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_0$</td>
<td>Light speed in free space</td>
<td>$\frac{\text{m}}{\text{s}}$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>299 792 458</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.3** Electromagnetic quantities and their units.
Maxwell equations invariant. This example will prepare the discussion of symmetry properties of Maxwell’s equations in Section 2.8. The symmetry properties are closely related to the invariance of Maxwell’s equations.

Assume the fields \((E, H, D, B)\) are a specific solution to Maxwell’s equations in the Cartesian coordinate system \((x, y, z)\). Next, we consider a new coordinate frame \((x', y', z')\), again Cartesian, which is obtained from the original one by replacing \(x\) by \(-x\). That is, we define the symmetry operation, called reflection

\[
T : (x, y, z) \rightarrow (x', y', z') = (-x, y, z).
\]

The solution of Maxwell’s equation in the system \((x', y', z')\) will be denoted by \((E', H', D', B')\).

Since the Maxwell equations must be independent of the chosen coordinate system, they have the same form in the primed system as in the unprimed one. Thus

\[
- \text{rot}' E' = \partial B'/\partial t, \quad \text{rot}' H' = \mathcal{J}' + \partial D'/\partial t, \quad \text{div}' D' = \rho', \quad \text{div}' B' = 0,
\]

where \(\text{rot}' = \nabla' \times\) and \(\text{div}' = \nabla'\) with \(\nabla' = (\partial/\partial x', \partial/\partial y', \partial/\partial z')\). That is, the derivatives are with respect to the primed coordinates.

Obviously, there is a relation between \(\partial/\partial x'\) and \(\partial/\partial x\) due to the transformation \(T\), which is

\[
\frac{\partial E'}{\partial x'} = \frac{\partial E}{\partial x} \cdot \frac{\partial x}{\partial x'} = \frac{\partial E}{\partial x} \cdot (-1) = -\frac{\partial E}{\partial x}
\]

and of course

\[
\frac{\partial E'}{\partial y'} = \frac{\partial E}{\partial y} \quad \text{and} \quad \frac{\partial E'}{\partial z'} = \frac{\partial E}{\partial z}.
\]

The following question arises: Can we find also a relation between \((E', H', D', B')\) and \((E, H, D, B)\)?

To answer this question we will apply the principle of invariance. For this, we consider exemplarily the equation

\[
\text{rot}' H' = \mathcal{J}' + \partial D'/\partial t.
\]

From (2.32) follows

\[
\text{rot}' = \begin{pmatrix} \partial/\partial x' \\ \partial/\partial y' \\ \partial/\partial z' \end{pmatrix} \times \begin{pmatrix} -\partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{pmatrix} \times .
\]

Thus, we obtain
16 The Maxwell equations

\[
\begin{bmatrix}
\frac{\partial H'_x}{\partial y} - \frac{\partial H'_y}{\partial z} \\
\frac{\partial H'_y}{\partial z} - \frac{\partial H'_z}{\partial x} \\
\frac{\partial H'_z}{\partial x} - \frac{\partial H'_x}{\partial y}
\end{bmatrix}
= \begin{bmatrix}
J'_x \\
J'_y \\
J'_z
\end{bmatrix}
+ \begin{bmatrix}
\frac{\partial D'_x}{\partial t} \\
\frac{\partial D'_y}{\partial t} \\
\frac{\partial D'_z}{\partial t}
\end{bmatrix}.
\] (2.35)

Note, we consider the primed solution \((E', H', D', B')\) in the unprimed coordinate system \((x, y, z)\).

It is clear that the vector on the left hand side of (2.35) does not correspond to the \(\text{rot}\) operator which should write

\[
\begin{bmatrix}
\frac{\partial H'_x}{\partial y} - \frac{\partial H'_y}{\partial z} \\
\frac{\partial H'_y}{\partial z} - \frac{\partial H'_z}{\partial x} \\
\frac{\partial H'_z}{\partial x} - \frac{\partial H'_x}{\partial y}
\end{bmatrix}
\] . (2.36)

In order that it represents a \(\text{rot}\) operator we have to change the second and third equation. For this, assume the following relation between \((E', H', D', B')\) and \((E, H, D, B)\)

\[
(H'_x, H'_y, H'_z) = (H_x, -H_y, -H_z),
\] (2.37)

\[
(D'_x, D'_y, D'_z) = (-D_x, D_y, D_z),
\] (2.38)

\[
(J'_x, J'_y, J'_z) = (-J_x, J_y, J_z).
\] (2.39)

Inserting these relations into (2.35) we get

\[
\begin{bmatrix}
\frac{\partial H'_x}{\partial y} + \frac{\partial H'_y}{\partial z} \\
\frac{\partial H'_y}{\partial z} - \frac{\partial H'_z}{\partial x} \\
\frac{\partial H'_z}{\partial x} - \frac{\partial H'_x}{\partial y}
\end{bmatrix}
= \begin{bmatrix}
-J_x \\
J_y \\
J_z
\end{bmatrix}
+ \begin{bmatrix}
\frac{\partial D'_x}{\partial t} \\
\frac{\partial D'_y}{\partial t} \\
\frac{\partial D'_z}{\partial t}
\end{bmatrix}. \]
(2.40)

After multiplying the first equation by \((-1)\), the Maxwell equation

\[
\text{rot } H = J + \partial D/\partial t \]
(2.41)

for the unprimed solution is recovered.

The same procedure can be applied for the other Maxwell equations.

(2.37) is one possible mapping between \((E', H', D', B')\) and \((E, H, D, B)\). Another one is given by
\( (H'_x, H'_y, H'_z) = (-H_x, H_y, H_z), \)  
\( (D'_x, D'_y, D'_z) = (D_x, -D_y, -D_z), \)  
\( (J'_x, J'_y, J'_z) = (J_x, -J_y, -J_z). \)

It is easily verified by inserting (2.42) into Maxwell’s equation that it satisfies them, too.

The mapping (2.37) is called \textit{even} solution and (2.42), accordingly, \textit{odd} solution. This naming convention, however, is arbitrary.

Invariance of Maxwell’s equations means that they have the same form in the primed and unprimed coordinate system. This principle was the basis to find the mappings (2.37) and (2.42) between the two solutions \((E', H', D', B')\) and \((E, H, D, B)\). The example of reflection is only one example of symmetry operation that leaves the Maxwell equation invariant. Others will be discussed in the next section.

\section*{2.8 Symmetry properties of Maxwell’s equations}

Many electromagnetic problems exhibit symmetry properties. Often, symmetry is used in system applications to design devices with certain properties or characteristics. In many cases, the symmetry is a geometrical one, that is, the space or object that is considered has some specific properties with respect to transformations like rotation, translation or reflection.

For example consider the infinite tube in Figure 2.4. It possesses several symmetry properties: its cross section is constant along its longitudinal axis. This symmetry is called \textit{translational} symmetry. Since the electromagnetic field “sees” along the tube’s axis always the same boundary conditions, it cannot vary with the coordinate, let’s call it \(z\), associated with the tube’s axis. This means, the field solution should be independent of \(z\). In other words, we deal with a 2D mathematical problem but the field vectors and the geometry themselves are 3D quantities.

The section of the tube itself has another symmetry: it is a rotational symmetry, i.e. the shape of the section is independent of the angle in circular coordinates (see also Figure 2.5(a))

Another kind of symmetry is presented in Figure 2.5(b). By turning the rectangle by 180° around its centre we recover the same geometry. Since the boundary conditions remain unchanged under a rotation of 180°, the field solution exhibits the same symmetry property. The rotation of 180° can also be represented by two consecutive reflections at the lines \(x' - x\) and \(y' - y\) as indicated in Figure 2.5(c)

As last example, the symmetry of the object in Figure 2.6(a) is considered. The device is a so-called \textit{T-junction} and used to divide signals (signal enters at the foot-point of the T) or merge them (signals enter at the two arms of the T).
The device exhibits two kinds of symmetries: a translational one along the height of the T-junction and reflection symmetry at the axis $x = 0$ (y-axis) as indicated in Figure 2.6(b). The translational symmetry is only valid inside the object, i.e., there are field distributions that do not vary with the height of the device.

Exploiting symmetry properties reduces in many cases tremendously the effort to solve Maxwell’s equations: either when an analytical solution is calculated or, if this one is not accessible, it reduces the numerical effort and therefore CPU time. Another advantage of symmetry is that it is an easy means to check the correctness of the calculated solution and to detect errors.

After this introductory considerations we shall examine a number of transformations with respect to the coordinates and the field vectors that leave Maxwell’s equations invariant. The study of these transformations is important for the study of especially waveguide transitions or junctions like the precedent T-junction. We will not consider all possible symmetry transformations but shall limit our study to the one covering most of the problems: reflections.

A) Reflection plane

Let us reconsider the example of Figure 2.6(b). The reflection plane is $yz$, i.e., $x$ is transformed into $-x$ whereas all other coordinates ($y$ and $z$) remain unchanged.
Chapter 2  
2.8 Symmetry properties of Maxwell’s equations

Figure 2.6 Example of reflection symmetry: the T-junction is symmetric under reflection at the axis $x = 0$ ($y$-axis).

It is the example that we have already encountered in Section 2.7 discussing the invariance of the Maxwell equations.

On the other hand, the components of the electric and magnetic fields must be so transformed as to ensure the invariance of Maxwell’s equations, that is the transformed field should also satisfy the equations. As seen in Section 2.7 this can be achieved by for example this transformation:

$$
x \rightarrow -x, \quad E_x \rightarrow -E_x, \quad H_y \rightarrow -H_y, \quad J_x \rightarrow -J_x, \quad J_z \rightarrow -J_z.
$$

The other components ($E_y, E_z, H_x, J_y, J_z, \rho$) remain unchanged. That Maxwell’s equations remain invariant can be verified by simple inspection. Stated in words, the above transformation replaces the electric field at the point $(x, y, z)$ by that at the point $(-x, y, z)$, changing the sign of its $x$-component. As already seen this is not the only possibility and we saw that a second transformation leaves Maxwell’s equations invariant, too, with respect to a reflection in the $yz$-plane.

The field thus transformed gives a new solution of Maxwell’s equations. In general, however, the solution thus obtained does not satisfy the original boundary conditions. A necessary condition for the boundary conditions to be fulfilled is that the geometrical structure considered will remain invariant for the given transformation, too. For example in the case of the T-junction of Figure 2.6(b) the boundary conditions are symmetric with respect to the axis $y$. Or, for the circular tube in Figure 2.4 the boundary conditions for the electromagnetic field do not change along the guide’s axis.

**B) Reflection operator**

The above defined reflection can mathematically be represented by the so-called reflection operator. This is just a way for writing shortly $x \rightarrow -x$ using the notation

$$
\mathcal{L}_x(x) = -x.
$$
The index $x$ of $\mathcal{L}$ means that we deal with a reflection in $x$-direction, i.e. with respect to the $yz$-plane. Analogously, for the $x$-component of the electric field it holds

$$\mathcal{L}_x(E_x(x,y,z)) = \alpha E_x(-x,y,z), \quad (2.46b)$$

with $\alpha$ a constant factor. In the example (2.45), $\alpha$ was $-1$.

The operator $\mathcal{L}_x$ is linear, i.e.

$$\mathcal{L}_x(\alpha u + \beta v) = \alpha \mathcal{L}_x(u) + \beta \mathcal{L}_x(v). \quad (2.47)$$

To find $\alpha$ we apply $\mathcal{L}_x$ a second time, but this time to the already transformed field, i.e.

$$\mathcal{L}_x(\mathcal{L}_x(E_x(x,y,z))) = \mathcal{L}_x(-x,y,z) = \alpha \mathcal{L}_x(E_x(x,y,z)) = \alpha^2 E_x(x,y,z). \quad (2.48)$$

Thus, we reflect the already reflected field again. But this means that the original field is reflected on itself and therefore it must hold

$$\mathcal{L}_x(\mathcal{L}_x(E_x(x,y,z))) = E_x(x,y,z), \quad (2.49)$$

from which follows that

$$\alpha^2 = 1 \quad \iff \quad \alpha = \pm 1. \quad (2.50)$$

The solutions associated with $\alpha = -1$ are called even, those for $\alpha = +1$ odd solutions. Thus, for the transformation $\mathcal{L}_x$ we find for the previous example

$$\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \longrightarrow \vec{r}' = \begin{pmatrix} -x \\ y \\ z \end{pmatrix},$$

even solution: $\mathbf{E}(\vec{r}) \longrightarrow \mathbf{E}'(\vec{r}') = \begin{pmatrix} -E_x \\ E_y \\ E_z \end{pmatrix}, \quad \mathbf{H}(\vec{r}) \longrightarrow \mathbf{H}'(\vec{r}') = \begin{pmatrix} H_x \\ -H_y \\ -H_z \end{pmatrix}, \quad (2.51)$

odd solution: $\mathbf{E}(\vec{r}) \longrightarrow \mathbf{E}'(\vec{r}') = \begin{pmatrix} E_x \\ -E_y \\ -E_z \end{pmatrix}, \quad \mathbf{H}(\vec{r}) \longrightarrow \mathbf{H}'(\vec{r}') = \begin{pmatrix} -H_x \\ H_y \\ H_z \end{pmatrix}.$

By convention we call a solution even if the tangential electric field is an even function with respect to the reflection plane. Similarly, for the odd solution they are odd functions.

If the solution is continuous on the plane of symmetry ($x = 0$) then we obtain from the above solutions for
Chapter 2  2.8 Symmetry properties of Maxwell’s equations

even solutions: \( E_x(0, y, z) = H_y(0, y, z) = H_z(0, 0, z) = 0 \), \( E_y(0, y, z) = E_z(0, y, z) = H_x(0, 0, z) = 0 \).

odd solutions: \( E_y(0, y, z) = E_z(0, y, z) = 0 \).

\[ (2.52) \]

C) Rotation

We also can introduce operators \( \mathcal{L}_y \) and \( \mathcal{L}_z \) that represent, respectively, reflections along \( y \) (xz-plane) and \( z \) (xy-plane). More generally, by noting, for example, that a reflection at \( yz \) followed by a reflection at \( xz \) is equivalent to a rotation by 180° about the \( z \)-axis, we are led to utilize a new rotation operator \( \mathcal{R}_z \) such that

\[ \mathcal{R}_z = \mathcal{L}_x \mathcal{L}_y = \mathcal{L}_y \mathcal{L}_x. \]

Similarly, we can define \( \mathcal{R}_x \) and \( \mathcal{R}_y \). Finally, the operator

\[ \mathcal{R} = \mathcal{L}_x \mathcal{L}_y \mathcal{L}_z \]

specifies that there is reflection with respect to the origin\(^3\).

For example, if we use the operator \( \mathcal{R}_z \) we obtain

\[ \vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \vec{r}' = \begin{pmatrix} -x \\ -y \\ z \end{pmatrix}, \]

\( \text{even solution: } \mathbf{E}(\vec{r}) \rightarrow \mathbf{E}'(\vec{r}') = \begin{pmatrix} -E_x \\ -E_y \\ E_z \end{pmatrix}, \quad \mathbf{H}(\vec{r}) \rightarrow \mathbf{H}'(\vec{r}') = \begin{pmatrix} -H_x \\ -H_y \\ H_z \end{pmatrix}, \)

\( \text{odd solution: } \mathbf{E}(\vec{r}) \rightarrow \mathbf{E}'(\vec{r}') = \begin{pmatrix} E_x \\ E_y \\ -E_z \end{pmatrix}, \quad \mathbf{H}(\vec{r}) \rightarrow \mathbf{H}'(\vec{r}') = \begin{pmatrix} H_x \\ H_y \\ -H_z \end{pmatrix}. \]

\[ (2.55a) \]

If the field has to be continuous at the rotation axis \( x = y = 0 \) then we get

\( \text{even solutions: } E_x(0, 0, z) = E_y(0, 0, z) = H_x(0, 0, z) = H_y(0, 0, z) = 0, \)
\( \text{odd solutions: } E_z(0, 0, z) = H_z(0, 0, z) = 0. \)

\[ (2.55b) \]

Similarly, Maxwell’s symmetric solutions are obtained by means of the opera-

\(^3\)Mathematicians call this group of reflections an Abelian group. The elements \( \mathcal{I}, \mathcal{R}_x, \mathcal{R}_y, \mathcal{R}_z \) form a subgroup, \( \mathcal{I} \) denoting the identity transformation, i. e. \( \mathcal{I}(x) = x. \)
tor \( \mathcal{R} \)

\[
\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \rightarrow \quad \vec{r}' = \begin{pmatrix} -x \\ -y \\ -z \end{pmatrix},
\]

**even** solution: \( \mathbf{E}(\vec{r}) \rightarrow \mathbf{E}'(\vec{r}') = \begin{pmatrix} E_x' \\ E_y' \\ E_z' \end{pmatrix} \), \( \mathbf{H}(\vec{r}) \rightarrow \mathbf{H}'(\vec{r}') = \begin{pmatrix} -H_x' \\ -H_y' \\ -H_z' \end{pmatrix} \),

**odd** solution: \( \mathbf{E}(\vec{r}) \rightarrow \mathbf{E}'(\vec{r}') = \begin{pmatrix} -E_x' \\ -E_y' \\ -E_z' \end{pmatrix} \), \( \mathbf{H}(\vec{r}) \rightarrow \mathbf{H}'(\vec{r}') = \begin{pmatrix} H_x' \\ H_y' \\ H_z' \end{pmatrix} \).

From the above result we can, without difficulty, deduce the solution according to the origin \( x = y = z = 0 \) when we use the operator \( \mathcal{R} \). In the case of continuity at the origin we have

**even solutions:** \( H_x(0,0,0) = H_y(0,0,0) = H_z(0,0,0) = 0 \),

**odd solutions:** \( E_x(0,0,0) = E_y(0,0,0) = E_z(0,0,0) = 0 \).

\[(2.56a)\]

### 2.9 The wave equation. Electromagnetic potentials

#### 2.9.1 Introduction

The electromagnetic fields of boundary-value problems are obtained as solutions to Maxwell’s equations, which are first-order partial differential equations. However, they are coupled, that is, each equation has more than one unknown field. These equations can be uncoupled only at the expense of raising their order. For each of the fields, following such a procedure leads to an uncoupled second-order partial differential equation that is usually referred to as the wave equation. Therefore electric and magnetic fields for a given boundary-value problem can be obtained either as solutions to Maxwell’s or to the wave equation. The choice of equations is related to individual problems by convenience and ease of use. In this section we will develop the vector wave equations for each of the fields.

We will limit the derivation of the wave equation to linear, isotropic media, which is the most important case. For more general cases, a similar procedure can be adopted.

#### 2.9.2 Time-varying electromagnetic fields

The first two of Maxwell’s equations (Table 2.1) in differential form, as given by (I) and (II), are first-order, coupled differential equations; that is, both the unknown fields \( \mathcal{E} \) and \( \mathcal{H} \) appear in each equation. Usually it is very desirable,
for convenience in solving for $E$ and $H$, to uncouple these equations. This can be achieved at the expense of increasing the order of the differential equations to second order.

Applying the rotational to the first and second equation we get

$$\text{rot rot } H = \text{rot } J + \frac{\partial D}{\partial t},$$  
(2.57a)

$$\text{rot rot } E = - \frac{\partial B}{\partial t}.$$  
(2.57b)

These equations together with the constitutive relations of linear, isotropic media yield two coupled equations that only contain $E$ and $H$, respectively,

$$\text{rot rot } H = \text{rot } J + \frac{\partial}{\partial t} (\varepsilon \text{ rot } E),$$  
(2.58a)

$$\text{rot rot } E = - \frac{\partial}{\partial t} (\mu \text{ rot } H).$$  
(2.58b)

Now we use again equations (I) and (II) of Maxwell's equations. (II) is used to replace $\text{rot } E$ in the first equation above and (I) for $\text{rot } H$ in the second one. This way, we obtain two decoupled partial differential equations: one for $E$ and a second one for $H$

$$\text{rot rot } H = \text{rot } J - \varepsilon \mu \frac{\partial^2 H}{\partial t^2},$$  
(2.59a)

$$\text{rot rot } E = - \mu \frac{\partial J}{\partial t} - \varepsilon \mu \frac{\partial^2 E}{\partial t^2}.$$  
(2.59b)

Rearranging the terms results in

$$\text{rot rot } H + \frac{1}{c^2} \frac{\partial^2 H}{\partial t^2} = \text{rot } J,$$  
(2.60a)

$$\text{rot rot } E + \frac{1}{c^2} \frac{\partial^2 E}{\partial t^2} = - \mu \frac{\partial J}{\partial t},$$  
(2.60b)

which is the 3D wave equation. The velocity of propagation $v$ is equal to $1/\sqrt{\varepsilon \mu}$. In an isotropic medium $v$ is equal to the velocity of light $c$. In free space $c = c_0 = 299792458$ m/s.

Until now, we did not consider the current $J$. As outlined in Section 2.2 $J$ is in general composed of an impressed current $J_i$ and another, induced part. In lossy isotropic media it is Ohm's law that relates electric field $E$ and current $J_{\text{Ohm}} = \sigma E$. We can insert this separation into the above wave equation.
The Maxwell equations

\[ \text{rot rot } \mathcal{H} + \frac{1}{c^2} \frac{\partial^2 \mathcal{H}}{\partial t^2} + \sigma \mu \frac{\partial \mathcal{H}}{\partial t} = \text{rot } \mathcal{J}, \]  
\[ \text{(2.61a)} \]

\[ \text{rot rot } \mathcal{E} + \frac{1}{c^2} \frac{\partial^2 \mathcal{E}}{\partial t^2} + \sigma \mu \frac{\partial \mathcal{E}}{\partial t} = -\mu \frac{\partial \mathcal{J}}{\partial t}. \]
\[ \text{(2.61b)} \]

Often one finds in textbooks only the form (2.60) of the wave equation. From the context must then be deduced if \( \mathcal{J} \) is only the impressed current \( \mathcal{J}_i \) or if it also includes the conduction term \( \sigma \mathcal{E} \). In the lossless case (2.61) becomes (2.60).

We will see later that the wave equation will take on a very simple form in the frequency domain that is the same for the lossless case (\( \sigma = 0 \)) and the lossy case (\( \sigma \neq 0 \)).

Now that we have derived the wave equation (2.60) respectively (2.61) we can ask why it is called wave equation. To illustrate the nature of the solutions of (2.60), we consider the case of a 1D problem. This is, \( \mathcal{E} \) and \( \mathcal{H} \) only depend on a single spatial coordinate, let’s call it \( z \). To further simplify the problem we will also assume that both \( \mathcal{E} \) and \( \mathcal{H} \) only have a single component, let’s say \( E_x \) respectively \( H_y \) (e. g. plane wave). In this case, the 3D vector wave equation reduces to a scalar ordinary differential equation of second order with constant coefficients

\[ \frac{d^2 E_x}{dz^2} - \frac{1}{c^2} \frac{d^2 E_x}{dt^2} = 0, \]
\[ \text{(2.62a)} \]

\[ \frac{d^2 H_y}{dz^2} - \frac{1}{c^2} \frac{d^2 H_y}{dt^2} = 0. \]
\[ \text{(2.62b)} \]

We have assumed here that we deal with a lossless, source-free region, i. e. \( \mathcal{J}_i = 0 \).

Both equations are of the form

\[ \frac{d^2 f}{dz^2} - \frac{1}{c^2} \frac{d^2 f}{dt^2} = 0. \]
\[ \text{(2.63)} \]

Any function of the form \( f(z \pm ct) \) is a solution of this equation as can be verified by simply inserting it into the wave equation.

The solution is illustrated in Figure 2.7 for the negative sign (\( - \)) and clearly shows a signal propagating in the positive \( z \)-direction with velocity \( c \). For the positive sign (\( + \)) we would find a signal travelling in opposite direction, i. e. in negative \( z \)-direction.

### 2.9.3 Wave equation in the frequency domain

For time harmonic signals we find in place of (2.61)
\( f(z) \)

\( t = 0 \)

\( f(z - c t_1) \)

\( t = t_1 \)

\( f(z - c t_2) \)

\( t = t_2 \)

**Figure 2.7** Propagation of a signal \( f(z - ct) \).

\[
\text{rot } \text{rot } \mathbf{H} + \gamma^2 \mathbf{H} = \text{rot } \mathbf{J}, \quad (2.64a)
\]

\[
\text{rot } \text{rot } \mathbf{E} + \gamma^2 \mathbf{E} = -j\omega \mu \mathbf{J}, \quad (2.64b)
\]

where

\[
\gamma^2 = -\frac{\omega^2}{\epsilon^2}(1 - j\frac{\sigma}{\omega\epsilon}). \quad (2.65)
\]

(2.64) is referred to as the *Helmholtz* equation or reduced wave equation. The constant \( \gamma \) is called the *propagation constant*. It is a complex number for a lossy medium \((\sigma \neq 0)\). This results in energy loss because of heating. The loss of energy is represented by the real part of \( \gamma \) which acts like an attenuation or damping factor. The damping term is proportional to \( \sigma \) and the first time derivative of the magnetic field. In metals, excluding ferro-magnetic materials, the permittivity and permeability are essentially equal to their free space values, at least for frequencies up to and including the microwave range.

It can be shown that a finite conductivity \( \sigma \) is equivalent to an imaginary term in the permittivity \( \epsilon \). In fact we can write

\[
\gamma^2 = -\omega^2 \epsilon \mu (1 - j\frac{\sigma}{\omega\epsilon}) = -\omega^2 \mu (\epsilon - j\frac{\sigma}{\omega}) = -\omega^2 \mu (\epsilon - j\epsilon''). \quad (2.66)
\]

In metals the conduction current \( \sigma \mathbf{E} \) is generally very much larger than the displacement current \( \omega \epsilon_0 \mathbf{E} \), so that the latter may be neglected. For example, \( \sigma \) is equal to \( 5.8 \cdot 10^7 \) S/m for copper, and at a frequency of \( 10 \) GHz, \( \omega \epsilon_0 = 0.56 \) S/m, which is much smaller than \( \sigma \). Only for frequencies in the optical range will the two become comparable.
As already mentioned, Maxwell’s equations is a set of coupled first-order partial differential equations relating electric and magnetic fields. To solve them it is often convenient to introduce potentials, obtaining a smaller number of second-order equations.

For example consider the special case of electrostatics. In this case all derivatives with respect to time vanish, Maxwell’s equations decouple and one finds for the electric field in a linear, isotropic space

\[
\text{div } \mathbf{E} = \frac{\rho}{\varepsilon}.
\]  

(2.67a)

This differential equation only contains derivatives of the first order. For instance in Cartesian coordinates \(\partial/\partial x, \partial/\partial y, \partial/\partial z\).

In electrostatics, the electric field \(\mathbf{E}\) is related to the potential \(\varphi\) by

\[
\mathbf{E} = -\nabla \varphi.
\]  

(2.67b)

and by considering the relation \(\text{div } \nabla = \triangle\) we obtain the second-order differential equation

\[
\triangle \varphi = -\frac{\rho}{\varepsilon}.
\]  

(2.67c)

(2.67c) is simpler than (2.67a) since it contains only a single scalar quantity as unknown \(\varphi\) whereas (2.67a) has three different unknowns: \(E_x, E_y, E_z\) in Cartesian coordinates.

This special case of electrostatics can be generalized and also applied to the dynamic case \((\partial/\partial t \neq 0)\). The price to pay is that a second potential, the so-called vector potential enters “into the game” as we will see.

To start we consider the vector relation

\[
\text{div } \text{rot } \mathbf{A} = 0.
\]  

(2.68)

This relation holds always and means that any rotational field \(\text{rot } \mathbf{A}\) has no divergence, i.e. no sources are present.

Inspecting Maxwell’s equations we can use this relation to define the magnetic flux

\[
\mathbf{B} = \text{rot } \mathbf{A}.
\]  

(2.69)

\(A\) is called the vector potential.

We can proceed similarly with Faraday’s law (II in Table 2.1) by rewriting it as

\[
\text{rot } \left( \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0
\]  

(2.70)

and using the vector relation

\[
\text{rot } \nabla \varphi = 0.
\]  

(2.71)
i. e. the rotation of any lamellar field (\( \text{grad} \phi \)) is zero.

This means that the quantity between parentheses in (2.70) can be expressed using the scalar potential \( \phi \):

\[
\mathcal{E} + \frac{\partial A}{\partial t} = -\text{grad} \phi \quad \text{or} \quad \mathcal{E} = -\text{grad} \phi - \frac{\partial A}{\partial t}.
\]

Inserting (2.69) and (2.72) in Ampère-Maxwell’s law (I) and Gauß’ law (IV) (Table 2.1) and using the constitutive equation \( \mathcal{D} = \varepsilon \mathcal{E} \) we obtain

\[
\Delta \phi + \frac{\partial}{\partial t} (\text{div} A) = -\frac{\rho}{\varepsilon}, \tag{2.73a}
\]

\[
\text{rot rot} A + \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} + \frac{1}{c^2} \frac{\partial}{\partial t} \text{grad} \phi = \mu J. \tag{2.73b}
\]

The complete set of Maxwell’s equations has thus been reduced to two equations. But they are still coupled. To uncouple them we will exploit the arbitrariness involved in the definition of the potentials. Since \( \mathcal{B} \) is defined through (2.69) in terms of \( A \), the vector potential \( A \) is unique up to a gradient field \( \text{grad} \Lambda \) since \( \text{rot grad} \Lambda = 0 \). We can therefore replace \( A \) by

\[
A \rightarrow A' = A + \text{grad} \Lambda. \tag{2.74a}
\]

This implies that \( \phi \) has to be replaced by

\[
\phi \rightarrow \phi' = \phi - \frac{\partial \Lambda}{\partial t} \tag{2.74b}
\]

to leave the field \( \mathcal{B} \) (2.69) and \( \mathcal{E} \) (2.72) invariant.

The freedom implied by (2.74a) and (2.74b) allows us to chose \((A, \phi)\) such that (2.73) takes on special forms that either simplify the equation or that are well-adapted to the considered problem.

Choosing \( \text{div} A \) is called choosing a gauge and changing \( A \) by adding \( \text{grad} \Lambda \) is called a gauge transformation. The invariance of the fields under such transformations is called gauge invariance.

In the following we will present two gauge transformations that are frequently encountered in literature and in solving electromagnetic problems.

A) Coulomb gauge

This gauge is given by

\[
\text{div} A = 0. \tag{2.75}
\]

From (2.73a) follows immediately

\[
\Delta \phi = -\frac{\rho}{\varepsilon}, \tag{2.76}
\]
i. e. the scalar potential \( \phi \) satisfies the Poisson equation. \( \phi \) is the Coulomb potential due to the charge density \( \rho(\vec{r},t) \). \( \phi \) is the instantaneous response to \( \rho \) since \( \triangle \) is independent of \( t \).

By means of the vector relation \( \text{rot rot} = \text{grad div} - \triangle \) the equation for the vector potential (2.73b) is transformed into

\[
\triangle A - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} = -\mu J + \frac{1}{c^2} \frac{\partial (\text{grad} \phi)}{\partial t}.
\]

The scalar potential \( \phi \) appears as additional source term in the differential equation for \( A \). The vector potential satisfies the inhomogeneous wave equation. Opposite to the scalar potential \( \phi \), which propagates with infinite speed, \( A \) propagates with finite speed due to the implied finite propagation velocity \( c \).

The Coulomb gauge is often used when no sources are present, i. e. \( \rho = J = 0 \). In this case \( \text{grad} \phi = \text{const.} \), since \( \triangle \phi = \text{div grad} \phi = 0 \), and therefore \( \partial (\text{grad} \phi)/\partial t = 0 \). Thus, \( A \) satisfies the homogeneous wave equation.

### B) Lorenz gauge

The Lorenz\(^4\) gauge relates the vector potential \( A \) to the scalar potential \( \phi \) by means of

\[
\text{div} A = -\frac{1}{c^2} \frac{\partial \phi}{\partial t}.
\]

This transforms (2.73) into an especially simple form

\[
\triangle \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\varepsilon},
\]

\[
\triangle A - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} = -\mu J.
\]

What an amazing set of equations! They are amazing, first, because they are decoupled and, second, because \( \phi \) and \( A \) satisfy the same type of equation: the inhomogeneous wave equation. Furthermore, \( \phi \) is the response to the electric charge and \( A \) goes with the electric current \( J \). A third important property of this gauge is that the equations are independent of the chosen coordinate system and therefore it fits naturally into the theory of special relativity where the Lorentz transformation of coordinates is an important issue.

Different to the Coulomb gauge also \( \phi \) propagates with finite speed, i. e. with the speed of light.

### 2.9.5 Solution of the wave equation

The differential equation that we have encountered in the case of the Lorenz gauge is called in mathematical literature d’Alembert’s inhomogeneous equation.

\(^4\)On the spelling of Lorenz see Appendix F.
A direct solution can be obtained that is presented here without derivation. In the case of the scalar potential, $\phi$ is given by

$$
\phi(\vec{r}_0, t) = \frac{1}{4\pi\varepsilon} \int_{\mathcal{V}} \rho(\vec{r}, t') \frac{dV}{|\vec{r}_0 - \vec{r}|} \quad (2.81a)
$$

with

$$
t' = t - \frac{|\vec{r}_0 - \vec{r}|}{c}. \quad (2.81b)
$$

$\vec{r}_0$ is the point in the space where we would like to compute the potential, $\vec{r}$ the location of the source. The integration has to be done over the complete volume $\mathcal{V}$ in which the source is located. This can be also the infinite space.

$t'$ is the so-called **retarded time**. It is responsible for the propagation of the potential with finite speed. In other words, a change in the source at the time instant $t$ is recognized at the observation point $\vec{r}_0$ only $\frac{|\vec{r}_0 - \vec{r}|}{c}$ seconds later. This means that the reaction cannot come before the action. The retarded time is the basis of a causal system.

A second solution is possible as we have seen in Section 2.9.2 with

$$
t' = t + \frac{|\vec{r}_0 - \vec{r}|}{c}. \quad (2.81c)
$$

This time is called **advanced time** and represents a non-causal system: the reaction comes before the action.

The solution for the Coulomb gauge is included in the above integral by letting $c \rightarrow +\infty$. In this case we have $t' = t$ thus a change in the source is **instantaneously** recognized at any point $\vec{r}_0$ in the space.

For the vector potential we have a similar equation (since the differential equations for $\phi$ and $A$ are the same). $\rho/\varepsilon$ is replaced by $\mu J$ and we obtain

$$
A(\vec{r}_0, t) = \frac{\mu}{4\pi} \int_{\mathcal{V}} \frac{J(\vec{r}, t')}{|\vec{r}_0 - \vec{r}|} dV. \quad (2.81d)
$$

Remark: Although we have an explicit expression for the scalar and vector potential, the calculation of the integral is often cumbersome or even impossible. Only for special source distributions of $\rho$ or $J$ a direct solution of the integral is possible.

### 2.9.6 Hertz’s electric vector. Polarization potentials

It is sometimes useful to utilize potentials other than the standard scalar and vector potentials $\phi$ and $A$, respectively, as auxiliary field from which the fundamental electromagnetic fields can be derived. The most important of these are the **polarization potentials** or **Hertz’s vectors**. As the name suggests, these potentials put the electric and magnetic polarization densities to the fore.
Let us consider linear, isotropic media with an external electric polarization density $P_{\text{ext}}$, but no separate macroscopic charge or current. The macroscopic electric field is written

$$D = \varepsilon_0 E_0 + P_{\text{ext}}.$$  

(2.82)

With the definitions (2.69) and (2.72) of the fields in terms of scalar and vector potentials, Maxwell’s equations yield the wave equations

$$\triangle \phi - \frac{1}{c_0^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{\text{div } P_{\text{ext}}}{\varepsilon_0},$$

(2.83a)

$$\triangle A - \frac{1}{c_0^2} \frac{\partial^2 A}{\partial t^2} = -\mu_0 \frac{\partial P_{\text{ext}}}{\partial t},$$

(2.83b)

when $A$ and $\phi$ satisfy the Lorenz condition. Comparing with (2.79) we find

$$J = \frac{\partial P_{\text{ext}}}{\partial t} \quad \text{and} \quad \rho = -\text{div } P_{\text{ext}}.$$  

(2.84)

The divergence of $P_{\text{ext}}$ plays the role of the electric charge and the time derivative $\partial P_{\text{ext}}/\partial t$ represents the electric current.

Regarding the right hand sides of the wave equations (2.83) we can introduce a new vector $\Pi_e$ such that $A$ and $\phi$ are written in a similar way

$$A = \mu_0 \frac{\partial \Pi_e}{\partial t} \quad \text{and} \quad \phi = -\frac{\text{div } \Pi_e}{\varepsilon_0}. $$

(2.85)

If we insert this definition into either the first or the second equation of (2.83) we find that both are satisfied and that they reduce to the same single vector equation for $\Pi_e$

$$\triangle \Pi_e - \frac{1}{c_0^2} \frac{\partial^2 \Pi_e}{\partial t^2} = -P_{\text{ext}}.$$  

(2.86)

While it is rather evident to use the second equation of (2.83) to derive the last one, it is a bit more difficult to derive (2.86) from the first one. To do so, we have used the vector relation

$$\triangle \text{div } \Pi_e = \text{div } \text{grad } \text{div } \Pi_e = \text{div } (\triangle \Pi_e + \text{rot rot } \Pi_e) = \text{div } \triangle \Pi_e,$$

(2.87)

since the divergence of a rotational field is always identical zero.

The second point is that the relation (2.86) is only unique up to a rotational field when deriving from (2.83a), again because $\text{div rot } C \equiv 0$. However, we can always use an appropriate gauge transformation to remove the rotational field, similar to the choice of an arbitrary integration constant. For this reason (2.86) derives also from (2.83a).

---

5In linear, isotropic media the electric polarization $P_{\text{ext}}$ and the electric field $E$ are related by the electric susceptibility $\chi_e$, i.e. $P_{\text{ext}} = \varepsilon_0 \chi_e E_{\text{ext}}$ which yields the electric permittivity $\varepsilon_r = 1 + \chi_e$. 

Finally, the electric and magnetic field are given by

\[ E = \frac{1}{\varepsilon_0} \text{grad div } \Pi_e - \mu_0 \frac{\partial^2 \Pi_e}{\partial t^2} = \frac{1}{\varepsilon_0} \text{rot rot } \Pi_e - \frac{P_{\text{ext}}}{\varepsilon_0} \tag{2.88a} \]

\[ H = \text{rot } \frac{\partial \Pi_e}{\partial t} \tag{2.88b} \]

using (2.69) and (2.72).

### 2.9.7 Duality Principle. Symmetric Maxwell equations

Maxwell’s equations possess a certain duality in \( E \) and \( H \), except for the source terms \( J \) and \( \rho \) since no magnetic sources are known to exist in nature. To study this duality we therefore consider first source-free regions. Later we will see that this duality can be extended to regions containing sources by introducing artificial magnetic charges and currents.

#### A) Source-free region

Let’s assume a region free of current \((J = 0)\) and charge \((\rho = 0)\) where we know two field configurations \((E, H)\) and \((E', H')\) each satisfying Maxwell’s equations

\[
\begin{align*}
\text{rot } H &= \varepsilon \frac{\partial E}{\partial t} \\
\text{and} \quad -\text{rot } E' &= \mu \frac{\partial H'}{\partial t} \\
-\text{rot } E &= \mu \frac{\partial H}{\partial t} \\
\text{rot } H' &= \varepsilon \frac{\partial E'}{\partial t}
\end{align*}
\tag{2.89}
\]

and we ask for a mapping from \((E, H)\) to \((E', H')\).

Inspecting the first row of equations we find that \( H \) can be replaced by \(-E'\) and \( \varepsilon E \) by \( \mu H' \). We are interested in the simplest transformation from unprimed fields to primed fields. Since Maxwell’s equations are linear, we will try to find a linear mapping. To do so, we write

\[
E' = -\zeta H \quad \text{and} \quad H' = \chi \frac{\varepsilon}{\mu} E. \tag{2.90}
\]

Inserting these relations into the equations with unprimed quantities in (2.89) we find Maxwell’s equations satisfied for the primed field if and only if

\[ \zeta = \chi. \]

We thus have found the mapping

\[
E' = -\zeta H \quad \text{and} \quad H' = \frac{\zeta}{\mu} E, \tag{2.91}
\]

where \( \zeta \) is an arbitrary number.

We now can ask if there are values of \( \zeta \) that yield a more symmetric transformation between primed and unprimed fields. For this, we first interpret the
previous result geometrically. In Figure 2.8 the fields \((\mathcal{E}, \mathcal{H})\) and \((\mathcal{E}', \mathcal{H}')\) are represented.

The transformation (2.91) turns the field \((\mathcal{E}, \mathcal{H})\) by \(90°\) and scales the axes. In matrix notation this can be written as

\[
\begin{pmatrix}
\mathcal{E}' \\
\mathcal{H}'
\end{pmatrix} = \zeta \begin{pmatrix}
0 & -1 \\
\frac{\varepsilon}{\mu} & 0
\end{pmatrix}
\begin{pmatrix}
\mathcal{E} \\
\mathcal{H}
\end{pmatrix}.
\] (2.92)

If we reapply the same transformation on the primed fields then we have the situation of Figure 2.8(b). We now can demand that \((\mathcal{E}'', \mathcal{H}'') = (-\mathcal{E}, -\mathcal{H})\), thus

\[
\begin{align*}
\mathcal{E}'' &= -\zeta_0^2 \frac{\varepsilon}{\mu} \mathcal{E} = -\mathcal{E}, \\
\mathcal{H}'' &= -\zeta_0^2 \frac{\varepsilon}{\mu} \mathcal{H} = -\mathcal{H},
\end{align*}
\] (2.93)

from which follows that

\[
\zeta_0 = \pm \sqrt{\frac{\mu}{\varepsilon}}.
\] (2.94)

For this value of \(\zeta\) the mapping (2.91) has the form

\[
\begin{align*}
\mathcal{E}' &= \mp \sqrt{\frac{\mu}{\varepsilon}} \mathcal{H} \\
\mathcal{H}' &= \pm \sqrt{\frac{\varepsilon}{\mu}} \mathcal{E}.
\end{align*}
\] (2.95)

The dimension of \(\zeta\) is \(\Omega\) and \(\zeta_0\) has the value \(120\pi \Omega\) for free space, which is the field impedance.
2.9 The wave equation. Electromagnetic potentials

B) Region with sources

We will now extend the duality of the previous section to regions with sources. For this we first insert the mapping (2.91) into Maxwell’s equations with electric current and charge and obtain

\[
\begin{align*}
- \text{rot} E' &= \zeta J + \mu \frac{\partial H'}{\partial t} \quad \text{div} E' = 0 \\
\text{rot} H' &= \varepsilon \frac{\partial E'}{\partial t} \quad \text{div} H' = \frac{\rho_m}{\mu}.
\end{align*}
\]

(2.96)

Formally these relations are Maxwell’s equations for an electromagnetic field \((E', H')\) produced by the magnetic current density \(M' = \zeta J\) and the magnetic charge density \(\rho'_m = \zeta \rho\). These considerations suggest that complete duality is achieved by generalizing Maxwell’s equations as follows:

\[
\begin{align*}
\text{rot} H &= J + \varepsilon \frac{\partial E}{\partial t} \quad \text{div} H = \frac{\rho_m}{\mu} \\
- \text{rot} E &= M + \mu \frac{\partial H}{\partial t} \quad \text{div} E = \frac{\rho}{\varepsilon}.
\end{align*}
\]

(2.97)

It remains to determine how \((M, \rho_m)\) is transformed to \((J', \rho')\). For this, the mapping (2.91) is applied to the symmetric set of Maxwell’s equations and it can be easily verified that under the duality transformation

\[
\begin{align*}
H' &= \zeta \frac{\varepsilon}{\mu} E \quad J' = -\zeta \frac{\varepsilon}{\mu} M \quad \rho' = -\zeta \frac{\varepsilon}{\mu} \rho_m \\
E' &= -\zeta H \quad M' = \zeta J \quad \rho'_m = \zeta \rho
\end{align*}
\]

(2.98)

Maxwell’s equations remain unchanged.

For the special value \(\zeta_0\) we find the transformations

\[
\begin{align*}
H' &= Y E \quad J' = -Y M \quad \rho' = -Y \rho_m \quad Z = \frac{1}{Y} = \pm \sqrt{\frac{\mu}{\varepsilon}}. \\
E' &= -Z H \quad M' = Z J \quad \rho'_m = Z \rho
\end{align*}
\]

(2.99)

### 2.9.8 Hertz’s magnetic vector

The symmetric Maxwell equations (2.97) contain the electric current \(J\) and the magnetic current \(M\). Due to the superposition principle, the total field consists of the field due to \(J\) and the field due to \(M\). The field due to \(J\) was already obtained in Section 2.9.6 in terms of the electric Hertz vector \(\Pi_e\). Similarly, the field due to \(M\) can be computed using the duality transformations (2.98) and by replacing the electric Hertz vector \(\Pi_e\) with the magnetic Hertz vector \(\Pi_m\). The duality transformations for the Hertz vectors are

\[
\Pi'_m = \zeta \Pi_e \quad \text{and} \quad \Pi'_e = -\zeta \frac{\varepsilon_0}{\mu_0} \Pi_m.
\]

(2.100)
Thus we have the following vector equation analogue to (2.86)

$$\Delta \Pi_m - \frac{1}{c_0^2} \frac{\partial^2 \Pi_m}{\partial t^2} = -Q_{\text{ext}} \quad \text{and} \quad \mathcal{M} = \frac{\partial Q_{\text{ext}}}{\partial t}, \rho_m = -\text{div} \ Q_{\text{ext}},$$

(2.101)

where $Q_{\text{ext}}$ is the magnetic polarization vector and represents the magnetic dipole moment per unit volume. The field is

$$\mathcal{H} = \frac{1}{\mu_0} \text{grad} \ \text{div} \ \Pi_m - \varepsilon_0 \frac{\partial^2 \Pi_m}{\partial t^2} = \frac{1}{\mu_0} \text{rot} \ \text{rot} \ \Pi_m - \frac{Q_{\text{ext}}}{\mu_0},$$

(2.102a)

$$\mathcal{E} = -\text{rot} \ \frac{\partial \Pi_m}{\partial t}.$$  

(2.102b)

### 2.10 Summary

The next two pages summarize the derivations of the preceding sections: from Maxwell’s basic equations over the duality principle and the symmetric Maxwell equations, the introduction of potentials as solutions to Maxwell’s equations and gauge transformations to the Hertz Vector in case that we apply the Lorenz gauge.

---

*We defined $Q_{\text{ext}}$ such that $B = \mu_0 H + Q_{\text{ext}}$. Thus, $Q_{\text{ext}}$ represents a magnetic flux. In the literature one encounters frequently the definition $Q_{\text{ext}} = B/\mu_0 - \mathcal{H}$ which is the magnetization of the medium and represented by the symbol $\mathcal{M}$. Since we have defined $\mathcal{M}$ to be the magnetic current density and to preserve the symmetry in Maxwell’s equations we have chosen $Q_{\text{ext}}$ to be the equivalent to $P_{\text{ext}}$ and it is therefore called magnetic polarization vector.*
The basic Maxwell Equations
\[ \text{rot } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad \text{rot } \mathbf{E} = \frac{\partial \mathbf{B}}{\partial t}, \]
\[ \text{div } \mathbf{B} = 0, \quad \text{div } \mathbf{D} = \rho \]

Duality Principle (Sect. 2.9.7)

The symmetric Maxwell Equations
\[ \text{rot } \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad \text{rot } \mathbf{E} = \mathbf{M} + \frac{\partial \mathbf{B}}{\partial t}, \]
\[ \text{div } \mathbf{B} = \rho_m, \quad \text{div } \mathbf{D} = \rho \]

Solution of Maxwell's Equations (Sect 2.9.4)

**Potentials**

<table>
<thead>
<tr>
<th>electric sources</th>
<th>magnetic sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (\mathbf{A}, \phi) )</td>
<td>( (\mathbf{F}, \psi) )</td>
</tr>
<tr>
<td>( \mathbf{E} = -\text{grad } \phi - \frac{\partial \mathbf{A}}{\partial t} )</td>
<td>( \mathbf{H} = -\text{grad } \psi - \frac{\partial \mathbf{F}}{\partial t} )</td>
</tr>
<tr>
<td>( \mathbf{B} = \text{rot } \mathbf{A} )</td>
<td>( \mathbf{D} = -\text{rot } \mathbf{F} )</td>
</tr>
</tbody>
</table>

**Gauge Transformations** (Sect 2.9.4)

- Lorenz gauge
- Coulomb gauge

**Lorenz gauge**

1. **electric sources** \( (\mathbf{M}, \rho_m) = (0, 0) \)
   - \( \text{div } \mathbf{A} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0 \)
   - \( \Delta \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\varepsilon} \)
   - \( \Delta \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu \mathbf{J} \)

2. **magnetic sources** \( (\mathbf{J}, \rho) = (0, 0) \)
   - \( \text{div } \mathbf{F} + \frac{1}{c^2} \frac{\partial \psi}{\partial t} = 0 \)
   - \( \Delta \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = -\frac{\rho_m}{\mu} \)
   - \( \Delta \mathbf{F} - \frac{1}{c^2} \frac{\partial^2 \mathbf{F}}{\partial t^2} = -\varepsilon \mathbf{M} \)

**Coulomb gauge**

1. **electric sources** \( (\mathbf{M}, \rho_m) = (0, 0) \)
   - \( \text{div } \mathbf{A} = 0 \)
   - \( \Delta \phi = \frac{\rho}{\varepsilon} \)
   - \( \Delta \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu \mathbf{J} + \frac{1}{c^2} \text{grad } \frac{\partial \phi}{\partial t} \)

2. **magnetic sources** \( (\mathbf{J}, \rho) = (0, 0) \)
   - \( \text{div } \mathbf{F} = 0 \)
   - \( \Delta \psi = \frac{\rho_m}{\mu} \)
   - \( \Delta \mathbf{F} - \frac{1}{c^2} \frac{\partial^2 \mathbf{F}}{\partial t^2} = -\varepsilon \mathbf{M} + \frac{1}{c^2} \text{grad } \frac{\partial \psi}{\partial t} \)
The Maxwell equations

Lorenz Gauge

\[ A = \mu_0 \frac{\partial \Pi_e}{\partial t} \]
\[ \phi = -\frac{\text{div } \Pi_e}{\varepsilon_0} \]
\[ \triangle \Pi_e - \frac{1}{c_0^2} \frac{\partial^2 \Pi_e}{\partial t^2} = -P_{\text{ext}} \]
\[ J = \frac{\partial P_{\text{ext}}}{\partial t} \]
\[ \rho = -\text{div } P_{\text{ext}} \]

Hertz's Vector

\[ F = \varepsilon_0 \frac{\partial \Pi_m}{\partial t} \]
\[ \psi = -\frac{\text{div } \Pi_m}{\mu_0} \]
\[ \triangle \Pi_m - \frac{1}{c_0^2} \frac{\partial^2 \Pi_m}{\partial t^2} = -Q_{\text{ext}} \]
\[ M = \frac{\partial Q_{\text{ext}}}{\partial t} \]
\[ \rho_m = -\text{div } Q_{\text{ext}} \]

Electromagnetic Field

\[ E_e = \frac{1}{\varepsilon_0} \text{grad div } \Pi_e - \mu_0 \frac{\partial^2 \Pi_e}{\partial t^2} \]
\[ = \frac{1}{\varepsilon_0} \text{rot rot } \Pi_e - \frac{P_{\text{ext}}}{\varepsilon_0} \]
\[ H_e = \text{rot } \frac{\partial \Pi_e}{\partial t} \]

\[ E_m = \frac{1}{\mu_0} \text{grad div } \Pi_m - \varepsilon_0 \frac{\partial^2 \Pi_m}{\partial t^2} \]
\[ = \frac{1}{\mu_0} \text{rot rot } \Pi_m - \frac{Q_{\text{ext}}}{\mu_0} \]
\[ H_m = -\text{rot } \frac{\partial \Pi_m}{\partial t} \]

Superposition Principle

\[ \mathcal{E} = E_e + E_m \]
\[ \mathcal{H} = H_e + H_m \]
CHAPTER 3

Velocities of propagation

In this chapter we will study different velocities that are associated with the propagation of electromagnetic waves. The properties of the medium through which the wave travels are characterized by the propagation constant and its frequency dependency, i.e. by the dispersion relation.

One of the most important definitions refers to the group velocity which we will study in detail. It seems to have been first discovered by Lord Rayleigh, who characterized this velocity in sound waves. It is now known to apply to practically all kinds of waves. Let us use the vocabulary of radio engineers and consider a carrier wave with a modulation. The phase velocity yields the motion of the elementary wavelets in the carrier, while the group velocity gives the propagation of the modulation. Lord Rayleigh considered that the group velocity corresponds to the velocity of energy or signals.

This however raised difficulties with the theory of relativity which states that no velocity can be higher than c, the velocity of light in vacuum. Group velocity, as originally defined, became larger than c or even negative within an absorption band. Such a contradiction had to be resolved and was extensively discussed in publications about 1910. Sommerfeld stated the problem correctly and proved that no signal velocity could exceed c.

We will see that apart from the group velocity we have to introduce the terms velocity of energy flow and signal velocity. These three velocities are identical for nonabsorbing (non-dissipative or lossless) media, but they differ considerably in dissipative media.

3.1 Phase velocity

A quantity that can be directly deduced from the phase constant $\beta$ (the imaginary part of the propagation constant $\gamma$) is the phase velocity $v_{ph}$. It is defined as the ratio $\omega/\beta$.

For a plane wave in unbounded space, $\beta$ is a linear function of the frequency $f$ and therefore the phase velocity is constant and does not change with frequency. In bounded space, e.g. a waveguide, waves can be dispersive, that is their phase constant is not a linear function of $f$. For this reason, their phase velocity is not constant over frequency. Moreover, $v_{ph} > c$, i.e. the phase velocity of a dispersive wave is larger than the velocity of light in a corresponding unbounded medium. Therefore they are called fast waves. We will see this in more detail in the Chapter 4.
3.2 Group velocity

The most simple way to transmit some information is to switch on and off the source, hence some sort of binary signal. Such a signal cannot be represented by a monochromatic wave but only by a (infinite) number of waves of different frequencies. Even in the most sharply tuned radio transmitter or most monochromatic light source, waves with finite frequency ranges are generated and transmitted. Furthermore, any causal signal must consist of “wave trains” of finite extent or waves with finite frequency spectra. Since Maxwell’s equations are linear, any time-dependent process can be treated by a superposition of sinusoidal waves of different frequencies respectively wave numbers.

In Chapter 4 we will see that in a waveguide the phase velocity of a wave depends on its frequency. Consequently, different frequency components of a wave travel with different speeds and tend to change phase with respect to one another during the propagation and give rise to phase distortion of the waveform. If we deal with a signal\(^1\) we can use the superposition principle and decompose it into a linear combination of monochromatic waves each propagating with its own phase velocity. For such a decomposed signal we can ask at what time the signal will arrive at the other end of e. g. a waveguide. Thus we come to the question of the signal velocity.

3.2.1 Wave train of two monochromatic waves

To answer the question of signal velocity let us consider the case of two monochromatic waves of frequencies \(f_1\) and \(f_2\). A schematic sketch of the spectrum is shown in Figure 3.1(a). The two waves are given by

\[
\begin{align*}
    w_1(u, v, z, t; \omega_1) &= A_1(u, v) e^{j(\omega_1 t - \beta(\omega_1)z)}, \\
    w_2(u, v, z, t; \omega_2) &= A_2(u, v) e^{j(\omega_2 t - \beta(\omega_2)z)}.
\end{align*}
\]

(3.1a)

(3.1b)

\(A_1\) and \(A_2\) represent the amplitudes of the monochromatic waves and stand for any of the electric and magnetic field components of the waves. The total wave is given by virtue of superposition as

\[
w(u, v, z, t; \omega_1, \omega_2) = w_1(u, v, z, t; \omega_1) + w_2(u, v, z, t; \omega_2).
\]

(3.1c)

For our further considerations it does not matter if the waves have different amplitudes or not. We will therefore assume \(A_1 = A_2 = A\) which simplifies the following derivations.

We can rewrite this equation in a more convenient way by introducing the frequencies and wave numbers

\(^1\)The word signal implies that we are transferring information thus we deal with a superposition of monochromatic waves since a single monochromatic wave cannot transmit any information.
Chapter 3

3.2 Group velocity

(a) The frequency spectrum of two monochromatic waves.

(b) A sketch of the signal for which we have used $\omega_m = 20 \cdot \omega_0$.

Figure 3.1 The frequency spectrum of two monochromatic waves.

\[
\omega_m = \frac{\omega_2 + \omega_1}{2} \quad \text{and} \quad \omega_0 = \frac{\omega_2 - \omega_1}{2},
\]

(3.2)

\[
\beta_m = \frac{\beta(\omega_2) + \beta(\omega_1)}{2} \quad \text{and} \quad \beta_0 = \frac{\beta(\omega_2) - \beta(\omega_1)}{2}.
\]

(3.3)

$\omega_m$ represents the mean value of the two frequencies and $\pm \omega_0$ can be interpreted as the modulation of $\omega_m$ to generate the two monochromatic waves. In a similar manner, $\beta_m$ is the mean value of the two phase constants $\beta(\omega_1)$ and $\beta(\omega_2)$. We thus have

\[
w(u, v, z, t; \omega_1, \omega_2) = Ae^{i(\omega_m t - \beta_m z)}\left(e^{i(\omega_0 t - \beta_0 z)} + e^{-i(\omega_0 t - \beta_0 z)}\right).
\]

(3.4)

The real part of such a signal is sketched in Figure 3.1(b).

So far we did not change anything with respect to the original expression. However, the last expression allows us to interpret the superposition of the two monochromatic waves in a different way: We seem to have a wave (carrier) which travels with the mean frequency $\omega_m$ and the phase constant $\beta_m$, i. e. the mean value of the two propagation constants. What is new is that its strength is varying with a form that depends on the difference frequency $\omega_0$ and the difference wave number $\beta_0$.

From $\omega_0$ and $\beta_0$ we can find the velocity of the modulation. The modulating term is

\[
2 \cos(\omega_0 t - \beta_0 z)
\]

(3.5)

which is of the general form

\[
f(\omega_0 t - \beta_0 z).
\]
But the propagation velocity of such a signal is

\[ v_0 = \frac{\omega_0}{\beta_0} = \frac{\omega_2 - \omega_1}{\beta(\omega_2) - \beta(\omega_1)} . \]  

(3.6)

Hence, the modulation travels with the speed \( v_0 \).

Now we make an important assumption that we have to keep in mind for all further interpretations. We assume that the two frequencies \( f_1 \) and \( f_2 \) are not too “far away” from each other. Mathematically, this means that we can approximate \( \beta(\omega_m \pm \omega_0) \) by

\[ \beta(\omega_m \pm \omega_0) \approx \beta(\omega_m) \pm \frac{\partial \beta(\omega)}{\partial \omega} \, |_{\omega=\omega_m} \cdot \omega_0 . \]  

(3.7)

Comparing this with (3.6) shows that

\[ v_0 \approx \frac{1}{\frac{\partial \beta}{\partial \omega} |_{\omega=\omega_m}} . \]  

(3.8)

The smaller the difference between \( f_1 \) and \( f_2 \) the better the approximation in the last equation. In a strict sense, the inverse derivative of \( \beta \) with respect to \( \omega \) and evaluated at \( \omega_1 \) or \( \omega_2 \) or at the mean value of both represents the speed of the modulation for a infinitely small spectrum, i.e. \( \omega_0 \rightarrow 0 \).

The inverse derivative of \( \beta \) is given the name group velocity since it represents the speed of propagation of a packet (group) of monochromatic waves:

\[ v_{gr} = \frac{1}{\partial \beta / \partial \omega} . \]  

(3.9)

As mentioned above, this normally is only true for an infinitely small spectrum \( (\omega_0 \rightarrow 0) \). In other words, for the slowest modulation there is a definite speed at which it travels and which is not the same as the phase velocity of the waves.

If we deal with a narrow but finite spectrum, i.e. \( \omega_0 \neq 0 \) but small, then \( v_{gr} \) can still be interpreted as the velocity of the modulation but we are free to choose the frequency at which we evaluate the derivative. Of course, for a signal of finite spectrum from \( \omega_1 \) to \( \omega_2 \) it does not make sense to evaluate \( v_{gr} \) at a frequency outside of the range \([\omega_1, \omega_2]\). But we are free to take any frequency within this range. From this it also becomes clear that \( v_{gr} \) is only approximately the speed of the modulation and this approximation is the more coarse the larger the difference between \( \omega_1 \) and \( \omega_2 \), that is the bigger the modulating frequency \( \omega_0 \).

In practical applications we deal with modulated signals: a lowpass signal is used to modulate a carrier. Let’s say the lowpass signal has a spectrum \([0, \omega_0]\) and the carrier the frequency \( \omega_m \). If we use the technique of amplitude modulation (AM), then the bandpass signal has the spectral range \([\omega_m - \omega_0, \omega_m + \omega_0]\). To find the group velocity of this bandpass signal it makes sense to use as reference the frequency of the carrier, i.e. \( \omega_m \). However, we can chose any other frequency within this range as well.
3.2.2 Wave train with continuous spectrum

The study of the last paragraph can be generalized to wave packets or wave trains of finite length but arbitrary spectrum. An example is shown in Figure 3.2. The appropriate tool for the analysis of such signals is the Fourier integral.

Any component of the electric field can be represented by the following Fourier integral:

\[ w(z,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} A(\omega) e^{j(\omega t - \beta(\omega)z)} d\omega, \quad (3.10) \]

where \( A(\omega) \) is the amplitude of the monochromatic wave with frequency \( \omega \), which describes the properties of the linear superposition of the waves with different frequencies. The sum in (3.1c) is replaced by an integral over all frequencies. \( A(\omega) \) is given by the transform of the function \( w(z,t) \) evaluated at \( z = 0 \):

\[ A(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} w(0,t) e^{-j\omega t} dt. \quad (3.11) \]

In other words \( A(\omega) \) represents the amplitude at the beginning of the wave’s journey.

Before continuing, let’s look what happens with this integral if we consider a monochromatic wave of frequency \( f_0 \). The amplitude function for this kind of wave is given by Dirac’s distribution, i.e. \( A(\omega) = \sqrt{2\pi} A_0 \cdot \delta(\omega - \omega_0) \). Evaluating the above integral yields

\[ w(z,t) = A_0 e^{j(\omega_0 t - \beta(\omega_0)z)}, \]

which is the well-known propagation of a monochromatic wave.

To further proceed with the Fourier integral, we have to make some assumptions about the amplitude function \( A(\omega) \). When we considered the propagation
of two monochromatic waves, we assumed that the difference between its frequencies $f_1$ and $f_2$ is not too big which allowed us to introduce the notion of group velocity.

This assumption is now replaced by a narrow band spectrum. That is, the amplitude function $A(\omega)$ is confined around a frequency $\omega_m$. If $A(\omega)$ is of negligible magnitude outside the region $[\omega_m - \delta \omega, \omega_m + \delta \omega]$, we may replace the Fourier integral by

$$w(z, t) = \frac{1}{\sqrt{2\pi}} \int_{\omega_m - \delta \omega}^{\omega_m + \delta \omega} A(\omega) e^{j\omega t - \beta(\omega)z} d\omega,$$

which represents what is commonly called a wave packet. For the present we will assume that $\beta(\omega)$ is real that is the medium is lossless.

If $2\delta \omega$ is “sufficiently” small, we are allowed to replace $\beta(\omega)$ by its Taylor expansion around $\omega_m$ and truncate it after the linear term:

$$\beta(\omega) \approx \beta(\omega_m) + \frac{\partial \beta}{\partial \omega} \bigg|_{\omega_m} (\omega - \omega_m).$$

Inserting this expression into the Fourier integral over the interval $[\omega_m - \delta \omega, \omega_m + \delta \omega]$ and applying a change of variable and integration limits we finally obtain

$$w(z, t) = e^{j\omega_m t - \beta_m z} \cdot w(0, t - \frac{z}{v_{gr}}),$$

where we have introduced for convenience the constants $\beta_m = \beta(\omega_m)$ and $\frac{1}{v_{gr}} = \frac{\partial \beta}{\partial \omega} \bigg|_{\omega_m}$.

From this it is clear that also the wave packet is propagated with the group velocity as introduced for two monochromatic waves. We also recover the phase term which oscillates with $\omega_m$, which can be interpreted as the carrier frequency, and propagates with the phase velocity $\omega_m/\beta_m$.

It is apparent from the manner in which the group velocity was defined that this concept is wholly precise only when the wave packet is composed of elementary (monochromatic) waves lying within an infinitely narrow region of the spectrum, i.e. $\delta \omega \to 0$. As the interval $\delta \omega$ is increased, the spread in phase velocity of the harmonic components in a dispersive medium becomes more marked: the larger $\delta \omega$ the bigger the difference between the phase velocities $v_{ph}(\omega_m - \delta \omega)$ and $v_{ph}(\omega_m + \delta \omega)$. Thus the time delay between the phase fronts of each frequency component becomes bigger and bigger the longer the distance the wave packet
Chapter 3 3.2 Group velocity

travels. The packet is deformed rapidly, and the group velocity as a velocity of the whole packet loses its physical significance.

If \( \delta \omega \) becomes larger, we have to use in the Taylor expansion of \( \beta(\omega) \) also higher order terms. In order to be in the position to still evaluate the integral, special assumptions on the form of the amplitude function must be made. For example, we can take into account the quadratic term in the expansion, i.e.

\[
\beta(\omega) \approx \beta(\omega_m) + \frac{\partial \beta}{\partial \omega} \bigg|_{\omega_m} (\omega - \omega_m) + \frac{1}{2} \frac{\partial^2 \beta}{\partial \omega^2} \bigg|_{\omega_m} (\omega - \omega_m)^2.
\]

Let’s assume we deal with a Gaussian impulse for the amplitude function

\[
w(0, t) = e^{j \omega_m t} e^{-\frac{t^2}{2 \tau^2}} \quad \text{and} \quad A(\omega) = \sqrt{2 \pi} \tau e^{-\frac{(\omega - \omega_m)^2}{2 \tau^2}}.
\]

In a strict sense, the spectrum of the Gaussian impulse is not finite. A characteristic measure for the decay of the spectrum is \( 1/\tau \). After a certain multiple of \( 1/\tau \) we can assume the spectrum to be zero. To continue with the integral, we have to use the original Fourier integral for \(( -\infty, +\infty )\). Thus, after rearranging the terms we have

\[
w(z, t) = e^{j(\omega_m t - \beta_m z)} \frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{+\infty} \sqrt{2 \pi} \tau e^{-\frac{1}{2} (\tau^2 + j \beta_m^\prime z) \omega^2} e^{j[\omega(t - z/v_{gr})]} d\omega
\]

where

\[
\beta_m^\prime = \frac{\partial \beta}{\partial \omega} \bigg|_{\omega_m}.
\]

To ease notation we introduce a new characteristic measure

\[
\tau' = \tau^2 + j \beta_m^\prime z
\]

and the retarded time

\[
t' = t - \frac{z}{v_{gr}}.
\]

We thus find the shape of the Gaussian impulse after having travelled a distance \( z \) in the dispersive medium to be

\[
w(z, t) = \frac{\tau}{\tau'} e^{-\frac{\tau^2}{2 \tau'^2}} e^{j(\omega_m t - \beta_m z)}
\]

which is clearly different to its initial shape.

Again, we see that the phase term travels with the phase velocity and goes with the time \( t \), whereas the modulation or envelope is retarded (\( t' \)) and propagates with the group velocity.

However, different to the previous analysis, we learn something new: the new characteristic measure \( \tau' \) is not constant like its counterpart \( \tau \) but a function of
We can rewrite it in a slightly different way to better reveal its relation with the original $\tau$:

$$\tau'^2 = \tau^2 \left( 1 + \frac{\beta'' m z}{\tau'^2} \right) = \tau^2 \sqrt{1 + \left( \frac{\beta'' m z}{\tau'^2} \right)^2} e^{i \phi},$$

where $\phi$ is the phase of $\tau'$.

The important result is that the magnitude of $\tau'$ increases monotonically with the travelled distance $z$. This has two effects: As time goes by and the impulse propagates in the dispersive medium

1. its width becomes larger. After the distance $z$, its new characteristic length is

$$\tau \sqrt{1 + \left( \frac{\beta'' m z}{\tau'^2} \right)^2},$$

2. its amplitude is continuously diminished. After a distance $z$, the amplitude is

$$\frac{1}{\sqrt{1 + \left( \frac{\beta'' m z}{\tau'^2} \right)^2}},$$

which is just the factor by which the characteristic length is increased.

The expression under the root-sign is proportional to $1/\tau'^2$. This means, the sharper the impulse was at the beginning of its journey, i.e. the smaller $\tau$, the faster it is widened and the stronger its amplitude is diminished. The broadening of the impulse is quantified by

$$\text{broadening of the pulse} \approx \frac{\beta'' m z}{\tau'^2},$$

for an initial impulse width of $2\tau$.

There are equivalent expressions for the group velocity that are occasionally more convenient. One of them is

$$1 \frac{v_{\text{gr}}}{v_{\text{ph}}} = \frac{\partial \beta}{\partial \omega} = 1 \frac{v_{\text{ph}}}{v_{\text{ph}}} + \omega \frac{\partial}{\partial \omega} \left( \frac{1}{v_{\text{ph}}} \right)$$

using $v_{\text{ph}} = \omega/\beta$.

From this we can see that $v_{\text{gr}} = v_{\text{ph}}$ if $v_{\text{ph}}$ is constant. This is the case for non-dispersive media like for example vacuum. On the other hand, if the derivative of $1/v_{\text{ph}}$ with respect to frequency is positive then the group velocity is always smaller than the phase velocity. This is the case of normal dispersion.

Let us consider a specific example: In Chapter 4 we will see that for uniform waveguides

$$\beta(\omega) = \sqrt{\frac{\omega^2}{c^2} - k_c^2}$$

(3.16)
3.2 Group velocity

Figure 3.3 Normalized group velocity (solid line) and normalized phase velocity (dashed line) for a wave with $k_c > 0$ versus normalized frequency. The frequency has been normalized to $f_c = k_c/(2\pi \cdot c)$. For $f/f_c \rightarrow \infty$ the phase and group velocity converge towards the speed of a wave in an unbounded medium.

holds, where $k_c$ is a real-valued constant.

In Figure 3.3 we have plotted the phase and group velocity and we see that the phase velocity is always greater than the group velocity. From Figure 3.3 we learn also that for large frequencies both velocities approach the velocity of a wave in an unbounded medium, i.e. the velocity of light.

If the dispersion is normal and moderate, that is $\beta$ does not vary too much with frequency, the group velocity is equal or at least approximately the velocity of energy since the energy of a modulated signal is presumed to be localized in the envelope of the modulated carrier. The group velocity is also interpreted as the signal velocity, since the information is located in the envelope of the modulated carrier.

However, there are cases where the group velocity loses its meaning as energy and signal velocity and one has to distinguish between them. In a normal dispersive medium an increase in frequency results in a decrease of phase velocity. Under these circumstances the group velocity is always less than the phase velocity. If, on the contrary, the dispersion is anomalous – as is the case in conducting media – the derivative $1/\partial\omega(1/v_{ph})$ is negative and the group velocity is greater than the phase velocity. There is in fact nothing wrong with that. The group velocity simply loses its meaning as energy velocity. Remember how we found and defined the group velocity: we in fact never used the term energy. We were only looking at a (modulated) signal. Its interpretation as a velocity that represents not only the speed of the modulation but also the energy velocity came only at the very end, where we associated the signal energy with
46  Velocities of propagation

the envelope of the modulated signal.

The same is true for its interpretation as signal velocity. The definition of
signal velocity is even more delicate than the definition of energy velocity.

3.3 Velocity of energy flow

As mentioned in the last paragraph, in the case of anomalous dispersion the
group velocity does not represent anymore the velocity of energy. There is in
fact no lack of examples to show that \( v_{gr} \) may exceed the velocity of light. Since
historically it was generally believed and accepted that the group velocity was
necessarily equivalent to the velocity of energy propagation, examples of this sort
were proposed in the first years following Einstein’s publication of the special
theory of relativity as definite contradictions to the postulate that a signal can
never be transmitted with a velocity greater than \( c \). It was Sommerfeld who
showed in a general manner that a signal cannot travel faster than light.

In order to derive the velocity of energy flow, we consider the energy stored in
the electromagnetic field that is transmitted. Let’s assume the electromagnetic
field has an average energy density \( \bar{w} \). In order to transfer the small portion of
energy \( \delta W = \bar{w} \delta V \) stored in the volume \( \delta V \) within the time period \( T \) we need an
average power of

\[
\bar{P} = \frac{\delta W}{T} = \frac{\bar{w} \delta V}{T}.
\]

Now suppose this small portion of energy must flow through the area \( A \). We
can imagine \( A \) as the cross section of a waveguide. Then the average power flow
density is given by

\[
\bar{S} = \frac{\bar{P}}{A} = \frac{\bar{w} \delta V/A}{T}.
\]

The quantity \( (\delta V/A)/T \) represents a velocity. It is the velocity with which we
must move the volume \( \delta V \) in order to transfer it through the area \( A \) within the
time \( T \). Since transferring \( \delta V \) means transferring the energy stored in it, this
quantity is the velocity of energy flow.

This formula allows us to express the velocity of energy propagation by basic
field quantities

\[
v_{energy} = \frac{\bar{S}}{\bar{w}},
\]  

(3.17)

where \( \bar{S} \) is the average power flow density, thus the average Poynting-vector
(average with respect to time and space), and \( \bar{w} \) the average (with respect to
space) energy density of the electromagnetic field which can be calculated using
(2.17).

It can be shown that in the case of normal dispersion this definition of energy
flow velocity leads the same result as the group velocity. But it will remain also
valid in the case of anomalous dispersion, e. g. in dissipative media.
3.4 Signal velocity

As in the case of energy velocity, the group velocity loses its meaning as signal velocity if we deal with the phenomenon of anomalous dispersion. The problem was first investigated by Sommerfeld and Brillouin. Whereas the definition of energy flow velocity is rather simple, the definition of a signal velocity is of much more delicate nature.

The particular problem that Sommerfeld considered was a signal at \( z = 0 \) defined by the function

\[
w(0, t) = \begin{cases} 
0, & \text{when } t < 0, \\
e^{-j\omega t}, & \text{when } t > 0.
\end{cases}
\]

The representation of the field at any point \( z \) can be represented by the Laplace integral

\[
w(z, t) = \frac{1}{j2\pi} \int_{\gamma-j\infty}^{\gamma+j\infty} e^{s(t-\beta(s)z)} \left( \frac{1}{s + j\omega} \right) ds, \quad \gamma > 0.
\]

By functional analytic techniques Sommerfeld could show that at a point \( z \) within the medium (of normal or anomalous dispersion) the field is zero as long as the time \( t < z/c \) and hence that the velocity of the wave front cannot exceed the constant \( c \).

We will not enter into the details of the theory and rather summarize the outcome. According to Brillouin, a signal is a disturbance in the form of a train of oscillations starting at a certain instant. In the course of propagation in a dispersive medium the signal is deformed. It was found that after penetrating to a certain depth into the medium the main body of the signal is preceded by a so-called forerunner or precursor with the velocity \( c \) in all media (see Figure 3.4). The first forerunner arrives with small period and zero amplitude, and then grows slowly both in period and in amplitude. The amplitude then decreases. Then a second forerunner arrives with a velocity smaller than \( c \). The period of the second forerunner is at first very large and then decreases, while the amplitude rises and then falls in a manner similar to that of the first forerunner. These two forerunners can partly overlap and their amplitudes are very small but increase rapidly as their periods approach that of the signal. With a sudden rise of amplitude the principle part of the disturbance arrives, travelling with a velocity \( v_{\text{signal}} \), which Brillouin defines as the signal velocity. An explicit and simple expression of \( v_{\text{signal}} \) cannot be given and its definition is associated somewhat arbitrarily with the method employed to evaluate the above Laplace integral.

Physically the meaning is quite clear: When we want to measure the signal velocity we need a detector which measures the disturbances associated with the signal. Since the first disturbances are infinitely small (they start from zero amplitude) we will need a detector of infinite sensitivity. In this case we would measure the arrival of the first forerunner, which travels with the velocity \( c \). In
practice, however, we will have only detectors of finite sensitivity. They will only detect an arriving signal if its amplitude is large enough. Thus for a detector with normal sensitivity we should measure a speed which is approximately $v_{signal}$, since its amplitude is large enough to be detected.

### 3.5 Concluding remarks

1. The phase velocity has no physical importance and is associated with steady states and may be either greater or less than c. Steady states means signals that last since ever and forever, i. e. a signal which exists for $-\infty < t < +\infty$.

2. The group velocity differs from the phase velocity only in dispersive media. If the dispersion is normal, the group velocity is less than the phase velocity; it is greater than the phase velocity when the dispersion is anomalous. In the neighbourhood of a resonance of the medium (e. g. an absorption line in a plasma) it may become infinite or even negative. Negative means that modulation and carrier travel in opposite directions.

3. The energy flow velocity coincides with the group velocity in regions of normal dispersion. It remains, however, inferior to c in media of anomalous dispersion contrary to the group velocity. The energy flow velocity has a relatively simple definition and is based on the electromagnetic field.

4. The signal velocity coincides like the energy flow velocity with the group velocity if we deal with normal dispersion but deviates markedly from it whenever the group velocity behaves anomalously. The signal velocity is always less than c, but becomes somewhat difficult and arbitrary to define in the neighbourhood of resonances in the dispersive medium.

In Figure 3.5 we give a summary of the behaviour of the different propagation velocities. $\omega_0$ denotes in this case a resonance in the dispersive media. For example an absorption line in a plasma. It is important to note that the vertical axis is proportional to $c/v$, where $v$ is one of the propagation velocities considered.
in this section. Thus, if the curves lie above 1, the velocity is *inferior* to the velocity of light.

Far away from the resonance, where the phase velocity behaves smoothly, all velocities are smaller than c. But for frequencies around and above they differ significantly. It becomes also clear from this illustration that the velocities of energy and signal remain inferior to the velocity of light in all circumstances.

**Figure 3.5** Illustration of the different propagation velocities: phase velocity (−−−), group velocity (--), velocity of energy flow (- - -) and signal velocity (···).
Velocities of propagation
CHAPTER 4

Electromagnetic Waveguides

4.1 Introduction

In low-frequency circuits, the connection between two devices is often made with wires, but this method does not work very well at high frequencies because the circuits would radiate energy into all the space around them, and it is hard to control where the energy will go. The electromagnetic field spreads out around the wires; the currents and voltages are not “guided” very well by the wires. In this chapter we want to look into the ways to interconnect objects at high frequencies. We will see what happens when oscillating fields are confined in one or more dimensions and discover the interesting phenomenon when the fields are confined in only two dimensions and allowed to go free in the third one, i.e., they propagate in form of waves. These are guided waves and the devices which guide them are called waveguides. The term waveguide can be defined as follows:

A waveguide is a device consisting of conducting and/or non-conducting media with a certain spatial structure and a special direction or axis, along which electromagnetic waves can be guided.

Let us consider an ordinary power transmission line like in Figure 4.1 that runs over the countryside. It radiates away some of its power, but the frequencies are so low (50 – 60 Hz) that this loss is not serious. The radiation could be stopped by surrounding the line with a metal pipe, but this method would not be practical for power lines because the voltages and currents used would require a very large, expensive and heavy pipe. Thus, simple “open lines” are used.

![Figure 4.1](image.png) Power transmission line that runs over the countryside.
For somewhat higher frequencies, say a few kHz, radiation can already be serious. However, it can be reduced by using “twisted-pair” transmission lines, as is done for short-run telephone connections. At higher frequencies, however, the radiation soon becomes intolerable, either because of power losses or because the energy appears in other circuits where it is unwanted (e.g., cross talking). For frequencies from a few kHz to some GHz, electromagnetic signals and power are usually transmitted via coaxial lines consisting of a wire inside a cylindrical outer conductor which acts as shield (see Figure 4.2(a)).

Now imagine, we take out the inner conductor. All that remains is a hollow, circular pipe as in Figure 4.2(b). Can this pipe be still used to transmit power?

From a theoretical viewpoint, Heaviside was the first to introduce the concept of metallic conductors acting as guide for an electromagnetic wave. He also developed the basis for what is today called transmission-line theory. In 1893, he considered various possibilities for waves along wire lines from a theoretical standpoint and concluded that single conductor lines were not feasible. Heaviside was convinced that guided waves needed two separate conductors or at least a “real” one and the earth (ground) acting as second conductor. Furthermore, he concluded that the transmission of plane waves within a single hollow tube was only of a theoretical interest, since needing both electric and magnetic conductance:

“It does not seem possible to do without inner conductor, for when it is taken away we have nothing left upon which the tubes of displacement can terminate internally, and along which they can run...

It would appear that the only way of completely solving the problem of the automatic transmission of plane waves within a single tube is a theoretical one, employing magnetic as well as electric conductance”.¹ ²

It was Lord Rayleigh who extended the theory of transmission of electromagnetic waves. He was studying cylindrical structures and solving the associated

---

¹Oliver Heaviside, *Electromagnetic Theory*, p. 400.
²Note the reference to plane waves. Heaviside did not consider other possibilities for a tube although he did consider a spherical wave bound to a single wire.
boundary value problem of Maxwell’s equations. For a space bounded by a cylin-
drical surface of arbitrary cross-sectional shape, he was able to show that waves
indeed can propagate within a hollow conducting cylinder. He found that such
waves existed only in a set of well-defined normal modes. Furthermore, he was
able to infer that two types of waves exist: one with a longitudinal component
in the electric field what we today call TM- or E-mode, and the other one with a
longitudinal component in the magnetic field, today called TE- or H-mode.
He found that a fundamental limitation on the existence of such waves was
that the frequency must exceed a lower limit depending on the mode number
and the cross-sectional dimensions of the cylinder. Thus, Rayleigh was able to
give specific solutions for the cases of cylinders of rectangular and circular cross
section.
Other contributions to guided waves in these early years were given by Som-
merfeld (1899), Hondros and Debye (1909-1910).
The first real demonstration of the use of hollow waveguides did not happen
until 1936 and is in general attributed to Southworth and Barrow, who inde-
dependently worked on the field of wave transmission through hollow pipes. Their
work was mainly focused on modes in circular hollow metal guides, for the fields,
the cutoff frequencies, the guide wavelengths and the attenuation constants.
After this first successful demonstration of circular waveguides, contributions
were published about rectangular and elliptical waveguide propagation. Soon
afterwards, discontinuities between waveguides were considered and important
contributions were done by Marcuvitz, Schwinger, Oliver, Kahn, Lewin, and
Whinnery.
During the postwar years, the technology developed in the preceding years
spread rapidly. The research and development activities resulted in major ad-
vances in microwave technology. New waveguide structures to replace rectan-
gular and circular guides were introduced.
Today, the telecommunication market is a fast growing market and there is
a continuously growing demand towards higher frequencies. Multimedia ap-
lications are planned in the near future for commercial use, remote learning
and on-line training, teleworking, mobile phones, fast Internet connections, etc.
These applications will use broadband communication systems to establish a
worldwide mobile information system for all information types and will require
wide bandwidths, and consequently high frequencies and transmission links of
high bit rates.
As we will see in this chapter, the waveguide is both, a problem of propaga-
tion and a boundary value problem. The spatial and time dependence of the
electromagnetic field in such waveguides are described by Maxwell’s equations
together with the constitutive equations of the involved materials.

4.2 Transmission line theory

A transmission line is the medium or structure that forms all or part of a path
from one place to another for directing the transmission of energy, such as elec-
Electromagnetic Waveguides

\[ z - \Delta z \quad i(z - \Delta z) \quad i(z + \Delta z) \quad u(z - \Delta z) \quad u(z + \Delta z) \quad z - \Delta z \quad z + \Delta z \]

**Figure 4.3** Currents and voltages on a transmission line.

Electromagnetic or acoustic waves, as well as electric power transmission. Components of transmission lines include wires, coaxial cables, dielectric slabs, optical fibres, electric power lines, and waveguides.

### 4.2.1 Review

Oliver Heaviside developed the transmission line model, also known as the telegrapher equations, that describes how electrical voltage and current vary along a conductor.

The theory applies to high-frequency transmission lines (such as telegraph wires and radio frequency conductors) but is also important for designing high-voltage energy transmission lines. The equations consist of two linear differential equations in time and position: one for \( v(z, t) \) and the other one for \( i(z, t) \).

The telegrapher equations can be understood as a simplified case of Maxwell’s equations. In a more practical approach, one assumes that the conductor is composed out of an infinite series of two-port elementary components, each representing an infinitesimally short segment of the transmission line.

To derive the transmission line equations, we consider the simplest coaxial line that has a central conductor and an outer conductor as in Figure 4.2(a). We assume that the electric conductors are perfect, i.e. without losses as well as the dielectric between the conductors.

We begin figuring out approximately how the line behaves at relatively low frequencies. Assume that we can measure what happens at the two neighbouring points \( z - \Delta z \) and \( z + \Delta z \) on the coaxial line (see Figure 4.3). The inner and outer conductor represent two coupled lines, that is, a current flowing from \( z - \Delta z \) to \( z + \Delta z \) on the inner conductor passes through the magnetic field of the outer conductor and vice versa as sketched in Figure 4.4. We apply Faraday’s law and integrate over the section \( S \) circumscribed by the lines A, B, C and D

\[
\iint_S \text{rot} \mathbf{E} \, d\mathbf{s} = -\frac{\partial}{\partial t} \iint_S \mathbf{B} \, d\mathbf{s}.
\] (4.1a)
The orientation of the surface element $dS$ is related to the path $\partial S = ABCD$ by the right hand rule. Applying Stokes’ theorem to the left hand side we obtain

$$
\int_{\partial S} \mathbf{E} \cdot d\mathbf{l} = \int_{A} \mathbf{E} \cdot d\mathbf{l} + \int_{B} \mathbf{E} \cdot d\mathbf{l} + \int_{C} \mathbf{E} \cdot d\mathbf{l} + \int_{D} \mathbf{E} \cdot d\mathbf{l} = u(z+\Delta z) + 0 - u(z-\Delta z) + 0 = -\frac{\partial}{\partial t} \int_{S} B \, dS. 
$$

(4.1b)

The integral over the magnetic flux density $B$ yields the total magnetic flux

$$
\Phi_m = \int_{S} B \, dS \approx B(z) \cdot 2\Delta z \cdot d, 
$$

(4.1c)

where $d$ is the distance between the inner and outer conductor. The magnetic flux $\Phi_m$ is therefore proportional to the distance $2\Delta z$.

On the other hand, we know that $B$ is generated by the electric current $i$. The ratio

$$
L_0 = \frac{\Phi_m}{i} 
$$

(4.1d)
Electromagnetic Waveguides

is the inductance, a term coined by Heaviside, thus we have

\[ \Phi_m \approx B(z) \cdot 2\Delta z \cdot d = L_0 i(z) \quad (4.1e) \]

and find

\[ B(z) = \frac{L_0}{2\Delta z} i(z) \quad (4.1f) \]

The short distance \(2\Delta z\) can therefore be represented by an inductance \(L_0\) and we can introduce the inductance per unit length \(L' = L_0/(2\Delta z)\). Inserting this result into (4.1b) we obtain

\[ u(z + \Delta z) - u(z - \Delta z) \approx 2\Delta z \cdot \partial u/\partial z = -L' \cdot 2\Delta z \cdot \partial i/\partial t \quad (4.1g) \]

If the current in the line is varying with respect to the time, then the inductance \(L' \cdot 2\Delta z\) will give us a voltage drop across the small portion of line from \(z - \Delta z\) to \(z + \Delta z\). Dividing both sides by \(2\Delta z\) and taking the limit \(\Delta z \to 0\), we get

\[ \frac{\partial u}{\partial z} = -L' \frac{\partial i}{\partial t} \quad (4.2) \]

The time-dependent current \(i\) produces a gradient of the voltage \(u\).

On the other side, the two lines also represent a capacity \(C_0\) between \(z - \Delta z\) and \(z + \Delta z\). To see this we consider Figure 4.5 and apply the law of charge conservation

\[ \iint_V \text{div} \mathcal{J} \, dV = \iint_V -\frac{\partial}{\partial t} \int\int \rho \, dV. \quad (4.3a) \]

The integral on the left hand side can be reduced to a surface integral over the volume \(V\) using Gauß’ theorem and the integral on the right hand side evaluates to the charge enclosed by \(V\). Thus we obtain

\[ \iint_{\partial V} \mathcal{J} \, d\mathcal{S} = \iint_{S_{z-\Delta z}} \mathcal{J} \, d\mathcal{S} + \iint_{S_{z+\Delta z}} \mathcal{J} \, d\mathcal{S} + \iint_M \mathcal{J} \, d\mathcal{S} = -\frac{\partial Q}{\partial t}. \quad (4.3b) \]

\(M\) is the lateral surface of the cylinder through which no current is flowing. We therefore have

\[ -i(z - \Delta z) + i(z + \Delta z) + 0 = -\frac{\partial Q}{\partial t}. \quad (4.3c) \]

The charge \(Q\) is given on the one hand by

\[ Q = C_0 \cdot u(z) \quad (4.3d) \]

and on the other hand, since the charge is located on the line, by
Chapter 4  4.2 Transmission line theory  57

Outer conductor

\[ i(z - \Delta z) \]

Inner conductor

\[ i(z + \Delta z) \]

\[ D \]

\[ z - \Delta z \]

\[ z + \Delta z \]

**Figure 4.5** Transmission line model of Figure 4.3. The outer and inner conductor represent also a capacitance.

\[ Q = \iiint_V \rho \, dV = \int \rho_L(z) \, dz \approx \rho_L(z) \cdot 2\Delta z. \quad (4.3e) \]

Thus, the charge must be proportional to \(2\Delta z\). Analogously to \(L'\), we can introduce the capacity per unit length \(C' = C_0/(2\Delta z)\). Combining the two last equations and inserting them into (4.3c) we get

\[ 2\Delta z \frac{\partial i}{\partial z} = i(z + \Delta z) - i(z - \Delta z) = -C'2\Delta z \frac{\partial u}{\partial t} \quad (4.3f) \]

and for the limit \(\Delta z \to 0\)

\[ \frac{\partial i}{\partial z} = -C' \frac{\partial u}{\partial t}. \quad (4.4) \]

Thus, the conservation of charge implies that the gradient of the current is proportional to the time rate-of-change of the voltage.

Equations (4.2) and (4.4) are the basic equations of a transmission line and represent a system of coupled differential equations of first order. The transmission line can be thus described as series-connected inductance, denoted by \(L'\) and shunt-connected capacitance per unit length, denoted by \(C'\) as shown in Figure 4.6(a).

We can enhance them by introducing losses on the lines by a resistance in series with \(L'\) and account for losses in the material between the lines (e.g. a dielectric filling) by adding a resistance parallel to the capacity \(C'\) and obtain the model of Figure 4.6(b).
4.2.2 The Telegraph Equations and their solutions

To decouple the two equations (4.2) and (4.4), we can differentiate one with respect to \( t \) and the other with respect to \( z \) and eliminating either \( u \) or \( i \). Then we have either

\[
\frac{\partial^2 u}{\partial z^2} = L'C \frac{\partial^2 u}{\partial t^2}
\]

or

\[
\frac{\partial^2 i}{\partial z^2} = L'C \frac{\partial^2 i}{\partial t^2}.
\]

We can recognize the wave equation (see Section 2.9.2) in \( z \). For an uniform waveguide, the voltage and current propagate along the line as wave. The voltage and the current along the line must be of the form, respectively, \( u(\pm)(z,t) = \hat{u}(\pm)(z \pm vt) \) and \( i(\pm)(z,t) = \hat{i}(\pm)(z \pm vt) \), or a sum of both waves travelling in opposite directions.

The velocity \( v \) of the wave can be deduced from the term in front of \( \partial^2 / \partial t^2 \), thus

\[
v = \frac{1}{\sqrt{L'C}}.
\]

Assuming a time-harmonic dependence of \( u \) and \( i \), the derivative with respect to \( t \) can be replaced by \( j\omega \). \( u \) and \( i \) are replaced by the phasors \( U \) and \( I \) and equations (4.2) and (4.4) become

\[
\frac{\partial^2 U}{\partial z^2} = -\frac{\omega^2}{c^2} U,
\]

\[
\frac{\partial^2 I}{\partial z^2} = -\frac{\omega^2}{c^2} I.
\]

The solutions to these differential equations of second order are complex exponentials, that is \( U(\pm) = \hat{U}(\pm)e^{\mp jkz} \) and \( I(\pm) = \hat{I}(\pm)e^{\mp jkz} \), with \( k = \omega/c \).
4.2.3 Example: coaxial cable and its transmission line parameters

To find the propagation speed $v$ of a transmission line, we have to know the inductance and capacitance per unit length. For the present example of a coaxial cable, these can be easily calculated. Again, we limit our considerations to the lossless case, thus for the model of Figure 4.6(a).

The capacitance derives from Gauss’ law. Suppose the inner conductor carries the charge density $\rho_L$. Due to the cylindrical symmetry, we can decompose the electric field into a radial and an angular component and integrate over a line of radius $r$. We get

$$\rho_L = \varepsilon_0 \int_0^{2\pi} E_r \cdot r \, d\varphi = 2\pi r \varepsilon_0 E_r.$$  \hfill (4.8)

With this, we can calculate the voltage $u$ between the inner conductor with radius $a$ and the outer conductor having radius $b$. The voltage is simply the integration of the electric field along a line. In the case of a coaxial cable, the voltage is independent of the path of integration, therefore we use a radial path and obtain

$$u = \int_a^b E_r(r) \, dr = \frac{\rho_L}{2\pi \varepsilon_0} \ln \frac{b}{a}.$$ \hfill (4.9)

Capacity is defined as the ratio of charge and voltage. Using the last equation, we find the capacitance per unit length

$$C' = \frac{\frac{2\pi \varepsilon_0}{\ln \frac{b}{a}}}{\ln \frac{b}{a}}.$$ \hfill (4.10)

To find the inductance, we proceed in a similar way. For this, we use the magnetic energy $W_m = \frac{1}{2} L_0 i^2$. This we set equal to the magnetic energy which we get by integrating the magnetic flux $B$ over a cylindrical shell of thickness $dr$ and of length $l$. For $B$ we have from Ampère’s law

$$B_\varphi = \frac{\mu_0 i}{2\pi r}.$$ \hfill (4.11)

We find for the magnetic energy

$$W_m = \int_a^b \int_0^{2\pi} \int_0^l \frac{1}{2} H_{\varphi} \cdot B_\varphi \, dz \, d\varphi \, dr = \frac{1}{2\mu_0} \int_a^b \left( \frac{\mu_0 i}{2\pi r} \right)^2 2\pi l r \, dr.$$ \hfill (4.12)

Carrying out the integral, we get

$$W_m = \frac{\mu_0 i^2 \cdot l}{4\pi} \ln \frac{b}{a}.$$ \hfill (4.13)

Setting this result equal to $\frac{1}{2} L_0 i^2$, we find

$$L' = \frac{L_0}{l} = \frac{\mu_0 \ln \frac{b}{a}}{2\pi}.$$ \hfill (4.14)
Combining (4.10) and (4.14) we finally see that the product $L'C'$ is just equal to $1/c_0^2$. Thus, $v$ is equal to $c_0$. The wave travels down the line with the speed of light. We have to point out that this result depends on two assumptions:

1. that there are no dielectric or magnetic materials in the space between the conductors, and
2. that we deal with perfect conductors.

It is interesting to note that as long as these two assumptions are fulfilled, the velocity $v$ is equal to $c_0$ for any pair of conductors.

4.3 Uniform waveguides

The waveguides to be analyzed in this chapter are all characterized by having axial symmetry. Their cross-sectional shape and electrical properties, e. g. boundary conditions, do not vary along the axis which is chosen as the $z$ axis. Such kind of waveguides are called uniform waveguides. We first define the term uniform and provide some examples of uniform waveguides.

4.3.1 Definition of uniform waveguides

By an uniform waveguide we understand a waveguide with the following properties:

1. The waveguide has the form of a general, infinitely long cylinder, with a finite cross section.
2. The properties of the materials (metal and dielectrics) do not depend on the position along the waveguide’s axis, nor on the time, the direction, or the electromagnetic field. The conductivity of the metal is denoted by the scalar $\sigma$, the permittivity of the dielectrics by $\varepsilon$ and their permeability by $\mu$.
3. In the ideal case, which we will assume in the sequel, the metal is a perfect electric conductor, i. e. $\sigma = \infty$ and the dielectrics without losses, which implies that $\varepsilon$ and $\mu$ are real-valued scalar constants and therefore frequency independent.

With this definition of an uniform waveguide we exclude all kind of waveguides that are curved or winded as well as those waveguides having a cross section that is varying along the waveguide’s axis. Examples of non-uniform waveguides are shown in Figure 4.7. Note, the bend is a non-uniform waveguide, although its cross section along the guide’s axis is constant. However, the waveguide’s axis is curved and therefore, the waveguide does not represent a general cylinder with a straight axis. The second example represents a waveguide with a continuously varying cross section, and the third one shows a waveguide with constant cross section and straight axis. However, the dielectric filling inside the waveguide is changing.
Chapter 4  4.3 Uniform waveguides

Figure 4.7 Examples of non-uniform waveguides. The left picture shows a waveguide bend. It is a non-uniform waveguide because its axis is curved. The centre picture shows a waveguide with a continuously varying cross section, and the right picture a waveguide with a dielectric filling that changes. The right waveguide can be, however, considered as a piecewise uniform waveguide.

After these examples of non-uniform waveguides, we will return to the problem of uniform waveguides. In the remainder, we always mean uniform waveguide, when we speak about waveguides, although we do not always explicitly mention the term uniform. Figure 4.1 shows various kinds of uniform waveguides. They can be classified into two groups: The first group contains waveguides, the electromagnetic field of which is bounded to a finite space by a metallic enclosure. For instance, this is the case for the rectangular, circular, or coaxial waveguide. The other group consists of waveguides that are "open", i.e. the electromagnetic field can radiate into the free space around the waveguide. This, for instance, is the case for the microstrip line, the slot line, or the coplanar line as well as for the optical fibre.

There are essentially two ways to force the electromagnetic field to propagate along a certain direction:

1. Either the field is “imprisoned” inside a metallic enclosure. Examples are hollow metallic waveguides such as rectangular, circular, or coaxial ones. These types of waveguides do not show any radiation effects. Or

2. The use of dielectrics of high permittivity as in the case of the optical fibre. The high permittivity dielectrics bind the electromagnetic field and force it to follow the waveguide's axis. For instance, in the case of the optical fibre, the effect of total reflection is used to guide the electromagnetic field.

It will not be possible to examine in detail all the different structures that have been introduced for waveguiding. We shall restrict the discussion to examining the basic theory of hollow metallic, cylindrical waveguides with specific emphasis on the rectangular, circular empty and circular-coaxial waveguide.
<table>
<thead>
<tr>
<th>Coupled Lines</th>
<th>Hollow Metallic Waveguides</th>
<th>Printed Circuits</th>
<th>Shielded Printed Circuits</th>
<th>Dielectric Waveguides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telegraph Line</td>
<td>Rectangular Waveguide</td>
<td>Microstrip Line</td>
<td>Suspended Strip Line</td>
<td>Optical Fibre</td>
</tr>
<tr>
<td>Coaxial Line</td>
<td>Circular Waveguide</td>
<td>Slot Line or Fin Line</td>
<td>Inverted Suspended Strip Line</td>
<td>Planar Dielectric or Slab-Dielectric Waveguide</td>
</tr>
<tr>
<td>Strip Line</td>
<td>Ridged Waveguide</td>
<td>Coplanar Line</td>
<td>Shielded Strip Line</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.1** Different examples and types of uniform waveguides.
4.3 Uniform waveguides

4.3.2 Cylindrical coordinate system

A suitable choice for analyzing uniform waveguides are cylindrical coordinates as shown in Figure 4.8. We denote the axis of the cylinder by \( z \) and the unit vector by \( \mathbf{e}_z \). The coordinates in the transverse plane are denoted by \( u \) and \( v \) and the corresponding unit vectors by \( \mathbf{e}_u \) and \( \mathbf{e}_v \). Further, we introduce the tangential unit vector \( \mathbf{t} \) and the outward normal \( \mathbf{n} \) with respect to the boundary in the transverse plane.

This special kind of coordinate system suggests to separate field quantities into components parallel to and transverse to the \( z \)-axis

\[
\mathcal{E} = \mathcal{E}_t + \varepsilon_z E_z \quad \text{and} \quad \mathcal{H} = \mathcal{H}_t + \varepsilon_z H_z,
\]

where the subscript \( t \) denotes the transverse part and \( E_z/H_z \) the longitudinal part of the field, i.e., the part of the field which is parallel to the guide’s axis. The same procedure can be applied to the mathematical operators

\[
\begin{align*}
\text{grad } \Phi &= \text{grad}_t \Phi + \varepsilon_z \frac{\partial \Phi}{\partial z}, \\
\text{div } \mathcal{A} &= \text{div}_t \mathcal{A}_t + \varepsilon_z \frac{\partial A_z}{\partial z}, \\
\text{rot } \mathcal{A} &= \varepsilon_z \times \left( -\text{grad}_t A_z + \frac{\partial A_t}{\partial z} \right) + \varepsilon_z \frac{\partial A_t}{\partial z}.
\end{align*}
\]

and for the nabla-operator

\[
\nabla = \nabla_t + \varepsilon_z \frac{\partial}{\partial z}.
\]
4.3.3 Electromagnetic waves in cylindrical systems

Before proceeding with the special case of cylindrical coordinates as in Figure 4.8 we will formulate the boundary-value problem. In the remaining part of this chapter, we only consider source-free regions and linear, isotropic materials. The complete boundary-value problem is therefore given by

\[
\begin{align*}
\text{rot} \, \mathbf{H} & = \varepsilon \frac{\partial \mathbf{E}}{\partial t}, & \text{rot} \, \mathbf{E} & = -\mu \frac{\partial \mathbf{H}}{\partial t}, \\
\text{div} \, \mathbf{H} & = 0, & \text{div} \, \mathbf{E} & = 0,
\end{align*}
\]

(4.17a)

\[
\begin{align*}
n \cdot \mathbf{H} & = 0 \quad \forall \mathbf{r} \in \partial S, \\
n \times \mathbf{E} & = 0 \quad \forall \mathbf{r} \in \partial S,
\end{align*}
\]

(4.17c)

where \( \partial S \) is the boundary of the waveguide.

In Section 2.9 we have learnt that this system of coupled differential equations can be decoupled. Taking into account that we deal with a source-free region, the vector wave equation can be transformed into

\[
\begin{align*}
\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} & = 0, \\
\nabla^2 \mathbf{H} - \frac{1}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} & = 0.
\end{align*}
\]

(4.18a)

(4.18b)

The electric and magnetic field have to satisfy, respectively, the vector wave equation.

Although, this is a significant simplification, the individual components of the electric or magnetic field still remain coupled. Except for Cartesian coordinates, the \( \nabla \)-operator relates the different components of the vector \( \mathbf{E} \) respectively \( \mathbf{H} \) with each other. We therefore deal with three complicated scalar partial differential equations. Hence, solving the above partial differential equations directly and generally will be quite difficult.

For this purpose, some methods for reducing the 3D-problem to a 1D-problem under certain conditions can be found in the literature. We will study one of them in the next section. It is based on the electric and magnetic Hertz vectors introduced in Section 2.9.6 and Section 2.9.8. Another method is explained in the appendix E and is called the method of longitudinal components. There, the equivalence of the two approaches is shown, too.

4.3.4 Method of Hertz Vectors

Since we deal with a source-free region, we can express the electric and magnetic field by means of vector potentials \( \mathcal{F} \) and \( \mathcal{A} \), respectively
\[ E = -\frac{1}{\varepsilon} \text{rot} F \quad \text{and} \quad \mathcal{H} = \frac{1}{\mu} \text{rot} A. \quad (4.19) \]

Since for an uniform waveguide we deal with a cylindrical coordinate system, we try to separate the longitudinal component \( F_z \) respectively \( A_z \) from the transverse one \( F_t \) and \( A_t \), respectively. By doing so, we obtain

\[ E = -\frac{1}{\varepsilon} \text{rot}(F_z \varepsilon_z) - \frac{1}{\varepsilon} \text{rot} F_t, \quad (4.20a) \]
\[ \mathcal{H} = \frac{1}{\mu} \text{rot}(A_z \varepsilon_z) + \frac{1}{\mu} \text{rot} A_t. \quad (4.20b) \]

Taking a closer look at these equations we observe that the first expression on the right hand sides yields only a transverse field. For example

\[ \text{rot}(F_z \varepsilon_z) = \nabla \times (F_z \varepsilon_z) = (\nabla F_z) \times \varepsilon_z = -\varepsilon_z \times \text{grad} F_z = -\varepsilon_z \times \text{grad}_t F_z \]
and analogously for \( A \).

Moreover, applying the Lorenz gauge, we have

\[ 0 = \text{div} F + \frac{1}{c^2} \frac{\partial \psi}{\partial t} = \text{div}_t F_t + \frac{\partial F_z}{\partial z} + \frac{1}{c^2} \frac{\partial \psi}{\partial t}. \quad (4.21a) \]

If we assume that \( F \) can be chosen such that

\[ \frac{\partial F_z}{\partial z} = -\frac{1}{c^2} \frac{\partial \psi}{\partial t} \quad (4.21b) \]

then we find that

\[ \text{div}_t F_t = 0 \quad (4.21c) \]

which allows us to express \( F_t \) by means of a single longitudinal component

\[ F_t = \text{rot} F^*_z \varepsilon_z. \quad (4.21d) \]

Analogously, we can proceed with \( \mathcal{H} \) and \( A \) and obtain finally

\[ E = -\frac{1}{\varepsilon} \text{rot}(F_z \varepsilon_z) - \frac{1}{\varepsilon} \text{rot} \text{rot}(F^*_z \varepsilon_z), \quad (4.22a) \]

purely transverse

\[ \mathcal{H} = \frac{1}{\mu} \text{rot}(A_z \varepsilon_z) + \frac{1}{\mu} \text{rot} \text{rot}(A^*_z \varepsilon_z). \quad (4.22b) \]

purely transverse

(4.22c)
Now we remember that we saw in Section 2.9.6 and Section 2.9.8 how the electric and magnetic field can be found once the electric and magnetic Hertz vectors are known. By superposition we obtained

\[ E = -\text{rot} \frac{\partial \Pi_m}{\partial t} + \frac{1}{\varepsilon} \text{rot rot} \Pi_e, \]  

\[ H = \text{rot} \frac{\partial \Pi_e}{\partial t} + \frac{1}{\mu} \text{rot rot} \Pi_m. \]  

If we compare this with (4.22) we see that \( F^*_z \varepsilon_z \) and \( A^*_z \varepsilon_z \) represent Hertzian vectors. We therefore set

\[ \Pi_e = -F^*_z \varepsilon_z \quad \text{and} \quad \Pi_m = A^*_z \varepsilon_z \]  

and find that

\[ F_z \varepsilon_z = \varepsilon \frac{\partial \Pi_m}{\partial t} \quad \text{and} \quad A_z \varepsilon_z = \mu \frac{\partial \Pi_e}{\partial t}. \]  

We come to the conclusion that to solve the electromagnetic field associated with uniform waveguides, it suffices to find an electric and a magnetic Hertz vector which only has a \( z \)-component, i.e. a component \textit{parallel} to the waveguide’s axis.

Let us define

\[ \Pi_e = \Phi \varepsilon_z \quad \text{and} \quad \Pi_m = \Psi \varepsilon_z. \]  

From Section 2.9.6 and Section 2.9.8 we know that the Hertzian vectors have to satisfy the wave equation. Since in the case of uniform waveguides they only have a \( z \)-component, the vector wave equation reduces to a scalar wave equation. Because no sources are present, we find that \( \Phi \) and \( \Psi \) are defined by

\[ \triangle \Phi - \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} = 0, \]  

\[ \triangle \Psi - \frac{1}{c^2} \frac{\partial^2 \Psi}{\partial t^2} = 0. \]  

Hence, the original vector wave equation and its solution can be reduced to a \textit{scalar} wave equation associated with the longitudinal components of Hertzian vectors.

### 4.3.5 Boundary conditions

The electromagnetic field has to satisfy the general boundary conditions of the electromagnetic field on the waveguide walls. Since we limit our study to hollow metallic waveguides assuming perfect electric conductors, the tangential electric...
field has to vanish on the surface of the waveguide. Denoting by $\partial S$ the boundary of the waveguide's cross section, we can formulate the boundary condition as

$$n \times E = 0 \quad \text{on} \quad \partial S$$  \hspace{1cm} (4.27)

where $n$ is defined as in Figure 4.8.

Using the method of Hertz vectors, the electric field is given by (4.23a). Applying the above boundary condition we find that $\Phi$ and $\Psi$ must satisfy the equation

$$\frac{1}{\varepsilon} \nabla \times \nabla t \frac{\partial \Phi}{\partial z} + n \times \left( \epsilon_z \times \nabla t \frac{\partial \Psi}{\partial t} \right) - \frac{1}{\varepsilon} \nabla \Phi (n \times \epsilon_z) = 0 \quad \text{on} \quad \partial S \quad \forall \quad t \quad \forall \quad z.$$  \hspace{1cm} (4.28a)

Making use of the tangential vector $t = \epsilon_z \times n$ (see Figure 4.8) the boundary condition can be reformulated as

$$\frac{1}{\varepsilon} \frac{\partial \Phi}{\partial t} \epsilon_z + \frac{\partial \Psi}{\partial t} \epsilon_z + \frac{1}{\varepsilon} \nabla \Phi t = 0 \quad \text{on} \quad \partial S \quad \forall \quad t \quad \forall \quad z$$  \hspace{1cm} (4.28b)

being $\partial/\partial t$ the derivative along the tangential vector $t$ and $\partial/\partial n$ the normal derivative.

Since the last equation must hold for all coordinates $z$ and times $t$ we obtain

$$\Phi = 0 \quad \text{and} \quad \frac{\partial \Psi}{\partial n} = 0 \quad \text{on} \quad \partial S \quad \forall \quad t \quad \forall \quad z$$  \hspace{1cm} (4.29)

because $\Phi$ is evaluated along $n$ on $\partial S$ but differentiated with respect to the tangential direction $t$ in the first term. In the second term $\Psi$ is differentiated with respect to time for which reason $\partial \Psi/\partial n = 0$ holds. The last term yields as well $\Phi = 0$ since $\nabla \Phi = -\frac{\partial^2 \Phi}{\partial z^2} + \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2}$.

### 4.3.6 Separation of variables

Using the methods given in the previous sections, the boundary value problems of vector wave equations may reduce to the problems of solving the scalar wave equation

$$\nabla^2 U - \frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} = 0$$  \hspace{1cm} (4.30a)

where $U$ represents either $\Phi$ or $\Psi$ with the boundary conditions, resp.,

$$U_{|\partial S} = 0 \quad \text{or} \quad \frac{\partial U}{\partial n}_{|\partial S} = 0$$  \hspace{1cm} (4.30b)

The method or separation of variables is an important and convenient way to solve scalar partial differential equations in mathematical physics. By choosing an appropriate orthogonal coordinate system, we can represent the solution by a product of three functions (in a 3D problem), one for each coordinate, and the 3D partial differential equation is reduced to three ordinary differential equations. The functions satisfying these ordinary differential equations are
orthogonal function sets called harmonics. The solution of the differential equation with specific boundary conditions is usually a (infinite) series of the specific harmonics set.

Equations involving the 3D Laplacian operator, for example Laplace’s equation and Helmholtz’s equation, are known to be separable in eleven different orthogonal coordinate systems, included in the following three groups:

1. Cylindrical coordinate systems
2. Rotational coordinate systems
3. General coordinate systems

A more detailed description of these coordinate systems can be found in Appendix B.

4.3.7 Types of electromagnetic waves in cylindrical systems

In an arbitrary cylindrical coordinate system, \( u, v, z \) the functions \( \Phi \) and \( \Psi \), resp., satisfy the same scalar wave equation

\[
\Delta U - \frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} = \Delta_t U + \frac{\partial^2 U}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} = 0. \tag{4.31}
\]

Applying the method of separation of variables, we set

\[
U(u,v,z,t) = U_t(u,v) \cdot Z(z,t), \tag{4.32}
\]

where \( U_t \) represents the transverse function of \( U \), i.e. it depends only on the transverse coordinate \( u \) and \( v \). \( Z \) on the other hand denotes the longitudinal part of \( U \), i.e. the part of \( U \) which varies with the waveguide’s axis. Substituting this ansatz into (4.31) and dividing by \( U \equiv 0 \) yields

\[
\frac{\Delta_t U_t}{U_t} + \frac{\partial^2 Z/\partial z^2}{Z} - \frac{1}{c^2} \frac{\partial^2 Z/\partial t^2}{Z} = 0 \quad \forall u,v,z,t. \tag{4.33}
\]

We always can assume \( U \neq 0 \) since \( U = 0 \) represents the trivial solution in which we are not interested.

The first term is a function of \( u \) and \( v \) only and the second and third term depend only on \( z \) and \( t \). Each of them must be equal to a constant so that the sum of them vanishes. That is

\[
\frac{\Delta_t U_t}{U_t} = -T^2, \quad \frac{\partial^2 Z/\partial z^2}{Z} - \frac{1}{c^2} \frac{\partial^2 Z/\partial t^2}{Z} = T^2 \quad \forall u,v,z,t. \tag{4.34a}
\]

Assuming a time harmonic dependency we can further separate \( Z(z,t) = Z_1(z) \cdot Z_2(t) \), with \( Z_2(t) = e^{j\omega t} \). Substituting this into the previous equations yields

\[
\frac{\Delta_t U_t}{U_t} = -T^2, \quad \frac{d^2 Z_1/\partial z^2}{Z_1} + \frac{\omega^2}{c^2} = T^2. \tag{4.34b}
\]
With the same argument as before \( \frac{\partial^2 Z_1/\partial z^2}{Z_1} \) must be equal to a constant. We will call this constant \( \gamma^2 \)

\[
\frac{\Delta_t U_t}{U_t} = -T^2, \quad \frac{d^2 Z_1/\partial z^2}{Z_1} = \gamma^2
\]  

(4.34c)

and

\[
\gamma^2 = T^2 - k^2 \quad \Rightarrow \quad \gamma = \sqrt{T^2 - k^2}, \quad k = \omega/c. 
\]  

(4.35)

Then we have

\[
\Delta_t U_t + T^2 U_t = 0, 
\]  

(4.36a)

\[
\frac{d^2 Z_1}{dz^2} - \gamma^2 Z_1 = 0. 
\]  

(4.36b)

The second equation (4.36b) is a 1D homogeneous scalar Helmholtz equation. We have encountered it already in Section 4.2.2 when we studied the Telegrapher Equation. The solutions of (4.36b) are two travelling waves propagating along \(+z\) and \(-z\), known as guided waves

\[
Z_1(z) = Z_a e^{-\gamma z} + Z_b e^{+\gamma z}, 
\]  

(4.37)

where \( \gamma \) is the propagation constant which is determined by \( k = \omega/c \) and \( T \) in (4.35).

The first equation (4.36a) is a 2D scalar Helmholtz equation which is known as the transverse wave equation and \( T^2 \) is the transverse eigenvalue which is determined by the boundary conditions of the system.

The guided waves in a bounded cylindrical system are classified according to the transverse eigenvalue \( T^2 \).

A) TEM, L- or Lecher wave (mode)

When \( T^2 = 0 \), then

\[
\gamma = j\beta = j\omega/c
\]  

(4.38)

and for the phase velocity holds \( v_{ph} = c \).

This is a wave with a velocity equal to the velocity of a plane wave in the unbounded medium. According to (E.13) this must be a wave with neither electric nor magnetic field in the direction of propagation, i.e. \( E_z = H_z = 0 \). It is known as transverse electromagnetic (TEM) wave. They are also called L- or Lecher waves in honour of Ernst Lecher (Wien, 1856-1925) who was the first to study the propagation of electromagnetic waves on coupled lines.

Under this condition the equations for the transverse fields \( E_t \) and \( H_t \) (E.5)
become
\[ \text{rot}_t \mathcal{E}_t = 0 \quad \text{and} \quad \text{rot}_t \mathcal{H}_t = 0. \]  
(4.39a)

The transverse fields are irrotational vector functions in the transverse cross-section. Moreover, at the beginning of the chapter we assumed source-free regions, i.e., \( \text{div} \mathcal{E} = \text{div} \mathcal{H} = 0 \). The fields are solenoidal vector functions and since \( E_z = H_z = 0 \) it holds as well that
\[ \text{div}_t \mathcal{E}_t = 0 \quad \text{and} \quad \text{div}_t \mathcal{H}_t = 0. \]  
(4.39b)

This means that the total fields \( \mathcal{E} \) and \( \mathcal{H} \) are solenoidal, and the transverse fields \( \mathcal{E}_t \) and \( \mathcal{H}_t \) are solenoidal and irrotational in the transverse cross section.

This resembles the static case of Maxwell’s equations where \( \mathcal{E} \) and \( \mathcal{H} \) decouple. In this case \( \mathcal{E} \) and \( \mathcal{H} \) can be expressed using gradients of scalar potentials.

For the case of TEM waves (modes), the Hertzian vectors \( \Phi \) and \( \Psi \) play the role of the potentials. We can choose either \( \Phi \neq 0 \) and \( \Psi = 0 \) or \( \Phi = 0 \) and \( \Psi \neq 0 \) or any linear combination of both cases. Since \( T^2 = 0 \), they satisfy both the Laplace equation, that is
\[ \triangle \Phi = 0 \quad \text{or} \quad \triangle \Psi = 0. \]  
(4.40)

The fields are given by
\[ \mathcal{E}_t = \frac{1}{\varepsilon} \text{grad}_t \frac{\partial \Phi}{\partial z}, \quad \mathcal{E}_t = \varepsilon_z \times \text{grad}_t \frac{\partial \Psi}{\partial t}, \]  
\[ \mathcal{H}_t = -\varepsilon_z \times \text{grad}_t \frac{\partial \Phi}{\partial t}, \quad \mathcal{H}_t = \frac{1}{\mu} \text{grad}_t \frac{\partial \Psi}{\partial z}. \]  
(4.41a)

The boundary conditions are
\[ \frac{\partial \Phi}{\partial t} = 0 \quad \text{or} \quad \frac{\partial \Psi}{\partial n} = 0, \quad \text{on } \partial S. \]  
(4.41b)

Thus the tangential derivative of \( \Phi \) or the normal derivative of \( \Psi \) have to vanish on the boundary \( \partial S \). \( \partial \Phi/\partial t = 0 \) means that \( \Phi \) is constant on the boundary \( \partial S \).

Since \[ \triangle \Phi = 0 \quad \text{or} \quad \triangle \Psi = 0 \] it is shown that the scalar potentials \( \Phi \) respectively \( \Psi \) of the TEM mode satisfy the 2D Laplace equations. The transverse distribution of the electric and magnetic fields of a TEM wave is the same as those of static fields.

We come to the conclusion that the TEM mode can exist only in a 2D system (cross section) that can support static fields. This means that the TEM mode cannot be supported by a single conductor or insulator, no matter what the configuration is, and only the system composed of at least two conductors isolated from each other can carry TEM waves. This kind of system is known as a transmission line and can be analyzed by means of the field approach as well as the
circuit approach.

To avoid any misunderstanding: The static field behaviour applies to the 2D cross section and transverse fields $\mathcal{E}_t$ and $\mathcal{H}_t$. The full 3D field is nevertheless time-dependent and propagating along the waveguide with the same speed as a plane wave in an unbounded medium.

**B) Fast wave modes**

Here $T^2 > 0$, $T$ is real. The fields in the system depend on the relation between $T^2$ and $k^2$ as follows:

1. $T^2 < k^2$, $\gamma = j\beta$, $\beta^2 = k^2 - T^2 > 0$, $\beta$ is real and $\beta < k$. Since $v_{\text{ph}} = \omega/\beta$ and $\omega/k = c$, we have

   \[
   v_{\text{ph}} = \frac{\omega}{\beta} = \frac{\omega}{\sqrt{k^2 - T^2}} = \frac{c}{\sqrt{1 - T^2/k^2}} > c. \quad (4.42)
   \]

   This is a travelling wave along $z$, and the phase velocity is greater than the phase velocity of a plane wave in the unbounded media. Thus it is a fast wave mode. The fact that the phase velocity is greater than the velocity of light does not violate the special theory of relativity, because the phase velocity does not bring any matter, energy, or signal with it.

   Since $T^2$ is a constant, the group velocity becomes

   \[
   v_{\text{gr}} = \frac{1}{d\beta/d\omega} = c\sqrt{1 - T^2/k^2} < c. \quad (4.43)
   \]

   So the group velocity is less than the velocity of a plane wave in the unbounded medium, and

   \[
   v_{\text{ph}}v_{\text{gr}} = c^2. \quad (4.44)
   \]

   This is the propagation state of a fast wave mode in common metallic waveguides. The modes in propagation state are called propagating modes.

2. $T^2 > k^2$, $\gamma = \alpha$, $\alpha^2 = T^2 - k^2 > 0$, $\alpha$ is real. The field is not a travelling wave but a damping field along $z$. This is the cutoff state of a waveguide mode. The modes in cutoff state are called cutoff modes or evanescent modes.

3. $T^2 = k^2$, $\gamma^2 = k^2 - T^2 = 0$. This is the critical state of a waveguide mode. Hence the transverse eigenvalue $T$ is also known as the critical angular wave number or cutoff angular wave number.

   \[
   k_c = \omega_c/c = T \quad (4.45)
   \]

   where $\omega_c$ denotes the cutoff angular frequency of the waveguide.
Fast waves cannot be TEM waves. However, since we decoupled Maxwell’s equations for cylindrical systems by means of Hertz vectors $\Pi = \Phi e_z$ and $\Pi_m = \Psi e_z$, the fields related to $\Phi$ are independent of the fields related to $\Psi$.

We can distinguish mainly two groups of fields:

1. Often the fields related to $\Phi$ and $\Psi$ respectively $E_z$ and $H_z$ satisfy independently the boundary conditions. In this case we are therefore in the position to distinguish the following two types of modes:

   a) **TE- or H-modes**: Transverse electric (TE) modes do not have an electric field component in propagation direction. On the contrary they have a magnetic field component, therefore they are also called H-modes. They are given by

   $$\Phi = 0 \quad \text{(see (4.23a))} \quad \text{or} \quad E_z = 0. \quad (4.46a)$$

   b) **TM- or E-modes**: Transverse magnetic (TM) modes have a longitudinal electric field component. Therefore they are also denoted as E-modes. On the other hand, the magnetic field has only transverse components with respect to the propagation direction. They are given by

   $$\Psi = 0 \quad \text{(see (4.23b))} \quad \text{or} \quad H_z = 0. \quad (4.47a)$$

2. In some cases, the fields of the TE or TM mode alone cannot satisfy the boundary conditions, and the only possible mode is the hybrid electric and magnetic mode denoted by HEM mode.

Finally, we remark that there is a special kind of modes called **Longitudinal Section Electric (LSE)** and **Longitudinal Section Magnetic (LSM)** modes (see 4.3.14 B) on 88).

In this case we deal with fields that are transverse electric or transverse magnetic with respect to a certain direction which is not the propagation direction. For example in the case of a rectangular waveguide partially filled with dielectric slabs parallel to one waveguide side (e.g. the shielded printed circuits in Table 4.1) such kind of modes are used to represent the field. Often the label $\text{TE}^{(r)}$ and $\text{TM}^{(r)}$ is used for them to emphasize that they are transverse electric or transverse magnetic with respect to the direction $r$.

We will see in more detail these types of modes when we study the field distribution in rectangular waveguides.

C) **Slow waves**

When $T^2 < 0$, then $T$ is imaginary, and $\gamma = j\beta$, $\beta^2 = k^2 - T^2 > k^2$, $\beta$ is real and $\beta > k$. So

$$v_{\text{ph}} = \frac{\omega}{\beta} = \frac{\omega}{\sqrt{k^2 - T^2}} = \frac{c}{\sqrt{1 - T^2/k^2}} < c \quad (4.48)$$
and the phase velocity along $z$ is less than the phase velocity of a plane wave in the unbounded medium. So it is known as a slow wave.

(4.36a) is a special kind of Sturm-Liouville problem (see Appendix D). It can be shown that in systems with homogeneous boundary conditions, the eigenvalue of a Sturm-Liouville problem must not be negative. So fast-wave or TEM-wave systems must be surrounded by short-circuit (perfect electric conductor or electric wall) or open-circuit (perfect magnetic conductor or magnetic wall) boundaries. In fast-wave and TEM-wave systems, $T^2 \geq 0$ always holds.

On the contrary, in slow wave systems, $T^2 < 0$, which cannot be the eigenvalue of Sturm-Liouville problems with homogeneous boundary conditions. Thus a system surrounded by smooth short-circuit or open-circuit boundaries cannot support slow waves. The slow wave systems are constructed by means of dielectric boundaries or corrugated metallic boundaries. For a slow wave, the eigenvalue $T^2$ is no longer constant, so (4.43) and (4.44) are no longer valid. The group velocity of a slow wave is still less than or equal to the velocity of light in unbounded space.

4.3.8 Fields related to TEM-, TE- and TM-modes

In this section we provide specific relations between the transverse components of the electric and magnetic field for the case of time-harmonic fields and sinusoidal waves along $z$, that is

$$\mathcal{E} = E e^{j \omega t \pm \gamma z} \quad \text{and} \quad \mathcal{H} = H e^{j \omega t \pm \gamma z}.$$

A) TEM modes

Since in this case $E_z = H_z = 0$ we deduce from the equations on the right side of (E.5) that

$$\pm \gamma e_z \times \mathcal{H}_t = j \omega \varepsilon \mathcal{E}_t, \quad (4.49a)$$

$$\mp \gamma e_z \times \mathcal{E}_t = j \omega \mu \mathcal{H}_t. \quad (4.49b)$$

The lower sign is valid for a wave travelling in $+z$-direction, the upper one for a mode propagating towards $-z$.

Since $\gamma = j \omega / c$ we obtain

$$\mp e_z \times \mathcal{E}_t = Z_{F}^{(\text{TEM})} \mathcal{H}_t. \quad (4.50)$$

Introducing the direction of propagation $e_z$, we can rewrite the last equation in a compact way and independently of the coordinate system:

$$e_z \times \mathcal{E}_t = Z_{F}^{(\text{TEM})} \mathcal{H}_t, \quad (4.51)$$
because for a wave travelling towards positive values of \( z \), \( e_\gamma = +e_z \) holds and for the opposite case \( e_\gamma = -e_z \).

The quantity \( Z_F^{(TEM)} \) is the modal impedance or characteristic field impedance of a TEM wave and defined as

\[
Z_F^{(TEM)} = \sqrt{\frac{\mu}{\varepsilon}}. \tag{4.52}
\]

Hence the same as for a plane wave travelling in an unbounded medium.

The fields are given in (4.41) and repeated here for convenience for time-harmonic sinusoidal waves along the waveguide’s axis (z-direction):

\begin{align*}
E_t &= -j\omega Z_F^{(TEM)} \nabla_t \Phi, \tag{4.53a} \\
H_t &= -j\omega e_z \times \nabla_t \Phi = \frac{1}{Z_F^{(TEM)}} e_z \times E_t, \tag{4.53b} \\
\Phi &= \text{const on the boundary of } \partial S \left( \frac{\partial \Phi}{\partial n} = 0 \right). \tag{4.53c}
\end{align*}

The same result is obtained using \( \Psi \) instead of \( \Phi \) with \( \partial \Phi/\partial n = 0 \) on \( \partial S \).

B) TE modes

Using \( \Phi = 0 \) in the method of Hertz vectors

\[
e_\gamma \times E_t = Z_F^{(TE)} H_t, \tag{4.54}
\]

which is similar to the TEM case except for \( Z_F^{(TE)} \) that is given by

\[
Z_F^{(TE)} = \frac{j\omega \mu}{\gamma}. \tag{4.55}
\]

Using the Hertz vector \( \Pi_m = \Psi e_z (\Phi = 0) \) the fields can be expressed as follows:

\begin{align*}
H_t &= \pm \frac{\gamma}{\mu} \nabla_t \Psi, \tag{4.56a} \\
H_z &= \frac{T^2}{\mu} \Psi, \tag{4.56b} \\
\frac{\partial \Psi}{\partial n} &= 0 \quad \text{on } \partial S. 
\end{align*}
The lower sign is valid for propagation in \( +z \)-direction, the upper one for the opposite direction. The electric field can be computed by means of (4.54)

\[
\mathcal{E}_t = j\omega \varepsilon_z \times \text{grad}_t \Psi. \tag{4.56c}
\]

\section*{C) TM modes}

For the TM case \((\Psi = 0)\) we have analogously to TE modes

\[
\varepsilon_{\gamma} \times \mathcal{E}_t = Z_{F}^{(TM)} \mathcal{H}_t, \tag{4.57}
\]

with

\[
Z_{F}^{(TM)} = \frac{\gamma}{j\omega \varepsilon}. \tag{4.58}
\]

Similar to the TE case the fields are given by

\[
\mathcal{E}_t = \pm \frac{\gamma}{\varepsilon} \text{grad}_t \Phi, \tag{4.59a}
\]

\[
E_z = \frac{T^2}{\varepsilon} \Phi, \tag{4.59b}
\]

\[
\mathcal{H}_t = -j\omega \varepsilon_z \times \text{grad}_t \Phi, \tag{4.59c}
\]

\(\Phi = 0\) on \(\partial S\).

In conclusion, the transverse components of the electromagnetic field for TEM-, TE- and TM-modes are related to each other like for a plane wave in an unbounded medium: \((\varepsilon_{\gamma}, \mathcal{E}_t, \mathcal{H}_t)\) form a rectilinear system. The scaling factor between the electric and magnetic field depends on the type of mode and is \(Z_{F}^{(TEM)}\), \(Z_{F}^{(TE)}\) or \(Z_{F}^{(TM)}\), respectively.
<table>
<thead>
<tr>
<th></th>
<th>fast waves</th>
<th>slow waves</th>
</tr>
</thead>
<tbody>
<tr>
<td>eigen value</td>
<td>$T^2 = 0$</td>
<td>$T^2 &gt; 0$</td>
</tr>
<tr>
<td>propagation constant</td>
<td>$\gamma = j\beta = j\omega$</td>
<td>$\gamma = \sqrt{T^2 - k^2}$, $\beta &lt; k$</td>
</tr>
<tr>
<td>phase velocity $v_{ph} = \frac{\omega}{\beta}$</td>
<td>$c$</td>
<td>$\frac{c}{\sqrt{1-T^2/k^2}} &gt; c$</td>
</tr>
<tr>
<td>group velocity $v_{gr} = \frac{1}{\partial \beta / \partial \omega}$</td>
<td>$c$</td>
<td>$c\sqrt{1-T^2/k^2} &lt; c$</td>
</tr>
<tr>
<td>mode types</td>
<td>TEM</td>
<td>TE</td>
</tr>
<tr>
<td>Hertz Vector</td>
<td>magnetic ($\Pi_m$) or electric ($\Pi_e$)</td>
<td>magnetic ($\Pi_m$)</td>
</tr>
<tr>
<td>longitudinal fields</td>
<td>$E_z = H_z = 0$ ($\Phi = 0$ or $\Psi = 0$)</td>
<td>$E_z = 0$ ($\Phi = 0$)</td>
</tr>
<tr>
<td>static problem</td>
<td>Laplace equation</td>
<td>Helmholtz equation</td>
</tr>
<tr>
<td></td>
<td>$\triangle_t \Phi = 0$</td>
<td>$\triangle_t \Psi + T^2 \Phi = 0$</td>
</tr>
<tr>
<td>boundary conditions</td>
<td>$\frac{\partial \Phi}{\partial t} = 0$ or $\frac{\partial \Psi}{\partial n} = 0$</td>
<td>$\frac{\partial \Psi}{\partial n} = 0$</td>
</tr>
<tr>
<td>transverse fields</td>
<td>$\epsilon \times \mathcal{E}_t = Z_F^{(i)} \mathcal{H}_t$</td>
<td></td>
</tr>
<tr>
<td>characteristic (field) impedance, modal impedance</td>
<td>$Z_F^{(i)} = Z_F^{(TEM)} = Z_0$</td>
<td>$Z_F^{(i)} = Z_F^{(TE)} = \frac{\eta \mu}{\gamma} = \frac{j k Z_0}{\gamma}$</td>
</tr>
<tr>
<td>($Z_0 = \sqrt{\frac{\mu}{\varepsilon}}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sinusoidal waves</td>
<td>$\mathcal{E}_t = j \omega Z_F^{(TEM)} \text{grad}_t \Phi$</td>
<td>$\mathcal{E}_t = j \omega \epsilon \times \text{grad}_t \Psi$</td>
</tr>
<tr>
<td></td>
<td>$\mathcal{H}_t = \frac{1}{Z_F^{(TEM)}} \epsilon \times \mathcal{E}_t$</td>
<td>$\mathcal{H}_t = \pm \frac{\mu}{\epsilon} \text{grad}_t \Psi$</td>
</tr>
<tr>
<td></td>
<td>$H_z = \frac{T^2}{\mu} \Psi$</td>
<td></td>
</tr>
</tbody>
</table>
4.3.9 Discussion of the three types of waveguide modes

In the last section, we have seen that the electromagnetic field distribution in uniform waveguides with homogeneous boundary conditions can in principle be classified into three types: TE-, TM-, and TEM-modes. Using these results, we will discuss in this section several properties of uniform waveguides and their modes.

Since we deal with a problem of potential theory in the case of TEM-modes, the electromagnetic field distribution in the transverse plane is irrotational \((\text{rot}_t = 0)\) and solenoidal \((\text{div}_t = 0)\). Hence, the transverse field is the static field of a cylindrical capacitor. This, however, implies that TEM-modes can only exist if the waveguide consists at least of two separate conductors. To see this, let us consider the different waveguide configurations of Figure 4.9.

![Figure 4.9](image)

**Figure 4.9** Three examples for different waveguide sections with only one (a), two (b), and three (c) separate conductors.

In the case (a) (hollow waveguide), no static transverse electric field can exist and therefore no TEM-mode. We only can have a constant potential \(\phi\) over the entire cross section, thus the electric field, which is the gradient of the potential, is zero everywhere inside the waveguide.

In the case (b), the inner and outer conductor can have different potentials \(\phi_1\) and \(\phi_2\), with \(\phi_1 \neq \phi_2\), thus an electrostatic field can exist. (b) is the typical case of a coaxial cable.

In the case (c), we have three separate conductors. (c) represents the case of coupled lines that are shielded. Whereas in the configuration (b) one TEM-mode is present, the structure in Figure (c) can carry two TEM-modes. In general, a waveguide consisting of \(k\) separate conductors has \(k - 1\) TEM-modes.

The two field configurations for the shielded lines in Figure (c) are shown in Figure 4.10. In Figure 4.10(a) both inner conductors have the same potential \(\phi_2 = U\) and the outer conductor is connected to the ground, i.e. \(\phi_1 = 0\). Figure 4.10(b) shows a different field configuration associated to the case where the inner conductors are now on different potentials, that is, \(\phi_3 = -\phi_2 = U\). Both modes have the same eigenvalue \(T = 0\). This means, that we deal with a degenerated case. The two field configurations are linearly independent.

Whereas for TEM-modes at least two separate conductors are necessary, TE- and TM-modes can also exist and propagate in waveguides consisting of only
Figure 4.10 Field distribution of the transverse electric field on shielded coupled lines.

One conductor as in Figure 4.9(a). However, they also exist in the case of TEM waveguides. This means, TE- and TM-modes are always present, whereas TEM-modes only in configurations of two or more conductors.

The field distributions of TE- and TM-modes can be easily found according to the following rule:

1. **TM-modes**: For TM-modes, the magnetic field lines are closed, plane lines that lie in the transverse plane. The lines of the electric field commence on the surface of the electric conductors, are perpendicular to it, and continue inside the waveguide parallel to the $z$-direction.

2. **TE-modes**: In contrast, for TE-modes the lines of the electric field lie completely in the transverse plane. The magnetic lines, on the other hand, are closed lines that can have both transverse and longitudinal components.

### 4.3.10 Eigenfunctions and eigenvalues

As shown in the previous sections, the problem of finding the electromagnetic field distribution in uniform waveguides is reduced to that of finding a function $\Psi(u, v)$ (TE case) or $\Phi(u, v)$ (TM case) and the constant $T$ satisfying the Helmholtz equation or, in the TEM case, Laplace’s equation with $T = 0$. Furthermore, the functions are subject to certain boundary conditions.

The function $\Phi$ or $\Psi$ satisfying (4.36a) ($T = 0$ is included as special case) is called *eigenfunction* or *eigensolution* and the constant $T$ is the corresponding *eigenvalue*. Thus, a mode is an eigenfunction or eigensolution, i.e., a specific field distribution that is maintained without excitation.

The problem of solving for electromagnetic wave propagation in an uniform waveguide with a perfectly conducting wall is therefore reduced to the problem of finding eigenfunctions and eigenvalues for the 2D Dirichlet ($\Phi = 0$ on the
boundary) or Neumann ($\partial \Psi / \partial n = 0$ on the boundary) problem, and once $T$ is obtained, the propagation constant $\gamma$ is given by (4.35).

In general, to satisfy (4.36a), the eigenvalue $T$ cannot be an arbitrary number, and it can take on only specific values. There is an infinite number of discrete eigenvalues.

From a mathematical point of view, the problem of solving Helmholtz’s or Laplace’s equation with given boundary conditions is a Sturm-Liouville problem (see Appendix D), which is a well-known mathematical problem with a large number of theorems for associated eigensolutions and eigenvalues.

### 4.3.11 Vector eigenfunctions and mode expansion

The vector eigenfunction set of the time-varying boundary-value problem forms a complete orthogonal set. Again we will concentrate on time-harmonic sinusoidal waves along the axis of the waveguide.

In a cylindrical system, $(u, v, z)$, suppose that the 2D vector eigenfunctions of two arbitrary modes are $E_m(u, v, z, t)$, $H_m(u, v, z, t)$ and $E_n(u, v, z, t)$, $H_n(u, v, z, t)$, the associated transverse 2D eigenvalues $T_m$, $T_n$. For time-harmonic sinusoidal waves they can be expressed as

$$E_n(u, v, z, t) = E_n(u, v) e^{j\omega t \pm j\gamma z}, \quad (4.60)$$

$$H_n(u, v, z, t) = H_n(u, v) e^{j\omega t \pm j\gamma z}. \quad (4.61)$$

The orthogonality of these two sets of vector eigenfunctions is given as

$$\int_S E^*_m(u, v) \times H_n(u, v) \, du \, dv = 0, \quad m \neq n, \quad (4.62)$$

where $S$ denotes an arbitrary cross-section of the uniform waveguide. Note that the integral is only over the cross-section of the waveguide. Thus, the modes $E_m$, $H_m$ and $E_n$, $H_n$ are orthogonal with respect to the cross-section of the uniform waveguide. The physical meaning of this orthogonality relation is that the electric field and magnetic field of two different modes do not carry any power flow. Hence, the total power flow of a multimode system is equal to the sum of the power flows of the individual modes. These modes are known as normal modes.

Remark 1: This orthogonality property is only valid if no losses in the waveguide’s wall are present, i.e. if we deal with perfect conductors, and if the modes are non-degenerate. In the case of ohmic losses, the modes become coupled and the total power flow is not anymore the sum of the individual power flows. When degenerate modes do not satisfy a power-orthogonality relation, then it is possible to define new modes which are linear combinations of the old ones and for which that power orthogonality holds.

$^3$normal in the sense of “non-degenerate”
Remark 2: A similar orthogonality condition holds for non-time-harmonic and non-sinusoidal waves along \( z \). In fact, the orthogonality with respect to the cross-section follows directly from the separation of variables. Splitting the dependency of the fields into a product of two functions where one depends only on the transverse variables \((u,v)\) and the second one on the longitudinal coordinate \( z \) and the time \( t \) implies that the eigenfunctions related to the transverse wave equation are orthogonal independently of \( z \) and \( t \).

However, for our purposes we may assume time-harmonic sinusoidal waves for the remainder.

Since the eigenfunctions (modes) are orthogonal, any fields over the cross section of a cylindrical system can thus be expanded into a series of vector eigenfunctions or modes

\[
\mathcal{E} = \sum_{n=1}^{\infty} A_n \mathcal{E}_n, \quad \mathcal{H} = \sum_{n=1}^{\infty} B_n \mathcal{H}_n. \tag{4.63a}
\]

The coefficients of the series may be obtained by the orthogonality principle

\[
A_n = \frac{\int_S \epsilon_z^r (\mathcal{E} \times \mathcal{H}_n^*) \, dS}{\int_S \epsilon_z^r (\mathcal{E}_n^* \times \mathcal{H}_n) \, dS}, \quad B_n = \frac{\int_S \epsilon_z^r (\mathcal{E}_n^* \times \mathcal{H}^*) \, dS}{\int_S \epsilon_z^r (\mathcal{E}_n^* \times \mathcal{H}_n) \, dS}. \tag{4.63b}
\]

We come to the conclusion that the solutions of the Helmholtz equations that satisfy specific boundary conditions form a complete set of an infinite number of normal modes. Depending on the frequency, a finite number of modes are propagating modes and the rest are cutoff or evanescent modes.

The orthogonality of the 3D vector eigenfunctions can be proven and any fields in a closed region can be expanded into a series of the 3D orthonormal vector eigenfunctions.

### 4.3.12 Normalized modes

The orthogonality condition (4.62) allows us to introduce normalized vector eigenfunctions, i.e. normalized modes. There are various ways to normalize them. All that we have to respect when normalizing them are the relations between the electric and the magnetic field and between transverse and longitudinal components as in Section 4.3.8. One possibility, but by far not the only one, is to define electric and magnetic Hertz vectors such that the normalized modes satisfy the following conditions:
\[ e_n(u, v, z, t) = e_n(u, v)e^{j\omega t \pm \gamma z}, \quad (4.64) \]
\[ h_n(u, v, z, t) = h_n(u, v)e^{j\omega t \pm \gamma z}, \quad (4.65) \]
\[ e_\gamma \times e_n(u, v) = h_t_n(u, v), \quad (4.66) \]
\[ e_t_n(u, v), h_t_n(u, v) \in \mathbb{R}^2. \quad (4.67) \]

Since any eigenfunction is unique up to a multiplying constant, i.e. its amplitude, we are always in the position to choose Hertzian vectors such that these relations hold.

The first two equations are the analogue to (4.60) and the third one reflects the relation (4.51), (4.54) or (4.57) for any of the mode types TEM, TE and TM.

Using (4.66), the power orthogonality (4.62) reduces to
\[
\int_S e_{m}^\text{tr} [e_m(u, v) \times h_n^*(u, v)]\, dS = \int_S e_{t_m}^\text{tr} e_{t_n}^* dS = \int_S h_{t_m}^* h_{t_n}^* dS = \delta_{m,n}. \quad (4.68)\]

From (4.66) it becomes clear, that the transverse electric field \( e_{t_n} \) serves as our reference. With the same right we could have chosen the transverse magnetic field \( h_{t_n} \) or the longitudinal components \( e_z \) and \( h_z \) or the Hertzian vectors \( \Phi \) and \( \Psi \). However, since the orthogonality of the modes is based on the transverse components of the electric and magnetic field we used the transverse field. Without loss of generality, we decided to use the electric field as basis.

\textit{A) TM modes}

To normalize the TM modes, we can introduce the electric Hertz vector \( \Pi_e = \Phi e_z \) with
\[
\Phi = \pm \frac{\varepsilon}{\gamma} \phi(u, v)e^{j\omega t \pm \gamma z} \quad (4.69a)\]
which yields
\[
\mathcal{E}_t = \text{grad}_t \phi e^{j\omega t \pm \gamma z}, \quad (4.69b)\]
\[
E_z = \pm \frac{T^2}{\gamma} \phi e^{j\omega t \pm \gamma z}, \quad (4.69c)\]
\[
\mathcal{H}_\ell = \mp \frac{1}{Z_{\text{TM}}} e_z \times \text{grad}_t \phi e^{j\omega t \pm \gamma z}. \quad (4.69d)\]

In order to normalize \( \mathcal{E}_t \) we make use of (4.68) which yields a condition for
\( \phi(u, v) : \)

\[
\int_S \phi^2(u, v) \, dS = \frac{1}{T^2} \quad . \tag{4.69e}
\]

With this, the normalized modes are

\[
e_n(u, v, z, t) = \left( e_n(u, v) \pm \frac{T^2}{\gamma_n} \phi_n e_z \right) e^{j\omega t \pm \gamma_n z} , \tag{4.69f}
\]

\[
h_n(u, v, z, t) = \mp h_n e^{j\omega t \pm \gamma_n z} , \tag{4.69g}
\]

with the phasors

\[
e_t = \text{grad}_t \phi_n(u, v) \quad \text{and} \quad h_t = e_z \times e_t . \tag{4.69h}
\]

B) TE modes

In a similar way we can proceed with the TE modes by using the magnetic Hertz vector \( \Psi_m = \Psi e_z \) with

\[
\Psi = \frac{\mu}{\gamma} \psi(u, v) e^{j\omega t \pm \gamma z} , \tag{4.70a}
\]

where \( \psi \) satisfies the same normalization condition (4.69e) as \( \phi \) in the case of TM modes.

This yields the following field:

\[
E_t = Z_{PE}^{(TE)} e_z \times \text{grad}_t \psi e^{j\omega t \pm \gamma z} , \tag{4.70b}
\]

\[
H_t = \pm \text{grad}_t \psi e^{j\omega t \pm \gamma z} , \tag{4.70c}
\]

\[
H_z = \frac{T^2}{\gamma} \psi e^{j\omega t \pm \gamma z} . \tag{4.70d}
\]

The normalized modes are

\[
e_n(u, v, z, t) = e_n e^{j\omega t \pm \gamma_n z} , \tag{4.70e}
\]

\[
h_n(u, v, z, t) = \left( \mp h_n(u, v) + \frac{T^2}{\gamma_n} \psi_n e_z \right) e^{j\omega t \pm \gamma_n z} , \tag{4.70f}
\]

and the phasors
\begin{align*}
\phi_n(u, v) & \quad \psi_n(u, v) & \quad \phi_n(u, v) \\
\triangle \phi_n + T_n^2 \phi_n = 0 & \quad \triangle \psi_n + T_n^2 \psi_n = 0 & \quad \triangle \phi_n = 0, T_n = 0 \\
||\phi_n||^2 = \frac{1}{T^2} & \quad ||\psi_n||^2 = \frac{1}{T^2} & \\
\phi_n = 0 & \quad \partial \psi_n / \partial n = 0 & \quad \partial \phi_n / \partial t = 0 \\
e_t = \varepsilon_z \times h_t & \quad e_t = \varepsilon_t \times e_z & \quad e_t = \text{grad}_t \phi_n \\
e_z = \frac{T_n^2}{\gamma_n} \phi_n e_z & \quad e_z = 0 & \quad e_z = 0 \\
h_t = \varepsilon_z \times e_t & \quad h_t = -\text{grad}_t \psi_n & \quad h_t = \varepsilon_z \times e_t \\
h_z = 0 & \quad h_z = \frac{T_n^2}{\gamma_n} \psi_n e_z & \quad h_z = 0 \\
e_n = (e_t \pm e_z) e^{j\omega t \pm \gamma_n z} & \\
h_n = (\mp h_t + h_z) e^{j\omega t \pm \gamma_n z} & 
\end{align*}

**Table 4.2** Normalized eigenfunctions (modes) of TE, TM and TEM modes and relations.

\begin{align*}
e_t = \varepsilon_z \times h_t & \quad \text{and} \quad h_t = -\text{grad}_t \psi_n(u, v). \quad (4.70h)
\end{align*}

Note the minus sign in front of the gradient of the vector phasor $h_t$. This is necessary to achieve a consistent sign definition for $H_t$ with respect to TM-modes.

**C) TEM modes**

TEM modes are easily normalized using the Hertz vector $\Pi_e = \Phi \varepsilon_z$ with

\begin{align*}
\Phi = \frac{1}{j\omega} \frac{1}{Z_F^{(\text{TEM})}} \phi(u, v) e^{j\omega t \pm \gamma_n z}. \quad (4.71a)
\end{align*}

$\phi$ satisfies in this case the Laplace equation and the phasors are given by

\begin{align*}
e_t = \text{grad}_t \phi(u, v) & \quad \text{and} \quad h_t = \varepsilon_z \times e_t. \quad (4.71b)
\end{align*}

A summary of normalized eigenfunctions and the relations between transverse and longitudinal components is provided in Table 4.2.
4.3.13 Field representation in uniform waveguides: modal expansion

We can use the normalized modes of Section 4.3.12 to represent the electromagnetic field in uniform waveguides since they define a complete set of functions. Each mode represents a specific field distribution, i.e., a characteristic field, and propagates independently of all the other ones (power orthogonality (4.62) respectively (4.68) between the modes). Hence the total field is the sum over all modes weighted with a multiplicative constant. The field representation by means of modes is as follows:

\begin{align}
\mathcal{E}_t &= \sum_{n=1}^{\infty} \left( U_n^{(+)\,} e^{-\gamma_n z} + U_n^{(-)\,} e^{+\gamma_n z} \right) e^{j\omega t} e_{t_n}(u,v), \quad (4.72a) \\
\mathcal{H}_t &= \sum_{n=1}^{\infty} \left( I_n^{(+)\,} e^{-\gamma_n z} - I_n^{(-)\,} e^{+\gamma_n z} \right) e^{j\omega t} h_{t_n}(u,v), \quad (4.72b) \\
E_z &= - \sum_{n=1}^{\infty} \frac{T_n^2}{\gamma_n} \left( U_n^{(+)\,} e^{-\gamma_n z} - U_n^{(-)\,} e^{+\gamma_n z} \right) e^{j\omega t} \phi_{t_n}(u,v), \quad (4.72c) \\
H_z &= \sum_{n=1}^{\infty} \frac{T_n^2}{\gamma_n} \left( I_n^{(+)\,} e^{-\gamma_n z} + I_n^{(-)\,} e^{+\gamma_n z} \right) e^{j\omega t} \psi_{t_n}(u,v), \quad (4.72d) \\
U_n^{(\pm)} &= Z_{F_n} I_n^{(\pm)} , \quad Z_{F_n} = Z_F^{(TEM)} \text{ or } Z_F^{(TE)} \text{ or } Z_F^{(TM)} . \quad (4.72e)
\end{align}

$U_n^{(+)\,}, I_n^{(+)\,}$ represent a wave travelling in the positive $z$-direction and $U_n^{(-)\,}, I_n^{(-)\,}$ waves propagating in the opposite one.

$U$ and $I$ are called modal voltage and modal current, respectively. Their units are V and A, respectively. The unit of \( e_n \) and \( h_n \) is $1/m$ due to (4.68).

We used two different multiplicative constants for the electric and the magnetic field ($U$ respectively $I$) to underline that the electric field goes with the voltage and the magnetic field is associated with the electric current. In reality, they are related to each other which is expressed by the last equation (4.72e): This relation states the coupling between the electric and magnetic field as imposed by Maxwell’s equations. But it means also, that in an uniform, infinite waveguide the waves travelling in opposite directions are independent from each other, therefore the superscripts “+” and “-”.

Moreover, as seen at the beginning of this chapter, knowing the longitudinal components of the electric and the magnetic field, $E_z$ and $H_z$, we can find the transverse field distribution and vice versa. This is reflected by the above expansion: The same expansion coefficients $U$ and $I$ as used for the transverse fields reappear in the longitudinal components.
A last word related to the naming of the expansion constants $U$ and $I$: In non-TEM waveguides a voltage in the strict sense cannot be defined. For this reason the correct naming of $U$ and $I$ is modal voltage respectively modal current. $U_n$ and $I_n$ represent in fact the magnitude of the electric and magnetic field associated with the mode $n$. Hence they represent field quantities. In the case of TEM waveguides, however, it can be shown that there is a direct relation between the modal voltage/current representing the electromagnetic field and the classical definition of voltage and current as in circuit theory where voltage represents a difference of potential what, when closing the circuit, results in an electric current flow.

Finally, we can rewrite this expansion in a more compact way by introducing the functions

$$U_n(z) = U_n^{(+)} e^{-\gamma_n z} + U_n^{(-)} e^{+\gamma_n z}, \quad U_n(z, t) = U_n(z) e^{j\omega t}, \quad (4.73a)$$

$$I_n(z) = I_n^{(+)} e^{-\gamma_n z} - I_n^{(-)} e^{+\gamma_n z}, \quad I_n(z, t) = I_n(z) e^{j\omega t} \quad (4.73b)$$
as

$$E_t = \sum_{n=1}^{\infty} U_n(z, t) e_{t_n}(u, v), \quad (4.74a)$$

$$H_t = \sum_{n=1}^{\infty} I_n(z, t) h_{t_n}(u, v), \quad (4.74b)$$

$$E_z = -\sum_{n=1}^{\infty} \frac{T_n^2 Z_0}{jk} I_n(z, t) \phi_n(u, v), \quad (4.74c)$$

$$H_z = \sum_{n=1}^{\infty} \frac{T_n^2}{jk Z_0} U_n(z, t) \psi_n(u, v), \quad (4.74d)$$

where we have used (4.72e). Note that this expansion reflects the separation of variables introduced to solve Maxwell’s equations for cylindrical systems. Note also that $U_n^{(\pm)} = Z_{F_n} I_n^{(\pm)}$ holds but not $U_n(z) = Z_{F_n} I_n(z)$, i.e. the relation between modal voltage (transverse electric field) and modal current (transverse magnetic field) is only valid for waves propagating in one direction ($+z$ or $-z$) but not for their superposition.

### 4.3.14 Eigensolutions of Maxwell’s equations in rectangular waveguides

Let us consider the rectangular waveguide of Figure 4.11 with sides $a$ and $b$. 
Figure 4.11 Rectangular waveguide. Coordinate system and dimensions used to calculate its modes.

A) TE and TM modes

As outlined in Section 4.3.4 we can represent the field distribution in a cylindrical waveguide by means of an electric and a magnetic Hertz vector parallel to the waveguide’s axis. In Section 4.3.7 we then saw that TM modes can be associated with the electric Hertz vector $\Pi_e (\Psi = 0)$ and TE modes with the magnetic Hertz vector $\Pi_m (\Phi = 0)$.

**TM modes** Let us first consider TM modes in the rectangular waveguide of Figure 4.11. Using the method of separation of variables and assuming a time-harmonic dependency $e^{j\omega t}$ the electric Hertz vector (4.69a) is given by the eigenfunction

$$\phi_{\mu,\nu} = \frac{2}{\sqrt{ab}} q_{\mu,\nu} \sin \frac{\mu\pi x}{a} \sin \frac{\nu\pi y}{b}, \quad (4.75a)$$

and the eigenvalue

$$T^2 = q_{\mu,\nu}^2 = \left(\frac{\mu\pi}{a}\right)^2 + \left(\frac{\nu\pi}{b}\right)^2, \quad (4.76)$$

where $\mu = 1, 2, 3, \ldots$ and $\nu = 1, 2, 3, \ldots$. The choice of the amplitude $\frac{2}{\sqrt{ab}} q_{\mu,\nu}$ assures that (4.69e) is satisfied and in turn that the vector eigenfunctions (modes) $e_{\mu,\nu}$ and $h_{\mu,\nu}$ are normalized.

This mode is called the TM$_{\mu,\nu}$ mode or the E$_{\mu,\nu}$ mode. The propagation constant is given by

$$\gamma_{\mu,\nu} = \sqrt{q_{\mu,\nu}^2 - k^2}. \quad (4.77)$$

The normalized eigenfunctions (modes) can be calculated by means of (4.69f)-
(4.69h):

\[ \mathbf{e}_{t,\mu,\nu} = \text{grad}_t \phi_{\mu,\nu}, \quad (4.78) \]
\[ \mathbf{e}_{z,\mu,\nu} = \frac{q_{\mu,\nu}^2}{\gamma_{\mu,\nu}} \phi_{\mu,\nu} \mathbf{e}_z, \quad (4.79) \]
\[ \mathbf{h}_{t,\mu,\nu} = \mathbf{e}_z \times \text{grad}_t \phi_{\mu,\nu}. \quad (4.80) \]

**TE modes** In a similar manner, TE modes can be defined by the magnetic Hertz vector (4.70a) and the eigenfunction

\[ \psi_{\mu,\nu} = \sqrt{\frac{\varepsilon_{\mu} \varepsilon_{\nu}}{ab}} \cos \frac{\mu \pi x}{a} \cos \frac{\nu \pi y}{b}, \quad (4.81a) \]

where \( \mu = 0, 1, 2, \ldots \) and \( \nu = 0, 1, 2, \ldots \), but \( \mu + \nu \neq 0 \), i.e. \( \mu = \nu = 0 \) is excluded since it would lead to a zero field. The constant \( \varepsilon_{\mu} \) is given by

\[ \varepsilon_{\mu} = \begin{cases} 1 & \text{for } \mu = 0, \\ 2 & \text{otherwise}. \end{cases} \quad (4.82) \]

Again, the amplitude of \( \psi_{\mu,\nu} \) guarantees that (4.69e) is fulfilled and the corresponding modes are normalized.

The eigenvalue \( q_{\mu,\nu}^2 \) is the same as for the TM modes, that is (4.76). The modes are called TE\( \mu,\nu \) or H\( \mu,\nu \) modes and the corresponding normalized eigenfunctions are

\[ \mathbf{e}_{t,\mu,\nu} = \mathbf{e}_z \times \mathbf{h}_{t,\mu,\nu}, \quad (4.83) \]
\[ \mathbf{h}_{t,\mu,\nu} = -\text{grad}_t \psi_{\mu,\nu}, \quad (4.84) \]
\[ \mathbf{h}_{z,\mu,\nu} = \frac{q_{\mu,\nu}^2}{\gamma_{\mu,\nu}} \psi_{\mu,\nu} \mathbf{e}_z. \quad (4.85) \]

It can be shown that any TE\( \mu,\nu \) mode is orthogonal to any TM\( \mu',\nu' \) mode, that is TE and TM modes are mutually orthogonal. Since a TE\( m,n \) mode has the same eigenvalue as a TM\( m,n \) mode, we deal with a degenerate case.

An important special case is the mode with the smallest eigenvalue \( T \). In the case of a rectangular waveguide with \( a > b \) this is \( q_{1,0} \). The mode TE\( 1,0 \) is called the fundamental mode, since it has the lowest cutoff wavenumber \( k_c \) with

\[ k_{c,1,0} = \frac{\pi}{a}. \quad (4.86) \]

It is the first mode that changes from the state under cutoff \( (k < k_c) \) to the state above cutoff \( (k > k_c) \), i.e. it propagates before any other mode. This mode is used for most microwave waveguides if rectangular waveguides are used.
order of modes in terms of ascending cutoff wavenumber \( k_{c,\mu,\nu} = q_{\mu,\nu} \) depends on the ratio \( b/a \) of the rectangular cross-section.

A summary of TE and TM modes in a rectangular waveguide is given in Table 4.3.

B) LSE, LSM Modes

The metallic rectangular waveguide discussed in the previous section was filled with an uniform medium, in which any single TE- or TM-mode can exist independently. When a waveguide is filled with different media or partially filled with a dielectric medium, the electric and magnetic fields must satisfy, additionally to the boundary conditions on the waveguide’s wall, the boundary conditions on the interface between the different media.

Consider the rectangular, uniform waveguides of Figure 4.12 which are partially filled with a dielectric slab parallel to the \( yz \)-plane respectively \( xz \)-plane. We again assume that we can separate the problem into a problem depending only on the transverse coordinates \( x \) and \( y \) and a longitudinal problem which depends on \( z \). In other words, we again search for time-harmonic waves travelling along the \( z \) direction.

\[
E = E(x,y) e^{\omega \pm \gamma z} \quad \text{and} \quad H = H(x,y) e^{\omega \pm \gamma z}.
\]

In order to satisfy the boundary conditions at the interface between the two media it can be shown that only TE\(_{\mu,0}\) modes for the structure in Figure 4.12(a) are able to satisfy the additional interface boundary conditions. For the electric field, this means that it has no \( x \)-component \( (E_x = 0) \). In the case of Figure 4.12(b) only TE\(_{0,\nu}\) modes fulfil them, i.e. \( E_y = 0 \). This means that only TE modes having no component normal to the interface of the two media can exist in the waveguide with two different filling media and that the fields of the other TE- and TM-modes alone cannot satisfy the boundary conditions. Thus, the TE- and TM-modes do not form anymore a complete set of functions that allows to expand the electromagnetic field in the waveguide. We come to the conclusion
### 4.3 Uniform waveguides

#### Coordinate system for rectangular waveguide.

- **Transverse wave number:** $q_{\mu,\nu} = \sqrt{\left(\frac{\mu}{a}\right)^2 + \left(\frac{\nu}{b}\right)^2}$

#### TE modes:
- $\mu = 0, 1, 2, \ldots$, $\nu = 0, 1, 2, \ldots$, $\mu + \nu \neq 0$

#### TM modes:
- $\mu = 1, 2, 3, \ldots$, $\nu = 1, 2, 3, \ldots$

### Table 4.3 The rectangular waveguide and its TE and TM modes.

<table>
<thead>
<tr>
<th>$q_{\mu,\nu}$</th>
<th>$\psi_{\mu,\nu}$</th>
<th>$\phi_{\mu,\nu}$</th>
<th>$\varepsilon_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE modes: $q_{1,0}$</td>
<td>$q_{1,0}^{(TE)} = \frac{\pi}{a}$</td>
<td>$e_{1,0}^{(TE)} = -\varepsilon_y \sqrt{\frac{2ab}{\pi}} \sin \frac{\pi y}{a}$</td>
<td>1 for $i = 0$, 2 for $i \neq 0$</td>
</tr>
<tr>
<td>TE modes: $q_{2,0}$</td>
<td>$q_{2,0}^{(TE)} = \frac{2\pi}{a}$</td>
<td>$e_{2,0}^{(TE)} = -\varepsilon_y \sqrt{\frac{2ab}{2\pi}} \sin \frac{\pi y}{a}$</td>
<td></td>
</tr>
<tr>
<td>TM modes: $q_{0,1}$</td>
<td>$q_{0,1}^{(TE)} = \frac{\pi}{a}$</td>
<td>$e_{0,1}^{(TE)} = \varepsilon_x \sqrt{\frac{2ab}{\pi}} \sin \frac{\pi y}{b}$</td>
<td></td>
</tr>
<tr>
<td>TE modes: $q_{1,1}$</td>
<td>$q_{1,1}^{(TE)} = \sqrt{\frac{1}{a^2} + \frac{1}{b^2}}$</td>
<td>$e_{1,1}^{(TE)} = \varepsilon_x \sqrt{\frac{4a}{b(a^2 + b^2)}} \cos \frac{\pi y}{a} \sin \frac{\pi y}{b}$</td>
<td></td>
</tr>
<tr>
<td>TM modes: $q_{1,1}$</td>
<td>$q_{1,1}^{(TM)} = \sqrt{\frac{1}{a^2} + \frac{1}{b^2}}$</td>
<td>$e_{1,1}^{(TM)} = \varepsilon_x \sqrt{\frac{4b}{a(a^2 + b^2)}} \sin \frac{\pi y}{a} \cos \frac{\pi y}{b}$</td>
<td></td>
</tr>
</tbody>
</table>
that we need hybrid modes (HEM) to satisfy the additional boundary condition at the interface, i.e. modes that have a longitudinal electric field component $E_z$ as well as a longitudinal magnetic field component $H_z$.

In Section 4.3.7 we mentioned that besides TE-, TM-, and TEM-modes there are other types of modes called Longitudinal Section Electric (LSE) and Longitudinal Section Magnetic (LSM) modes. These type of modes are to be used if we deal with inhomogeneously filled waveguides. More precisely, for the problem of Figure 4.12(a) we need modes for which the electric or magnetic field normal to the interface vanishes. These modes are called LSE$(x)$ respectively LSM$(x)$ modes. An alternative notation for LSE$(x)$ and LSM$(x)$ is also TE$(x)$ and TM$(x)$ analogously to the conventional TE and TM modes. Similar, for the structure of Figure 4.12(b) we search for LSE$(y)$ and LSM$(y)$ modes respectively TE$(y)$ and TM$(y)$ modes. From this viewpoint, the ordinary TE and TM modes are actually TE$(z)$ and TM$(z)$ modes.

This analogy between TE$(x)$, TM$(x)$ and TE$(z)$, TM$(z)$ modes respectively between TE$(y)$, TM$(y)$ and TE$(z)$, TM$(z)$ modes can be further exploited to compute the LSE$(r)$ and LSM$(r)$ modes $(r = x, y)$. To compute the conventional TE and TM modes we used electric and magnetic Hertz vectors directed along the waveguides axis, i.e. $z$. In an arbitrary cylindrical system, the $z$-coordinate is a Cartesian coordinate. For rectangular waveguides the coordinates $x$ and $y$ are as well Cartesian coordinates. Hence we can compute the LSE$(x)$ (TE$(x)$) and LSM$(x)$

<table>
<thead>
<tr>
<th></th>
<th>LSE$(x)$</th>
<th>LSM$(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hertz vector</td>
<td>$\Pi_m = \Psi e_x$</td>
<td>$\Pi_e = \Phi e_x$</td>
</tr>
<tr>
<td>$\Psi = \psi_{m,n}(x,y)e^{j\omega t + \gamma_{m,n}z}$</td>
<td>$\Phi = \phi_{m,n}(x,y)e^{j\omega t + \gamma_{m,n}z}$</td>
<td></td>
</tr>
<tr>
<td>$\psi_{m,n} = A_{m,n} \sin \frac{m\pi x}{a} \cos \frac{n\pi y}{b}$</td>
<td>$\phi_{m,n} = A_{m,n} \cos \frac{m\pi x}{a} \sin \frac{n\pi y}{b}$</td>
<td></td>
</tr>
<tr>
<td>Boundary conditions</td>
<td>$\Psi = 0$ for $x = 0, a$</td>
<td>$\partial \Phi / \partial x = 0$ for $x = 0, a$</td>
</tr>
<tr>
<td></td>
<td>$\partial \Psi / \partial y = 0$ for $y = 0, b$</td>
<td>$\Phi = 0$ for $y = 0, b$</td>
</tr>
<tr>
<td>Electric field</td>
<td>$\mathcal{E}_t = E_y \varepsilon_y + E_z \varepsilon_z$, $\mathcal{H}_t = H_y \varepsilon_y + H_z \varepsilon_z$</td>
<td>$\mathcal{J}^{(\text{LSE}(x))} = \frac{1}{\varepsilon} \text{grad}_t \frac{\partial \Phi}{\partial x}$ $\mathcal{J}^{(\text{LSM}(x))} = \frac{T^2}{\varepsilon} \Phi$</td>
</tr>
<tr>
<td>$E_x$</td>
<td>$\frac{1}{T} \text{grad}_t \frac{\partial \Psi}{\partial x}$</td>
<td>$-j\omega \varepsilon_x \times \text{grad}_t \Phi$</td>
</tr>
<tr>
<td>$H_t$</td>
<td>$T^2 \mu \Psi$</td>
<td>-</td>
</tr>
<tr>
<td>$H_x$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>modal impedance</td>
<td>$Z_F^{(\text{LSE}(x))} = \frac{j\omega \varepsilon \gamma_{m,n}}{(\frac{m\pi}{a})^2 - k^2}$</td>
<td>$Z_F^{(\text{LSM}(x))} = \frac{(\frac{m\pi}{b})^2 - k^2}{j\omega \varepsilon \gamma_{m,n}}$</td>
</tr>
</tbody>
</table>

Table 4.4 LSE$(x)$ and LSM$(x)$ modes. The constant $A_{m,n}$ can be used to normalize the modes. The propagation constant $\gamma_{m,n}$ is the same as for ordinary TE and TM modes.
(TM\(^{(x)}\)) modes based on an \(x\)-directed magnetic respectively electric Hertz vector. Analogously, for LSE\(^{(y)}\) (TE\(^{(y)}\)) and LSM\(^{(y)}\) (TM\(^{(y)}\)) modes \(y\)-oriented magnetic and electric Hertz vectors, respectively, are used.

In fact, there is no need to do the complete derivation for LSE and LSM modes again. We just have to exchange the \(z\)-direction with the corresponding coordinate, i. e. for LSE\(^{(x)}\), LSM\(^{(x)}\) modes with \(x\) and for LSE\(^{(y)}\), LSM\(^{(y)}\) modes with \(y\).

In Table 4.4 we summarize the electric and magnetic Hertz vector as well as the corresponding field distribution for LSE\(^{(x)}\) and LSM\(^{(x)}\) modes. LSE\(^{(y)}\) and LSM\(^{(y)}\) modes can be found in a similar way.

The LSE and LSM modes fulfil the same power orthogonality condition (4.62) as ordinary TE and TM modes. Note, the LSE\(^{(x)}\) and LSM\(^{(x)}\) modes are orthogonal with respect to the cross-section of the waveguide, i. e. \(xy\)-plane and not with respect to the interface plane (\(yz\)-plane). Therefore in (4.62) we have to use \(e_\z\) for the direction of power transfer and not the direction \(e_x\) which has been used for the Hertzian vectors \(\Pi_e\) and \(\Pi_m\).

This orthogonality property can be used to normalize the LSE and LSM modes and implies that the modal expansion (4.72) respectively (4.74) is also valid for LSE and LSM modes. The eigenfunctions \(e(x, y)\) and \(h(x, y)\) represent in this case LSE and LSM modes instead of TE and TM modes. Furthermore, we have the same expansion coefficients \(U\) and \(I\), which of course have different values than for TE and TM modes, and the same relation between them using the modal impedances for LSE and LSM modes instead.

Finally we remark that LSE and LSM modes could also be expressed as superpositions of TE and TM modes and vice versa since both systems represent a complete set of functions.

### 4.3.15 Eigensolutions of Maxwell’s equations in circular waveguides

Using the method of Hertzian vectors we can find the TE and TM modes of circular waveguides as well. The procedure is exactly the same as for rectangular waveguides except that we deal with polar coordinates in the cross-section instead of Cartesian ones.

Compared to the rectangular waveguide, there is a small difference: In \(\varphi\) direction (circumference direction) we always have two solutions: \(\sin \mu \varphi\) and \(\cos \mu \varphi\). Thus all modes, except for \(\mu = 0\) are degenerate modes. To distinguish between the modes the superscripts \((s)\) and \((c)\) are introduced:

- If the TE (TM) mode has been derived from a Hertzian vector the angular dependency of which is \(\sin \mu \varphi\) we associate with this mode the superscript \((s)\). Thus the mode is called a TE\(^{(s)}\) (TM\(^{(s)}\)) mode.
- In case that the Hertzian vector has a \(\cos \mu \varphi\)-dependency we call it a TE\(^{(c)}\) (TM\(^{(c)}\)) mode.

The fundamental mode of a circular waveguide is the TE\(_{1,1}\) mode. The cutoff
frequencies of the modes are related to the zeroes of the Bessel function and its first derivative:

- The cutoff frequency $k_{c,\mu,\nu}$ of the $\text{TM}_{\mu,\nu}$ mode is related the $\nu$-th zero $\chi_{\mu,\nu}$ of the $\mu$-th order Bessel function $J_\mu(x)$, i.e. $J_\mu(\chi_{\mu,\nu}) = 0$.

- The cutoff frequency $k_{c,\mu,\nu}$ of the $\text{TE}_{\mu,\nu}$ mode is associated with the $\nu$-th zero $\chi'_{\mu,\nu}$ of the first derivative of the $\mu$-th order Bessel function $J'_\mu(x)$, i.e. $J'_\mu(\chi'_{\mu,\nu}) = 0$.

Different to the rectangular waveguide, the modes of a circular waveguide are always in the same order with respect to ascending cutoff wavenumbers $k_{c,\mu,\nu} = q_{\mu,\nu}$ since the order of the zeros $\chi_{\mu,\nu}$ and $\chi'_{\mu,\nu}$ does not change. In a rectangular waveguide, the order of the cutoff wavenumber of the modes depends on the ratio of the side lengths of the rectangular cross-section.

Table 4.5 presents a summary of TE and TM modes in a circular waveguide.

### 4.3.16 Eigensolutions of Maxwell’s equations in circular coaxial waveguides

A summary of the modes in a circular coaxial waveguide is presented in Table 4.6.

Different to the rectangular and circular waveguide, the circular coaxial waveguide consists of two separated conductors. It can therefore carry a TEM mode which is the fundamental and most important mode for applications from DC to microwaves.

Additionally to the Bessel functions we also need Neumann functions to describe the field distribution of a circular coaxial waveguide. Similar to the case of the circular waveguide, the cutoff wavenumbers of the modes are based on the zeroes of functions containing a combination of Bessel and Neumann functions or their derivatives. Moreover, we also have sine- and cosine-type modes: $\text{TE}^{(s)}$, $\text{TM}^{(s)}$, $\text{TE}^{(c)}$, $\text{TM}^{(c)}$. The order of cutoff wavenumbers depends, similar to the rectangular waveguide, on the ratio $c = R/r$ of the radii $R$ and $r$ of the outer and inner conductor, resp. The fundamental mode, however, is always the TEM mode since its cutoff wavenumber is always zero.

### 4.3.17 The equivalent transmission line of a waveguide

The transmission line theory is derived based on TEM wave systems, since only TEM waves allow a unique voltage and current definition. Nevertheless, the propagation of any mode (TEM, TE or TM) can be simulated using transmission line theory. The only problem for TE and TM waves: there is no unique definition for the voltage and current. In the literature there are various definitions for voltage and current in non-TEM waveguides and depending on this definition the transmission line models of a TE- or TM-mode differ one from another. This, in general, is not a problem as long as one bears in mind the definition.
Chapter 4 4.3 Uniform waveguides

Coordinate system for circular waveguide.

<table>
<thead>
<tr>
<th>Transverse wave number $q_{\mu,\nu}$</th>
<th>TE modes $\psi_{\mu,\nu} = \sqrt{\frac{\varepsilon_\mu}{\pi}} \frac{1}{\sqrt{\chi_{\mu,\nu}^2 - 1}} J_{\mu} \left( \frac{\chi_{\mu,\nu}}{r} \right) \left{ \cos \mu \phi \sin \phi \right}$</th>
<th>TM modes $\phi_{\mu,\nu} = \sqrt{\frac{\varepsilon_\mu}{\pi}} \frac{1}{\chi_{\mu,\nu} J_{\mu+1}(\chi_{\mu,\nu})} \left{ \cos \mu \phi \sin \phi \right} \varepsilon_i = \begin{cases} 1 &amp; \text{for } i = 0, \ 2 &amp; \text{for } i \neq 0 \end{cases}$</th>
</tr>
</thead>
</table>

- $\chi_{\mu,\nu}$ is the $\nu$-th root ($\chi_{\mu,\nu} \neq 0$) of the $\mu$-th order Bessel function, i.e. $J_{\mu}(\chi_{\mu,\nu}) = 0$.
- $\chi'_{\mu,\nu}$ is the $\nu$-th root ($\chi'_{\mu,\nu} \neq 0$) of the first derivative of the $\mu$-th order Bessel function, i.e. $J'_{\mu}(\chi'_{\mu,\nu}) = 0$.

**TE/TM modes:** $\mu = 0, 1, 2, \ldots; \nu = 1, 2, 3, \ldots$

**Fundamental mode**

- **TE**
  - $\psi_{1,1} = \sqrt{\frac{\varepsilon_\mu}{\pi}} \frac{1}{\sqrt{\chi_{1,1}^2 - 1}} J_{1} \left( \frac{\chi_{1,1}}{r} \right) \left\{ \cos \phi \sin \phi \right\}$
  - $\phi_{1,1} = \sqrt{\frac{\varepsilon_\mu}{\pi}} \frac{1}{\chi_{1,1} J_{1}(\chi_{1,1})} \left\{ \cos \phi \sin \phi \right\}$

**Degeneration:**

1. For $\mu \neq 0$: cosine and sine types $\text{TE}_{\mu,\nu}^{(c)}, \text{TM}_{\mu,\nu}^{(c)} - \text{TE}_{\mu,\nu}^{(s)}, \text{TM}_{\mu,\nu}^{(s)}$
2. $q_{1,\nu}^{(TM)} = q_{0,\nu}^{(TE)}$

**Table 4.5 TE and TM modes of the circular waveguide.**

- **TE**
  - $\psi_{1,1}^{(c)} = \sqrt{\frac{\varepsilon_\mu}{\pi}} \frac{1}{\sqrt{\chi_{1,1}^2 - 1}} J_{1} \left( \frac{\chi_{1,1}}{r} \right) \left\{ \sin \phi \right\}$
  - $\phi_{1,1}^{(s)} = \sqrt{\frac{\varepsilon_\mu}{\pi}} \frac{1}{\chi_{1,1} J_{1}(\chi_{1,1})} \left\{ \cos \phi \sin \phi \right\}$

- **TM**
  - $\psi_{0,1}^{(TM)} = \sqrt{\frac{\varepsilon_\mu}{\pi}} \frac{1}{\chi_{0,1} r J_{1}(\chi_{0,1})}$
We will study here a transmission line model of TE and TM waves where the term *characteristic impedance of the transmission line* has a clear physical meaning. The model is of great generality since it can be applied to any type of mode and waveguide type.

In circuit theory, the power flow along the transmission line is given by

\[ P = \frac{1}{2} \text{Re}\{U I^*\}. \]  

(4.87a)

\( U \) and \( I \) are solutions of the telegrapher equation.

In field theory, the power flow in a guided-wave system is carried out by the *transverse* component of the electric and magnetic fields. The power flow along the longitudinal direction of the waveguide is given by

\[ P = \frac{1}{2} \text{Re} \left\{ \int_S \mathbf{e}_t^v (\mathbf{E}_t \times \mathbf{H}_t^* ) \, dS \right\}. \]  

(4.87b)

Note that this expression has been also used to define the orthogonality of modes in waveguides (crossref. (4.62)).

Now, recall the expansion of the electromagnetic field in cylindrical systems by means of modes and modal voltage and current (4.72) or (4.74). In fact, as a consequence of Maxwell’s equations, the modal voltage \( U(z,t) \) and the modal current \( I(z,t) \) satisfy as well the telegrapher equation. We thus have found the *equivalent transmission line model* of a waveguide mode: The voltage and

\[
\begin{array}{|c|c|c|}
\hline
\text{TE}_{2,1}^{(c)} & q_{2,1}^{(TE)} = \frac{\chi_{2,1}}{r} & e_{2,1}^{(TE)} = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{\chi_{2,1}^2 - 4}} \left\{ \frac{2 \cdot J_1 \left( \frac{\chi_{2,1}}{r} \right)}{\rho J_1(\chi_{2,1})} \left\{ \begin{array}{l} \sin 2\phi \\ - \cos 2\phi \end{array} \right\} ight. \\
\hline
\text{TE}_{2,1}^{(s)} & q_{2,1}^{(TE)} = \frac{\chi_{2,1}}{r} & e_{2,1}^{(TE)} = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{\chi_{2,1}^2 - 4}} \left\{ \frac{\chi_{2,1} J_1 \left( \frac{\chi_{2,1}}{r} \right)}{r J_1(\chi_{2,1})} \left\{ \begin{array}{l} \cos 2\phi \\ \sin 2\phi \end{array} \right\} \right\} \\
\hline
\text{TM}_{1,1}^{(c)} & q_{1,1}^{(TM)} = \frac{\chi_{1,1}}{r} & e_{1,1}^{(TM)} = \sqrt{\frac{2}{\pi}} \left\{ - \frac{J_1' \left( \frac{\chi_{1,1}}{r} \right)}{\rho J_2(\chi_{1,1})} \left\{ \begin{array}{l} \cos \phi \\ \sin \phi \end{array} \right\} \right. \\
\hline
\text{TM}_{1,1}^{(s)} & q_{1,1}^{(TM)} = \frac{\chi_{1,1}}{r} & e_{1,1}^{(TM)} = \sqrt{\frac{2}{\pi}} \left\{ \frac{2 \cdot J_1 \left( \frac{\chi_{1,1}}{r} \right)}{\rho J_2(\chi_{1,1})} \left\{ \begin{array}{l} \sin \phi \\ - \cos \phi \end{array} \right\} \right\} \\
\hline
\text{TE}_{0,1} & q_{0,1}^{(TE)} = \frac{\chi_{0,1}}{r} & e_{0,1}^{(TE)} = \epsilon_r \sqrt{\frac{1}{\pi}} J_0 \left( \frac{\chi_{0,1}}{r} \right) \left\{ \begin{array}{l} \sin \phi \\ - \cos \phi \end{array} \right\} \\
\hline
\end{array}
\]

Table 4.5 TE and TM modes of the circular waveguide (continued).
Coordinate system for circular coaxial waveguide.

<table>
<thead>
<tr>
<th>TEM mode</th>
<th>( q_{0,0} = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE modes</td>
<td>( q_{\mu,1} = \frac{(c+1)\chi_{\mu,1}}{R+r} ) ( \mu = 1, 2, 3, \ldots )</td>
</tr>
<tr>
<td></td>
<td>( q_{\mu,\nu} = \frac{(c-1)\chi_{\mu,\nu}}{R-r} ) ( \mu = 0, 1, 2, \ldots ), ( \nu = 2, 3, 4, \ldots ) ( c = \frac{R}{r} )</td>
</tr>
<tr>
<td>TM modes</td>
<td>( q_{\mu,\nu} = \frac{(c-1)\chi_{\mu,\nu}}{R-r} ) ( \mu = 0, 1, 2, \ldots ), ( \nu = 1, 2, 3, \ldots )</td>
</tr>
</tbody>
</table>

\[ \chi_{m,n} \]

\[ J_{\mu}(c\chi_{\mu,\nu})N_{\mu}(\chi_{\mu,\nu}) - N_{\mu}(c\chi_{\mu,\nu})J_{\mu}(\chi_{\mu,\nu}) = 0 \]

\[ \chi'_{m,n} \]

\[ J'_{\mu}(c\chi'_{\mu,\nu})N'_{\mu}(\chi'_{\mu,\nu}) - N'_{\mu}(c\chi'_{\mu,\nu})J'_{\mu}(\chi'_{\mu,\nu}) = 0 \]

**Fundamental mode**

| TEM |

**Hertz vector**

\[ \phi = \frac{\ln \rho}{\sqrt{2\pi \ln R}} \]

\[ \psi_{\mu,\nu} = \zeta'_{\mu,\nu} \left( \chi'_{\mu,\nu} \frac{\rho}{r} \right) \left\{ \begin{array}{c} \cos \mu \phi \\ \sin \mu \phi \end{array} \right\} \]

\[ \phi_{\mu,\nu} = \zeta_{\mu,\nu} \left( \chi_{\mu,\nu} \frac{\rho}{r} \right) \left\{ \begin{array}{c} \cos \mu \phi \\ \sin \mu \phi \end{array} \right\} \]

\[ \zeta'_{\mu,\nu} \left( \chi'_{\mu,\nu} \frac{\rho}{r} \right) = \frac{\sqrt{2\pi \rho^2}}{2} \frac{J_{\mu}(\chi'_{\mu,\nu} \frac{\rho}{r})}{J_{\mu}(c\chi'_{\mu,\nu})} \left( \begin{array}{c} \chi'_{\mu,\nu} \frac{\rho}{r} \\ 1 \end{array} \right) \left[ 1 - \left( \frac{\mu}{c\chi'_{\mu,\nu}} \right)^2 \right] \]

\[ \zeta_{\mu,\nu} \left( \chi_{\mu,\nu} \frac{\rho}{r} \right) = \frac{\sqrt{2\pi \rho^2}}{2} \frac{J_{\mu}(\chi_{\mu,\nu} \frac{\rho}{r})}{J_{\mu}(c\chi_{\mu,\nu})} \left( \begin{array}{c} \chi_{\mu,\nu} \frac{\rho}{r} \\ 1 \end{array} \right) \left[ \frac{1}{J_{\mu}(c\chi_{\mu,\nu})} - 1 \right] \]

The modes of sine type can be obtained from the modes of cosine type by replacing \( \cos \mu \phi \) by \( \sin \mu \phi \) and \( \sin \mu \phi \) by \( -\cos \mu \phi \).

\( J_{\mu} \): Bessel function of the \( \mu \)-th order

\( N_{\mu} \): Neumann function of the \( \mu \)-th order

\( J'_{\mu} \): Derivative of the Bessel function of the \( \mu \)-th order

\( N'_{\mu} \): Derivative of the Neumann function of the \( \mu \)-th order

**Degeneration:**

For \( \mu \neq 0 \): cosine and sine types

\( TE_{\mu,\nu}^{(c)}, TM_{\mu,\nu}^{(c)} \) and \( -TE_{\mu,\nu}^{(s)}, TM_{\mu,\nu}^{(s)} \)

The zeros \( \chi_{\mu,\nu} \) and \( \chi'_{\mu,\nu} \) and therefore the eigenvalues \( q_{\mu,\nu} \) and the eigensolutions \( e_{\mu,\nu} \) and \( h_{\mu,\nu} \) are functions of the ratio \( c = R/r \).
TEM \(q_{0,0}^{(TEM)} = 0\)

\[ e_t^{(TEM)} = \varepsilon \frac{1}{\sqrt{2\pi \mu_0}} \frac{1}{r} \]

\( \text{TE}_{1,1} \)

\[ q_{1,1}^{(TE)} = \frac{(c+1)\chi_{1,1}}{R+r} \]

\[ e_t^{(TE)} = \varepsilon \frac{1}{\rho} \zeta_{1,1} \left( \frac{\rho}{r} \right) \left\{ \begin{array}{l} \sin \phi \\ -\cos \phi \end{array} \right\} + \varepsilon \frac{\chi_{1,1}}{r} \zeta'_{1,1} \left( \frac{\rho}{r} \right) \left\{ \begin{array}{l} \cos \phi \\ \sin \phi \end{array} \right\} \]

\( \text{TE}_{4,1} \)

\[ q_{0,1}^{(TM)} = \frac{(c-1)\chi_{0,1}}{R-r} \]

\[ e_t^{(TM)} = \varepsilon \frac{\chi_{0,1}}{r} \zeta'_{0,1} \left( \frac{\rho}{r} \right) \]

**Table 4.6** The TE and TM modes of the circular coaxial waveguide (continued).

The current of the transmission line are the modal voltage \(U(z,t)\) and the modal current \(I(z,t)\), respectively. The characteristic impedance \(Z_c\) of the transmission line is given by the modal impedance \(Z_F\) and the propagation constant by \(\gamma = \sqrt{T^2 - k^2}\).

This equivalent transmission line or circuit model of a waveguide has a clear physical meaning: the modal voltage and current represent the strength of the electric and magnetic field since we normalized the eigenfunctions (modes). And the characteristic impedance of the transmission line represents the ratio of the magnetic and the electric field of a wave travelling in one direction of the waveguide (transmission line).

Due to this analogy one often speaks of a transmission line when speaking about a waveguide.

As we have seen, there is an infinite number of modes present in any waveguide. For this reason the equivalent circuit model of a waveguide is an infinite number of transmission lines each representing one mode of the waveguide. This is a direct consequence of the power orthogonality (4.62) of the modes.

### 4.3.18 Frequency characteristics and behaviour

The last sections were mainly devoted to the solution of Maxwell’s equations in cylindrical systems and the representation of the electromagnetic field distribution by modes. We also discussed the types of modes. This section focuses on
the frequency characteristics of uniform waveguides. We have already seen a fundamental frequency property in Section 4.3.7: Depending on $T$ and $k$ we can distinguish between TEM, fast and slow waves. The purpose of this section is to study in more detail the frequency behaviour of uniform waveguides and related phenomena such as dispersion.

Recalling the electromagnetic field representation (4.72) or (4.74) we see that the frequency behaviour is determined by the part which goes with the longitudinal coordinate (the waveguide’s axis). On the contrary, the eigensolutions $e_{t_n}$ and $h_{t_n}$ respectively $\phi_n$ and $\psi_n$ do not depend on the frequency but only on the geometry of the cross section of the waveguide. This means that the frequency behaviour of uniform waveguides is uniquely determined by the modal voltages $U_n$ and modal currents $I_n$.

A) Transverse and cutoff wave number

**Transverse wave number** A key-role in the study of the frequency behaviour plays the eigen value $T$, which was introduced in Section 4.3.7 as a constant to separate the transverse and longitudinal variables. $T$ only depends on the geometry of the waveguide’s cross-section and takes on different values depending on the mode type and order. It is called transverse wave number and has the unit 1/m. When studying rectangular, circular and circular-coaxial waveguides, we introduced for it the symbol $q_{\mu,\nu}$ as well to emphasize that the value of $T$ depends on the modal indices $\mu$ and $\nu$. For example for rectangular waveguides we saw

$$T = q_{\mu,\nu} = \pi \sqrt{\left(\frac{\mu}{a}\right)^2 + \left(\frac{\nu}{b}\right)^2}.$$  

**Cutoff wave number** The propagation constant $\gamma$ is the analogue to $T$ for the 1D wave equation when separating $z$ and the time $t$ (see Section 4.3.7). It depends on $T$ and $k$ and is given by

$$\gamma = \sqrt{T^2 - k^2}.$$  

The wave number $k_c$ for which the argument of the square root vanishes is given by

$$k_c = T$$  

and is the critical state of a waveguide mode. $k_c$ is called the cutoff (angular) wave number and since $k_c = T$, $T$ is called cutoff wave number as well.

Since the different modes of a waveguide have different transverse wave numbers $q_{\mu,\nu}$, we can associate with each mode a cutoff wave number $k_{c,\mu,\nu}$.

$T$ or $k_c$ is a branch point of $\gamma$: depending on whether $k < k_c$ or $k > k_c$, $\gamma$ changes its characteristic: either it is real- or imaginary-valued.
**B) Cutoff frequency and cutoff wavelength**

**Cutoff frequency** From (4.88) follows the cutoff frequency

\[ f_c = \frac{c}{2\pi} k_c = \frac{c}{2\pi} T. \]  \hspace{1cm} (4.89)

This means, the larger the transverse wave number the higher the associated cutoff frequency of a mode. The mode with the smallest cutoff frequency respectively transverse wave number is the *fundamental mode* of a waveguide. If the smallest \( T \) is zero, then we deal with a TEM mode. Otherwise, the lowest *non-vanishing* eigen value \( T \) is *generally* associated with a TE-mode.

Analogously to the cutoff wave number, a mode \((\mu,\nu)\) has the cutoff frequency \( f_{c(\mu,\nu)} \).

**Cutoff wavelength** With the cutoff frequency \( f_c \) or cutoff wavenumber \( k_c \) we can associate the **cutoff wavelength**

\[ \lambda_c = \frac{2\pi}{k_c} = \frac{2\pi}{f_c} = \frac{2\pi}{T}. \]  \hspace{1cm} (4.90)

This is the wavelength of a *plane* electromagnetic wave in an unbounded medium at the frequency \( f_c \). To distinguish the cutoff wavelengths of different modes we write \( \lambda_{c(\mu,\nu)} \).

For example, for an empty rectangular waveguide we have for the fundamental mode \((\text{TE}_{1,0})\) \( \lambda_{c1,0} = 2a \), where \( a \) is the width of the waveguide. The cutoff wavelength can be interpreted as follows in this case: the width \( a \) has to be half of the *free space* wavelength.

Sometimes, the **transverse wavelength** \( \lambda_t \) is introduced as

\[ \lambda_t = \frac{2\pi}{T} \]  \hspace{1cm} (4.91)

and the cutoff wavelength is then \( \lambda_c = \lambda_t \).

**C) Operation states of a waveguide: propagating and evanescent modes**

The propagation constant \( \gamma \) changes its characteristics at the cutoff wavenumber \( k_c \). In the case of lossless media it changes from purely real to purely imaginary. This change has a fundamental impact on the propagation characteristic of a mode.

**Evanescent modes** In the case \( k < k_c \) which is equal to \( f < f_c \) we have

\[ \gamma = \alpha, \quad \alpha \text{ real}. \]  \hspace{1cm} (4.92)

The mode is called to be **under cutoff**. According to the propagation factor \( e^{\pm \gamma z} = e^{\pm \alpha z} \) we deal with damped exponentials in \( z \)-direction, i. e. in the direction of the waveguide’s axis. This means that all field components fade away. The
mode is therefore called an evanescent mode and $\gamma$ represents an attenuation constant in this case.

**Propagating modes** On the other hand, if the operating frequency is large enough such that $f > f_c$ or $k > k_c$, then

$$\gamma = j\beta, \quad \beta \text{ real} \quad (4.93)$$

and $e^{\pm \gamma z} = e^{\pm j\beta z}$. This mode propagates along the waveguide axis and is called a propagating mode. Its state is above cutoff and $\gamma$ represents a phase constant in this case.

Since the cutoff frequencies of modes differ, the state of operation for a given operating frequency $f$ is different: For finite frequencies there will be a limited number of modes that are above cutoff and all others are under cutoff. If $f$ is smaller than the lowest cutoff frequency in a non-TEM waveguide, i.e. smaller than the cutoff frequency of the fundamental mode, then all modes are below cutoff. In TEM waveguides there is always one mode propagating: the TEM one.

There is a specific frequency range where the waveguide is operated in a monomode regime: The operating frequency $f$ is between the cutoff frequencies of the fundamental and the first higher order mode. In this case only the fundamental mode is propagating, all others are evanescent modes.

Let us consider a specific example. Figure 4.13 shows the two operating ranges for a rectangular waveguide versus the normalized cutoff frequency. The reference frequency is the cutoff frequency of the fundamental mode, which is the mode with the lowest eigenvalue $T$. In our example this is the mode TE$_{1,0}$. The operating range “propagating modes” is represented by the shaded region. The other one is the range “evanescent modes”. The figures at the ordinate represent the order of the modes with respect to ascending eigenvalues $T$ and cutoff frequencies $f_c$, respectively. As an example, let us consider the normalized frequency range $2.7 < f_c/f_{c_{TE_{1,0}}} < 3.5$. In this range we find 8 propagating modes (TE$_{1,0}$, TE$_{2,0}$, TE$_{0,1}$, TE$_{1,1}$, TM$_{1,1}$, TE$_{3,0}$, TE$_{2,1}$, TM$_{2,1}$). All other ones are evanescent modes (non-shaded region). Further, we can see from the diagram that all modes TE$_{m,n}$ and TM$_{m,n}$ with $m \geq 1$ and $n \geq 1$ are degenerated.

In waveguides with at least two separated conductors (e.g., a coaxial waveguide), the mode with the lowest cutoff frequency is always the TEM-mode with $f_c = 0$. This means, that the TEM-mode is a propagating mode over the complete frequency range $0 < f < \infty$.

**D) Dispersion and dispersion-diagram**

As seen in the last section, we have to distinguish between two operation states of the waveguide. This results from the fact that the propagation constant $\gamma$ changes from purely real to purely imaginary at the cutoff frequency $f_c$. The propagation constant, and all quantities based on it (e.g. the modal impedance), depend thus on the frequency. This frequency dependency is called dispersion.
We can normalize the propagation constant of a mode using its cutoff frequency and its transverse wave number:

$$\frac{\gamma}{T} = \sqrt{1 - \left(\frac{f}{f_c}\right)^2}. \quad (4.94)$$

The normalized propagation constant is drawn in Figure 4.14 for an arbitrary mode with $T > 0$, i.e. a non-TEM mode.

For very small operating frequencies $f \ll f_c$ the attenuation constant $\alpha$ converges towards the transverse wavenumber $T$. In Section 4.3.10 we saw that the eigenvalues $T$ form an ascending sequence with $T \to \infty$ (crossover Sturm-Liouville problem (see Appendix D). From this follows that the attenuation constant becomes larger and larger. Hence, the higher the eigenvalue $T$ or the cutoff frequency $f_c$ the stronger the mode is attenuated for small frequencies. On the other hand, the attenuation of a mode does not exceed the value $\alpha = T$. From this it is clear that in the modal representation (4.72) of the electromagnetic field, we first must use the modes with the lowest cutoff wave numbers since they contribute the most to the electromagnetic field distribution. This is especially important in numerical techniques where the infinite sums (4.72) or (4.74) must be truncated after a finite number of terms.

Just below the cutoff frequency the attenuation constant is very small. In
this range the modulus of the field components decay only very slowly along the propagation direction $z$. Although the mode is still under cutoff, it already can contribute significantly to the electromagnetic field.

Just above the cutoff frequency the phase constant $\beta$ is very small and augments with increasing frequency. At the beginning, the increase is stronger than linearly.

For $f \gg f_c$, $\beta$ converges asymptotically towards the phase constant $\frac{\omega}{c}$ of a plane wave.

The asymptote represents as well the frequency dependency of a TEM mode for which $T = 0$ holds. In other words, $\gamma = j\beta$ depends linearly on the frequency for a TEM mode. Thus, a TEM mode has a further analogy with a plane wave in an unbounded medium: additionally to the fact that the electromagnetic field has only transverse field components, it shows also the same frequency behaviour.

The phenomenon that the propagation constant depends on the frequency is called dispersion and the diagram of Figure 4.14 the dispersion-diagram.

We can summarize two facts:

1. A TEM-mode is always non-dispersive since $\gamma$ varies linearly with $\omega$ and behaves like a plane wave in an unbounded medium with respect to frequency.

2. Any TE- or TM-mode in a hollow metallic waveguide is dispersive.

Dispersion follows from the fact that the propagation constant is non-linear. The stronger the frequency dependency deviates from a linear function the stronger the dispersion will be. From Figure 4.14 we see that $\beta$ converges asymptot-
ically towards the linear frequency behaviour of a plane wave which does not show any dispersion. From this we can conclude that the larger the difference between operating frequency and cutoff frequency of a mode the less dispersive the mode is. On the other hand, if this difference is small, that is the mode is operated close to its cutoff frequency, the mode suffers a larger dispersion. Thus it follows that for practical applications the operation frequency should be far away from the cutoff frequency of the used mode.

However, in practise there is an upper limit for the operating frequency. This is, when the next mode starts to propagate. In Figure 4.14 we have drawn the propagation constant for a single mode. Let’s look at the dispersion diagram of the fundamental mode and the first higher order mode. As example we use the rectangular waveguide of Figure 4.13. The fundamental mode is a TE$_{1,0}$ and the first higher mode a TE$_{2,0}$. The normalized propagation constants of both modes are shown in Figure 4.15. The frequency has been normalized to the cutoff frequency $f_{c1,0}$ of the fundamental mode and the propagation constant to the associated transverse wave number $q_{1,0}$. Since the first higher order mode is a TE$_{2,0}$ its cutoff frequency is $f_{c2,0} = 2 \cdot f_{c1,0}$ and $q_{2,0} = 2 \cdot q_{1,0}$.

Often, in waveguide-based systems, the waveguides are operated in the monomode range, that is only the fundamental mode is a propagating mode and all others are evanescent ones. For the waveguide represented by Figure 4.15 this frequency range is $1 < f/f_{c1,0} < 2$. Further above we mentioned that on the one hand we should remain as far away as possible from the cutoff frequency of the operating mode in order to reduce the dispersion. For our example the operating mode would be the fundamental mode and $f$ should be therefore far away from $f_{c1,0}$. On the other hand we saw, that if the operating frequency is approaching the cutoff frequency of a mode, then this mode, although still attenuated, will contribute already significantly to the electromagnetic field and therefore influence the functioning of the system. Therefore, the operating frequency $f$ should not be too close to $f_{c2,0}$ for the example of ref[fig:normalized propagation constant versus frequency for fundamental and higher order mode].

These considerations lead to an effective operation range of a waveguide mode. How far above cutoff a mode is operated and how far away the operating frequency must remain from the next higher order mode depends in practise on the specifications of the system.

\section*{E) Guided wavelength}

For a mode above cutoff we can derive another frequency dependent quantity: the \textit{guided} wavelength. For a mode below cutoff this term is meaningless.

\textbf{Guided wavelength} The \textit{guided} wavelength is defined as the \textit{shortest} length after which a certain state of the propagating wave is repeated periodically. Since the propagation term is $e^{\pm j\beta z}$ the guided wavelength is given by the condition
Figure 4.15 Normalized propagation constants of the first two modes of the rectangular waveguide in Figure 4.13. The frequency has been normalized to the cutoff frequency $f_{c_{1,0}}$ of the fundamental mode and the propagation constants to the transverse wave number $q_{1,0}$.

$e^{\pm j \beta_{q_{1,0}}} = e^{\pm j 2\pi}$. From $\beta \lambda = 2\pi$ we can derive

$$\lambda_g = \frac{c}{\sqrt{f^2 - f_c^2}}.$$

(4.95a)

Inserting the cutoff wavelength (4.90), yields the normalized guided wavelength of a mode:

$$\lambda_g = \frac{\lambda_c}{\sqrt{(\frac{\lambda}{\lambda_c})^2 - 1}}.$$

(4.95b)

If we compare $\lambda_g$ with the wavelength $\lambda$ of a wave with the same frequency but in an unbounded medium, we obtain

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (\frac{\lambda}{\lambda_c})^2}}.$$

(4.95c)

This expression shows that for $f > f_c$ or, what is equivalent, for $\lambda < \lambda_c$ the guided wavelength $\lambda_g$ of a dispersive mode ($T > 0$) is always greater than the wavelength $\lambda$ in an unbounded medium. This is shown in Figure 4.16 by the asymptote for $f \to \infty$ ($\lambda \to 0$). For very high frequencies ($f \to \infty$), the wave behaves like a wave in an unbounded medium since the lateral dimensions of the waveguide become electrically large.

We also learn from Figure 4.16 that the cutoff frequency for a non-TEM mode ($T > 0$) is like a barrier and has the same meaning as $f = 0$ for a TEM mode or...
Figure 4.16 Normalized guided wavelength versus normalized frequency for a mode with $T > 0$ (solid line). The guided wavelength has been normalized to the free space wavelength. The dashed line is the asymptote of $\lambda_g$ when $f/f_c \to \infty$ and represents the wavelength of a wave with the same frequency but propagating in an unbounded medium.

A wave in an unbounded medium. The frequency $\sqrt{f^2 - f_c^2}$ ($f > f_c$) can be considered as something like an effective frequency. Using this effective frequency we can in fact reuse the formulas of free space for wavelength, $\beta$, etc., replacing $f$ by $\sqrt{f^2 - f_c^2}$ and find the equivalent expressions for a wave propagating in a medium the dispersion relation of which is $\gamma = \sqrt{k_c^2 - k^2}$.

F) Modal impedance

Another frequency dependent parameter is the modal impedance. It can be normalized to the modal impedance of a TEM mode:

$$Z_F^{(TE)} = jZ_F^{(TEM)} \frac{k}{\gamma},$$  \hspace{1cm} (4.96)$$

$$Z_F^{(TM)} = -jZ_F^{(TEM)} \frac{\gamma}{k},$$  \hspace{1cm} (4.97)$$

which shows that the frequency behaviour of the modal impedance associated with a TM mode is the inverse of the frequency dependency of a TE mode. The normalized modal impedances are plotted in Figure 4.17. Below the cutoff frequency ($f/f_c < 1$) $Z_F^{(TE)}$ and $Z_F^{(TM)}$ are purely imaginary in the lossless case. This is indicated by the broken line. Above cutoff they are real-valued indicated by the solid lines. The modal impedance of a TEM wave is always real-valued and
Chapter 4  4.3 Uniform waveguides

\[ Z_{\text{TEM}} F \]

\[ Z_{\text{TE}} F \]

\[ Z_{\text{TM}} F \]

\[ f/f_c \]

**Figure 4.17** Normalized modal impedances of TE, TM and TEM modes. Below cutoff \((f/f_c < 1)\) the modal impedances of TE and TM modes are imaginary. Above cutoff they are real-valued and converge towards the modal impedance of a TEM wave for high frequencies.

constant.
APPENDIX A

Exercices

Exercice 1

*english version:*

The dimensions of a rectangular waveguide operating in the X-band are 
\( a = 22.86 \text{ mm} \) and \( b = 10.16 \text{ mm} \).

a) Calculate the cutoff frequency of the fundamental mode.

b) Which mode is the first higher order mode and what is its cutoff frequency.

c) What is the diameter of a circular waveguide that should operate in the same frequency band?

d) How large is the frequency band in which the circular waveguide can be operated in a mono-mode configuration?

*version française:*

Les dimensions d'un guide d'ondes rectangulaire dans la bande X sont 
\( a = 22.86 \text{ mm} \) et \( b = 10.16 \text{ mm} \).

a) Calculer la fréquence de coupure du mode fondamental.

b) Quel est le premier mode supérieur et quelle est sa fréquence de coupure ?

c) Quel devrait être le diamètre d'un guide d'ondes circulaire si l'on travaille dans la même bande de fréquence ?

d) Donner la largeur de la bande de fréquence pour ce guide d'ondes circulaire travaillant dans une configuration mono-mode.

**Solution 1**

a) 6.56 GHz.

b) \( \text{TE}_{20} \), 13.1 GHz.

c) \( D = 26.8 \text{ mm} \).
d) 6.56-8.57 GHz.
**Exercice 2**

*english version:*

A waveguide of arbitrary cross section is air-filled and operated at 30 GHz. The cutoff frequency of the propagating TM mode is 18 GHz.

a) Determine the wavelength in the waveguide.
b) Calculate the phase constant.
c) What is the phase velocity of the mode?
d) Compute the intrinsic (characteristic) impedance of the mode.
e) If the mode were an TE mode, what would be the characteristic impedance?
f) Assuming the mode is used as carrier for a modulated signal, what would be the group velocity of the signal?
g) Given an operating frequency of 15 GHz, determine the distance over which the fields of this mode will be reduced in magnitude by 20 dB.

*version française:*

Un guide d’ondes à une section quelconque travail à 30 GHz. La fréquence de coupure du mode TM propagant est 18 GHz.

a) Déterminer la longueur d’onde dans le guide d’ondes.
b) Calculer la constante de phase.
c) Quelle est la vitesse de phase de ce mode ?
d) Calculer l’impédance intrinsèque (caractéristiques) de ce mode.
e) Si le mode est un mode TE, quel serra son impédance caractéristique ?
f) En supposant que le mode est utilisé comme porteuse pour un signal modulé, quelle serra la vitesse de groupe du signal ?
g) Donnant une fréquence de fonctionnement de 15 GHz, déterminer la distance à partir de laquelle le champ serra réduit (en amplitude) par 20 dB.

**Solution 2**

a) $\lambda_g = 1.25$ cm.
b) $\beta = 160\pi$ rad/m.
c) $v_p = 3.75 \cdot 10^8$ m/s.
d) $Z_F^{(TM)} = 301.2\, \Omega$.
e) $Z_F^{(TE)} = 471.2\, \Omega$.
f) $v_g = 2.4 \cdot 10^8$ m/s.
g) $l = 11.05$ mm.
Exercices
Orthogonal coordinate systems

B.1 Cylindrical coordinate systems

B.1.1 Rectangular coordinates

Consists of three sets of mutual orthogonal parallel planes.

Figure B.1 Rectangular coordinates \((x, y, z)\). Coordinate surfaces are the planes: \(x = \text{const}, y = \text{const}, z = \text{const}\).
B.1.2 Circular-cylinder coordinates

Consists of a set of coaxial circular cylinders, a set of half planes rotated around the axis, and a set of parallel planes perpendicular to the axis. The circular-cylinder coordinate system is also a rotational coordinate system.

![Circular-cylinder coordinates](image)

**Figure B.2** Circular-cylinder coordinates \((r, \psi, z)\). Coordinate surfaces are circular cylinders \((r = \text{const})\), half-planes \((\psi = \text{const})\) intersecting on the \(z\)-axis, and parallel planes \((z = \text{const})\).
B.1.3 Elliptic-cylinder coordinates

Consists of a set of confocal elliptic cylinders, a set of confocal hyperbolic cylinders perpendicular to the elliptic cylinders, and a set of parallel planes perpendicular to the axis.

Figure B.3 Elliptic-cylinder coordinates \((\eta, \psi, z)\). Coordinate surfaces are elliptic cylinders \((\eta = \text{const})\), hyperbolic cylinders \((\psi = \text{const})\), and planes \((z = \text{const})\).
B.1.4 Parabolic-cylinder coordinates

Consists to two sets of mutual orthogonal parabolic cylinders and a set of parallel planes perpendicular to the axis.

Figure B.4 Parabolic-cylinder coordinates \((\mu, \nu, z)\). Coordinate surfaces are parabolic cylinders \((\mu = \text{const}, \nu = \text{const})\), and planes \((z = \text{const})\).
B.2 Rotational coordinate systems

B.2.1 Spherical coordinates

Consists of a set of concentric spheres, a set of cones perpendicular to the spheres, and a set of half planes rotated around the polar axis.

Figure B.5 Spherical coordinates \((r, \theta, z)\). Coordinate surfaces are spheres \((r = \text{const})\), circular cones \((\theta = \text{const})\), and half planes \((\psi = \text{const})\).
B.2.2 Prolate spheroidal coordinates

Consists of a set of confocal prolate spheroids, a set of confocal hyperboloids of two sheets, and a set of half planes rotated around the polar axis.

Figure B.6 Prolate spheroidal coordinates \((\eta, \theta, \psi)\). Coordinate surfaces are prolate spheroids \((\eta = \text{const})\), hyperboloids of revolution \((\theta = \text{const})\), and half planes \((\psi = \text{const})\).
B.2.3 Oblate spheroidal coordinates

Consists of a set of confocal oblate spheroids, a set of confocal hyperboloids of one sheet and a set of half planes rotated around the polar axis.

Figure B.7 Oblate spheroidal coordinates \((\eta, \theta, \psi)\). Coordinate surfaces are oblate spheroids \((\eta = \text{const})\), hyperboloids of revolution \((\theta = \text{const})\), and half planes \((\psi = \text{const})\).
B.2.4 Parabolic coordinates

Consists of two sets of mutual orthogonal paraboloids of revolution and a set of half planes rotated around the polar axis.

**Figure B.8** Parabolic coordinates \((\mu, \nu, \psi)\). Coordinate surfaces are paraboloids of revolution \((\mu = \text{const}, \nu = \text{const})\), and half planes \((\psi = \text{const})\).
B.3 General coordinate systems

B.3.1 Conical coordinates

Consists of a set of concentric spheres and two sets of mutual orthogonal elliptic cones.

Figure B.9 Conical coordinates \((r, \theta, \lambda)\). Coordinate surfaces are spheres \((r = \text{const})\), and elliptic cones \((\theta = \text{const}, \lambda = \text{const})\).
B.3.2 Ellipsoidal coordinates

Consists of a set of ellipsoids, a set of hyperboloids of one sheet, and a set of hyperboloids of two sheets.

Figure B.10 Ellipsoidal coordinates \((\eta, \theta, \lambda)\). Coordinate surfaces are ellipsoids \((\eta = \text{const})\), and hyperboloids \((\theta = \text{const}, \lambda = \text{const})\).
B.4 Paraboloidal coordinates

Consists of two sets of mutually orthogonal elliptic paraboloids, and a set of hyperbolic paraboloid.

Figure B.11 Paraboloidal coordinates \((\mu, \nu, \lambda)\). Coordinate surfaces are elliptic paraboloids \((\mu = \text{const}, \nu = \text{const})\), and hyperbolic paraboloids \((\lambda = \text{const})\).
APPENDIX C

The method of separation of variables

In this chapter we will study a concrete problem that is solved by means of the method known as Separation of Variables.

Let us consider a rectangular waveguide of cross section $a \times b$. We are interested in computing the electric Hertz vector $\Pi_e$, and seek for solutions of the form

$$\Pi_e(x, y, z) = \Phi(x, y) e^{\pm \gamma z}.$$  \hspace{1cm} (C.1)

The electric Hertz vector must satisfy the Helmholtz equation

$$\triangle \Pi_e + k^2 \Pi_e = 0,$$  \hspace{1cm} (C.2a)

and the boundary conditions

$$\Phi(x, y) = 0 \quad \text{on the boundary.}$$  \hspace{1cm} (C.2b)

The problem can be reformulated as the following boundary-value problem:

$$\triangle_t \Phi + \lambda \Phi = 0,$$  \hspace{1cm} (C.3a)

$$\Phi = 0 \quad \text{on the boundary,}$$  \hspace{1cm} (C.3b)

where the dependency on $z$ has been split off.

The method of separation of variables assumes now that the dependency of $\Phi$ on $(x, y)$ can be expressed by two terms either of each depending only on a single variable. In our case, we assume that $\Phi$ can be written as product

$$\Phi(x, y) = X(x) \cdot Y(y).$$  \hspace{1cm} (C.4)

Note, $X$ only depends on $x$ and $Y$ only on $y$.

Inserting (C.4) into (C.3) yields

$$\frac{\partial^2 X}{\partial x^2} \cdot Y + X \cdot \frac{\partial^2 Y}{\partial y^2} + \lambda X \cdot Y = 0$$  \hspace{1cm} (C.5)

Since we search for non-trivial solutions, i. e. $\Phi \neq 0 \forall (x, y)$, we can divide the last equation by $\Phi$ and obtain

$$\frac{X''}{X} + \frac{Y''}{Y} + \lambda = 0, \quad \forall 0 < x < a \text{ and } 0 < y < b.$$  \hspace{1cm} (C.6)
The method of separation of variables

The keypoint in the procedure is that this equation holds $∀$ $0 < x < a$ and $0 < y < b$. However, since $X$ is only a function of $x$ and $Y$ is depending only on $y$, the above equation can only be satisfied if $X''/X$ and $Y''/Y$ are constant $∀$ $0 < x < a$ and $0 < y < b$.

Remark: If the above equation held only for some single points, then we might be able to find some functions $X(x)$ and $Y(y)$ that can satisfy it.

Following these considerations, we can write

$$\frac{X''}{X} = -T_x \quad \text{and} \quad \frac{Y''}{Y} = -T_y.$$  \hfill (C.7)

We deal now with two ordinary differential equations (ODEs) with constant coefficients

$$X'' + T_x X = 0, \quad (C.8)$$
$$Y'' + T_y Y = 0, \quad (C.9)$$
$$-T_x - T_y + \lambda = 0. \quad (C.10)$$

$T_x$ and $T_y$ are arbitrary complex numbers. The reason, why we have used explicitly the minus sign, will become clear when we look at the solution of the ODEs and the boundary conditions that have to be imposed.

The solution $X$ and $Y$ are given by

$$X(x) = A \cos k_xx + B \sin k_xx, \quad k_x = \sqrt{T_x}, \quad (C.11)$$
$$Y(y) = C \cos k_yy + D \sin k_yy, \quad k_y = \sqrt{T_y}, \quad (C.12)$$

with $A$, $B$, $C$, and $D$ arbitrary complex constants.

Now, we impose the boundary conditions to determine $k_x$ and $k_y$. Let’s start with $X(x)$. From (C.3) follows that

$$X(0) = 0 \quad \text{and} \quad X(a) = 0. \quad (C.13a)$$

From the first condition follows that

$$A = 0. \quad (C.13b)$$

Taking this into account, the second condition results in

$$B \sin k_xa = 0. \quad (C.13c)$$

This equation has two solutions

$$B = 0 \quad \text{or} \quad k_xa = m \cdot \pi, \quad m \in \mathbb{Z}. \quad (C.13d)$$

The first solution yields $X(x) \equiv 0 \forall x$. This trivial solution needs to be discarded and only the second represents a non-trivial solution to the boundary-value
problem.
Thus
\[ X(x) = B \sin \frac{m \pi x}{a}, \quad m \in \mathbb{Z}. \] (C.13e)

The same procedure is used to obtain
\[ Y(y) = D \sin \frac{n \pi y}{b}, \quad n \in \mathbb{Z}. \] (C.14)

We finally have found a particular solution
\[ \Phi_{m,n}(x,y) = E_{m,n} \sin \frac{m \pi y}{a} \sin \frac{n \pi y}{b}, \quad m, n \in \mathbb{Z}. \] (C.15)

\( E_{m,n} \) is an arbitrary constant that can be determined by imposing a further condition on \( \Phi_{m,n} \). For example, we could require that the integral of \( \Phi_{m,n}^2(x,y) \) over the cross-section \( a \times b \) equals one.

Finally, we can compute \( \gamma \) in (C.1). From (C.8) follows that
\[ \lambda = T_x + T_y \iff \gamma_{m,n} = \left( \frac{m \pi}{a} \right)^2 + \left( \frac{n \pi}{b} \right)^2 - k^2, \quad m, n \in \mathbb{Z}. \] (C.16)
The method of separation of variables
Sturm-Liouville problem and related theorems

By separation of variables in an appropriate coordinate system, the Laplace or Helmholtz equation reduces to some ordinary differential equations of the following general form

\[
\frac{d}{dx} \left[ p(x) \frac{dy(x)}{dx} \right] - [q(x) - \lambda \rho(x)] y(x) = 0. \tag{D.1a}
\]

For example, in the case of Cartesian coordinates (rectangular waveguide) \( p(x) = 1 \), \( q(x) = 0 \), and \( \rho(x) = 1 \) This equation is known as the Sturm-Liouville equation. The function \( y(x) \) satisfies the homogeneous boundary conditions of the first kind \( y(x) = 0 \) on the boundary) or the second kind \( \frac{dy(x)}{dx} = 0 \) on the boundary). They are also known under the name Dirichlet and Neumann conditions, respectively. For example if the boundary is given by \( x = a \) and \( x = b \) then the boundary condition of the first kind is

\[
y(x)|_{x=a} = 0, \quad y(x)|_{x=b} = 0, \tag{D.1b}
\]

and of the second kind

\[
\frac{dy(x)}{dx} \bigg|_{x=a} = 0, \quad \frac{dy(x)}{dx} \bigg|_{x=b} = 0. \tag{D.1c}
\]

A more general form of the homogeneous boundary conditions is

\[
\left[ \alpha \frac{dy(x)}{dx} - \beta y(x) \right] \bigg|_{x=a,x=b} = 0, \tag{D.1d}
\]

where \( \alpha \) and \( \beta \) are two constants, including \( \alpha = 0 \) for the Dirichlet or first kind boundary condition and \( \beta = 0 \) for the Neumann or second kind boundary condition.

(D.1) is known as the Sturm-Liouville problem or eigenvalue problem. The following are four basic theorems about the Sturm-Liouville problems.

**Theorem 1** Only specific or discrete (real) values of \( \lambda \) are allowed for a nontrivial solution of the Sturm-Liouville equation satisfying the specific set of boundary conditions. These allowed values \( \lambda \) are called eigenvalues. The eigenvalues, ordered with respect to magnitude form a denumerable sequence
\[ \lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k, \ldots \]

For each specific eigenvalue \( \lambda_k \), there is a corresponding function \( y_k(x) \) that satisfies the differential equation and the boundary conditions (D.1). These functions \( y_k(x) \) are called the **eigenfunctions** of the problem. The complete set of eigenfunctions is

\[ y_1(x), y_2(x), y_3(x), \ldots, y_k(x), \ldots \]

Each eigenvalue corresponds to one eigenfunction or a number of linearly independent eigenfunctions. If more than one eigenfunction is allowed for a particular eigenvalue the problem is said to be **degenerate**.

Frequently, the problem of uniform waveguides is a degenerate problem. Often TE and TM modes have the same eigenvalue \( T \).

**Theorem 2**  All eigenvalues are nonnegative for \( q \leq 0 \), and with the homogeneous boundary conditions of (D.1),

\[ \lambda_k \geq 0. \]

**Theorem 3. Orthogonality Theorem**  The eigenfunction set is a complete orthogonal set. The eigenfunctions \( y_m(x) \) and \( y_n(x) \) corresponding to eigenvalues \( \lambda_m \) and \( \lambda_n \), respectively, are orthogonal with weight \( \rho(x) \):

\[ \int_a^b \rho(x)y_m(x)y_n^*(x) \, dx = 0, \quad m \neq n. \]

**Theorem 4. Expansion Theorem**  Every continuous function \( f(x) \) which has piecewise continuous first and second derivatives and satisfies the boundary conditions of the eigenvalue problem can be expanded in an absolutely and uniformly convergent series in terms of the eigenfunctions

\[ f(x) = \sum_{n=1}^{\infty} a_n y_n(x). \]

The coefficients \( a_n \), can be obtained by using the orthogonality property of the eigenfunctions:

\[ a_n = \frac{\int_a^b \rho(x')f(x')y_n^*(x') \, dx'}{\int_a^b \rho(x')|y_n(x')|^2 \, dx'}. \]
Theorem 5. Normalization  The orthogonality theorem can be used to normalize the eigenfunctions $y_k(x)$. The normalized eigenfunction $\hat{y}_k(x)$ is given by

$$\hat{y}_k(x) = \frac{1}{\sqrt{A_k}} y(x) \quad \text{with} \quad A_k = \int_a^b \rho(x) |y_k(x)|^2 \, dx.$$
APPENDIX E

Uniform waveguides

In section Section 4.3.4 we have seen how the 3D problem of finding the electromagnetic field distribution in a uniform waveguide can be reduced to a 1D problem using the Method of Hertz Vectors. In this appendix, we will study the Method of Longitudinal Components to achieve the same simplification. We will demonstrate, too, the equivalence of both methods.

E.1 Method of Longitudinal Components

In this section we will look at another possibility to reduce the original vector problem to a scalar one.

An arbitrary 3D vector function may be decomposed into a transverse 2D vector function and a longitudinal scalar function. So the electric and magnetic fields are expressed as follows:

\[ E = E_t + E_z e_z, \quad H = H_t + H_z e_z. \]  \hspace{1cm} (E.1)

In any cylindrical system the Laplacian operator \( \triangle \) can be applied separately on the transverse and on the longitudinal part of a vector \( A \), i.e.

\[ \triangle A = \triangle A_t + e_z \triangle A_z. \]  \hspace{1cm} (E.2)

Hence, the vector wave equations for \( E \) and \( H \) (4.18) are decomposed into the following two vector wave equations and two scalar wave equations

\[ \triangle E_t - \frac{1}{c^2} \frac{\partial^2 E_t}{\partial t^2} = 0, \quad \triangle H_t - \frac{1}{c^2} \frac{\partial^2 H_t}{\partial t^2} = 0, \]  \hspace{1cm} (E.3a)

\[ \triangle E_z - \frac{1}{c^2} \frac{\partial^2 E_z}{\partial t^2} = 0, \quad \triangle H_z - \frac{1}{c^2} \frac{\partial^2 H_z}{\partial t^2} = 0. \]  \hspace{1cm} (E.3b)

The equations for the longitudinal components are scalar Helmholtz equations. We can solve them first and then find the transverse components by means of Maxwell’s equations.

To solve for the transverse parts \( E_t \) and \( H_t \) we make use of (4.16) decomposing (4.17) into transverse and longitudinal part. For the rotational equations of (4.17) we obtain
and for the divergence equations

\[
\begin{align*}
\text{div} \, \mathbf{E}_t &= -\frac{\partial E_z}{\partial z}, \quad (E.4d) \\
\text{div} \, \mathbf{H}_t &= -\frac{\partial H_z}{\partial z}. \quad (E.4e)
\end{align*}
\]

We can now separate between longitudinal and transverse parts in the first two equations

\[
\begin{align*}
\text{rot}_t \, \mathbf{H}_t &= e_z \times \left( \frac{\partial \mathbf{H}_t}{\partial z} - \text{grad}_t H_z \right) = \varepsilon \frac{\partial E_z}{\partial t} + \varepsilon \frac{\partial \mathbf{E}_t}{\partial t}, \quad (E.5a) \\
- \text{rot}_t \, \mathbf{E}_t &= e_z \times \left( \frac{\partial \mathbf{E}_t}{\partial z} - \text{grad}_t E_z \right) = \varepsilon \frac{\partial H_z}{\partial t} + \mu \frac{\partial \mathbf{H}_t}{\partial t}. \quad (E.5b)
\end{align*}
\]

It is evident from these equations that if \( E_z \) and \( H_z \) are known the transverse components of \( \mathbf{E} \) and \( \mathbf{H} \) are determined. The equations are coupled for \( \mathbf{E}_t \) and \( \mathbf{H}_t \). To decouple them, we proceed analogously to Section 2.9.2: we have to differentiate them again in order to be able to separate \( \mathbf{E}_t \) and \( \mathbf{H}_t \).

We concentrate on the two equations on the right side and rewrite them as follows:

\[
\begin{align*}
\frac{\partial \mathbf{H}_t}{\partial z} + e_z \times \frac{\partial \mathbf{E}_t}{\partial t} &= \text{grad}_t H_z, \quad (E.6a) \\
\frac{\partial \mathbf{E}_t}{\partial z} - e_z \times \frac{\partial \mathbf{H}_t}{\partial t} &= \text{grad}_t E_z. \quad (E.6b)
\end{align*}
\]

We differentiate the first equation with respect to \( t \) and multiply it by \( \mu \) and
differentiate the second one with respect to $z$

\[
\mu \frac{\partial^2 \mathcal{H}_t}{\partial t \partial z} + \mu \varepsilon z \times \frac{\partial^2 \mathcal{E}_t}{\partial t^2} = \mu \nabla_t \frac{\partial H_z}{\partial t}, \quad (E.7a)
\]

\[
\frac{\partial^2 \mathcal{E}_t}{\partial z^2} - \varepsilon z \times \mu \frac{\partial^2 \mathcal{H}_t}{\partial z \partial t} = \nabla_t \frac{\partial E_z}{\partial z}. \quad (E.7b)
\]

We can now replace the term $\mu \frac{\partial^2 \mathcal{H}_t}{\partial z \partial t}$ in the second equation by means of the first one. After rearranging the terms and making use of $\varepsilon z \times (\varepsilon z \times \frac{\partial^2 \mathcal{E}_t}{\partial t^2}) = -\frac{\partial^2 \mathcal{E}_t}{\partial t^2}$ we get

\[
\frac{\partial^2 \mathcal{E}_t}{\partial z^2} - \mu \varepsilon \frac{\partial^2 \mathcal{E}_t}{\partial t^2} = \nabla_t \frac{\partial E_z}{\partial z} + \mu \varepsilon z \times \nabla_t \frac{\partial H_z}{\partial t}. \quad (E.8a)
\]

This is the wave equation for the transverse part $\mathcal{E}_t$ of the electric field. Assuming we know $E_z$ and $H_z$, which is the case since we can compute them independently by means of (E.3b), $\mathcal{E}_t$ can be found by solving the wave equation.

This wave equation is the transverse analogous to the 3D vector wave equation (2.61) and resembles the one of (2.62) where we have assumed for illustrative purposes that $\mathcal{E}$ and $\mathcal{H}$ have only one component.

A similar equation can be found for $\mathcal{H}_t$. For this, we make use of the following symmetries: by inspecting (E.6) we see that $\mathcal{E}$ and $\mathcal{H}$ can be exchanged and $\varepsilon$ and $-\mu$. Applying this symmetry to (E.8a) yields

\[
\frac{\partial^2 \mathcal{H}_t}{\partial z^2} - \mu \varepsilon \frac{\partial^2 \mathcal{H}_t}{\partial t^2} = \nabla_t \frac{\partial H_z}{\partial z} - \varepsilon \varepsilon z \times \nabla_t \frac{\partial E_z}{\partial t}. \quad (E.8b)
\]

Assuming we deal with time-harmonic signals and sinusoidal waves travelling along the waveguide’s axis $z$

\[
\mathcal{E} = E_0 e^{j\omega t \pm \gamma z} \quad \text{and} \quad \mathcal{H} = H_0 e^{j\omega t \pm \gamma z}
\]

then we obtain an explicit expression for the transverse fields ($k = \omega/c$)

\[
\mathcal{E}_t = \frac{1}{\gamma^2 + k^2} \left( \pm \gamma \nabla_t E_z + j\omega \mu \varepsilon z \times \nabla_t H_z \right),
\]

\[
\mathcal{H}_t = \frac{1}{\gamma^2 + k^2} \left( \pm \gamma \nabla_t H_z - j\omega \varepsilon \varepsilon z \times \nabla_t E_z \right).
\]

**E.2 Equivalence between both methods**

Comparing the methods of Hertzian vectors and longitudinal components shows some similarities. In both methods a $z$-directed vector component, i.e. a vector component parallel to the waveguide’s axis, is used to compute the complete electric and magnetic field. This similarity suggests that the expressions for the
Uniform waveguides

electromagnetic field found by both methods are equivalent but not necessarily identical. This can be easily shown. We will show this for the electric field.

From (4.23a) and (4.25) we find

\[ E_z = -\frac{1}{\varepsilon} \Delta_t \Phi, \quad \text{(E.9a)} \]

\[ \mathcal{E}_t = \frac{1}{\varepsilon} \text{grad}_t \frac{\partial \Phi}{\partial z} + \varepsilon \times \text{grad}_t \frac{\partial \Psi}{\partial t}. \quad \text{(E.9b)} \]

The last equation is already quite similar to (E.8a). Using (4.26a) \( E_z \) can be expressed as

\[ E_z = \frac{1}{\varepsilon} \left( \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \Phi. \quad \text{(E.10a)} \]

For \( H_z \) we can proceed in a similar way and obtain

\[ H_z = \frac{1}{\mu} \left( \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \Psi. \quad \text{(E.10b)} \]

Using this relation between the \( z \)-directed components \( \Phi \) and \( \Psi \) in the method of Hertz vectors and the longitudinal components \( E_z \) and \( H_z \) it can be easily shown that the solutions for the transverse components of the electric and magnetic field are the same for both methods. For example for the electric field we use (E.9b), insert it into (E.8a) and obtain

\[ \left( \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathcal{E}_t = \left( \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \left( \frac{1}{\varepsilon} \text{grad}_t \frac{\partial \Phi}{\partial z} + \varepsilon \times \text{grad}_t \frac{\partial \Psi}{\partial t} \right). \quad \text{(E.11a)} \]

Using (E.10a) we finally have

\[ \left( \frac{\partial^2}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathcal{E}_t = \text{grad}_t \frac{\partial E_z}{\partial z} + \mu \varepsilon \times \text{grad}_t \frac{\partial H_z}{\partial t}, \quad \text{(E.11b)} \]

which is equation (E.8a). Thus both methods yield the same transverse electric field. For the transverse magnetic field \( \mathcal{H}_t \) the demonstration is similar.

In conclusion, the difference between the longitudinal components of the Hertzian vectors and the longitudinal field components is given by (E.10).

If we assume a time-harmonic regime and sinusoidal waves propagating along \( z \) the relation (E.10) boils down to a scaling factor

\[ T^2 = \gamma^2 + k^2 \quad \text{(E.12)} \]

and reduces to

\[ E_z = \frac{1}{\varepsilon} T^2 \Phi \quad \text{and} \quad H_z = \frac{1}{\mu} T^2 \Psi. \quad \text{(E.13)} \]
APPENDIX F

Lorenz or Lorentz?
Lorenz or Lorentz?

The name of the Dutch physicist, Hendrik Antoon Lorentz, is familiar to anybody interested in Electromagnetism. Particularly well-known are his article, "Über die Grundgleichungen der Elektrodynamik für bewegter Körper," published in 1880 in the *Annalen der Physik und Chemie,* his extended paper on "La théorie électromagnétique de Maxwell et son application aux corps mouvants," written in 1892 and published in the *Archives néerlandaises des Sciences exactes et naturelles;* and his book on the *Theory of Electrons,* which came out in 1909 (and has been reprinted by Dover Publications). Lorentz was present at the birth of Relativity, and published his famous transformation formulas, which connect the 4-coordinates in an inertial frame to those in another frame, at the turn of the century. Even more familiar are the "Lorentz retarded potentials," based on the "Lorentz condition." In this instance, however, a case of mistaken paternity seems to have taken place. The undersigned was recently glancing through Whittaker's monumental *History of the Theories of Aether and Electricity,* and read with interest, on p. 268 of Volume 1, that the paternity for the retarded potentials should really be assigned to L. Lorentz, a Danish physicist who introduced them in three articles written in 1867 (Lorentz was 14 years old at the time). One of the articles, entitled "On the Identity of the Vibrations of Light with Electrical Currents," was written in English, and appeared in Volume XXXIV of the *Philosophical Magazine.* It is most interesting. The author first quotes the Kirchhoff expression for the electric field, viz.

\[ u = -2a \frac{dE}{dx} + 4 \frac{d\phi}{dt} \]
\[ v = -2a \frac{dE}{dy} + 4 \frac{d\phi}{dt} \]
\[ w = -2a \frac{dE}{dz} + 4 \frac{d\phi}{dt} \] (1)

We nowadays use the more compact form

\[ \tilde{e} = -\nabla \phi + \frac{\partial \tilde{E}}{\partial t} \] (2)

The quantity \( c \) in (1) is a given constant. Lorentz then writes the scalar potential as

\[ D = \int \frac{dx\,dy\,dz}{r} \epsilon' \left[ \frac{-\tilde{e}}{a} \right] + \int \frac{dx\,dy\,dz}{r} \epsilon' \left[ \frac{-\tilde{e}}{a} \right] \] (3)

where \( \epsilon' \) and \( \epsilon' \) stand for the volume and surface charge densities, respectively, and he gives corresponding expressions for the vector potential \( (a, \beta, \gamma) \). The relativistic effect appears clearly in (3), where the symbol \( a \) is the velocity of light. Lorentz assumes that \( c \) in (1) is \( 2a^2 \), and mentions that the best value of \( c \) known at the time was 264716 miles per second. Towards the end of the article, he proves that

\[ \frac{dE}{dt} = -2\frac{da}{dx} + 4\frac{d\phi}{dt} + \frac{d\phi}{dt} \] (4)

which is the well-known condition

\[ \nabla \cdot \tilde{a} + \frac{1}{c^2} \frac{\partial \phi}{\partial t} = 0 \] (5)

Lorentz's work is remarkable, since it was performed in parallel with that of Maxwell. The reader will probably agree that the paternity suit must be decided in Lorentz's favour. It appears that the various authors of textbooks who sinned against historical accuracy—the undersigned being regretfully one of them—should amend their references in future printings of their books?

J. Van Bladel
University of Ghent
St.-Pietersnieuwstraat 41
B-9000 Gent, Belgium

---

**The Radioscientist**

The "house" magazine of URSI, the International Union of Radio Science, includes news, views, reviews and scientific contributions of interest to radio scientists and engineers. Four issues are planned for 1991, six for 1992.

Introductory subscription rates include fast mail to anywhere in the world.

Next four issues: $10
Next ten issues: $20

Payment may be made on

VISA, bankcard, or MasterCard

by Fax: +64 3 4790964, giving signature, card number and expiry date.

Or send bank cheque/draft to

The Radioscientist
Physics Department
University of Otago
Dunedin
New Zealand

---

Lorenz or Lorentz?

J. Van Bladel
University of Ghent
St.-Pierreinnewstraat 41
B-9000 Gent, Belgium
Fax: (32) 91-64-35-93

This short (and modest) note has generated a number of reactions! The first one is the publication, in the present issue, of an interesting note by Dr. Sihvola. It concerns the paternity of another Lorenz-Lorentz common effort. The second reaction is a message by Professor J. Bach Andersen, Vice-President of URSI, who lets me know that an interesting review of Ludvig Valent Lorenz' career and scientific work, written by M. Pihl, can be found in Volume 1 of Electromagnetic Theory and Antennas.

Lorenz-Lorentz or Lorentz-Lorenz?

Ari Sihvola
Helsinki University of Technology
Electromagnetics Laboratory
Otakaari 5A, SF-02150 Espoo
Finland
Fax: 358-0-460-224

Inspired by the legitimate note by Professor Van Bladel [1], about the paternity of retarded potentials in electromagnetics, and, by the way, also the Lorenz gauge, I take the liberty to continue the discussion, and to further restore some of Lorenz's honor (and divest that from Lorentz, but on the other hand, he really cannot be claimed to be short of fame; and also, let it be clear that my admiration for him is not diminished).

My focus is on the connection between the polarizability, \( \alpha \), of a dielectric sphere, and the optical refractive index, \( n \), of the material consisting of a collection of a type of spheres. This connection,

\[
\frac{n^2 - 1}{n^2 + 2} = \text{proportional to } \alpha,
\]

has been known as the Lorenz-Lorenz formula. This label on the formula is due to the articles, both published in the year 1880, by Hendrik Antoon Lorentz [2] and Ludvig Valentin Lorenz [3]. These appeared in different issues of this respected journal, and because of this temporal order, the names also appear in the present sequence in the name of the formula. At least, this is the reason given in the famous book by Born and Wolf [4], and Classical Electrodynamics by Jackson [5] can be mentioned as an authoritative example that follows the same path.

However, the above-mentioned communication by Ludvig Lorenz was not his first one to present studies on the refractive power of molecules. As Lorenz mentions in the opening sentences in his article, his intention was to present, in the German language (to a broader scientific community), his research results which he had earlier published in Danish. His result (the expression above) can really be found in the 1869 paper Experimentale og theoretiske Undersøgelser [6]. This was one year before H. A. Lorentz matriculated at the University of Leiden, and therefore it seems clear that the "Lorenz-Lorentz formula" should be rechristened the "Lorenz-Lorenz formula."

References

Lorenz, Lorentz, and the Gauge

Robert Nevels
Department of Electrical Engineering, Texas A&M University
College Station, Texas 77843-3128 USA
Tel: +1 (979) 845-7591; Fax: +1 (979) 845-6259
E-mail: nevels@ee.tamu.edu

Chang-Seok Shin
Department of Electrical Engineering, Texas A&M University
College Station, Texas 77843-3128 USA
Tel: +1 (979) 845-8661; Fax: +1 (979) 845-6259
E-mail: shin@ee.tamu.edu

Keywords: History; electromagnetic analysis; electromagnetic fields; electromagnetic theory; potentials; gauge; Lorenz; Lorentz

Soon after development of the concept of the vector potential, \( \mathbf{A} \) and \( \phi \), and the scalar potential, \( \phi \), for which Hendrik A. Lorentz now gets credit [2]. This is especially curious, given the fact that the early researchers in our field of electromagnetics were few in number, and there were a limited number of journals in which to publish. How did it happen that such a mistake was made?

In the 1860s, the idea that the electromagnetic field could be expressed in terms of potentials was widely accepted. The scalar potential had been found in 1824 by Poisson, and the vector potential, by Neumann [3], in 1845, inferring what was later to be called the Coulomb gauge:

\[ \nabla \cdot \mathbf{A} = 0. \]

The Coulomb gauge was the basis of much of James Clerk Maxwell's work on potentials.

In 1867, Lorentz, a Dutch physicist, probably taking a hint from the earlier work of D’Alembert, published a paper [1] in which he proposed that the standard potentials of Neumann:

\[ \phi = \frac{1}{4\pi \epsilon_0} \int \frac{\rho(r) \, dr}{R}, \]

\[ \mathbf{A} = \frac{\mu_0}{4\pi} \int \frac{\mathbf{J}(r) \, dr}{R}, \]

which included the time of propagation from the source. Near the end of the same paper, he showed that a mathematical consequence of his retarded potentials was the condition given in Equation (1). However, many in the electromagnetics community at the time had strong reservations about adopting retarded potentials, because retardation implies that a physical effect – propagation at the speed of light – is being imposed on what was thought to be nonphysical and non-measurable quantities, \( \mathbf{A} \) and \( \phi \). Maxwell was less reluctant to accept the concept of retarded potentials, although in his famous treatise [4], only brief mention is made of Lorentz's proposal, concluding with the suggestion that he (Maxwell) had a print-publication claim on the idea. To quote Maxwell: "These conclusions (of Lorentz) are similar to those of this chapter, though obtained by an entirely different method. The theory given in this chapter was first published in the Phil. Trans. for 1865, pp. 459-512."

One person who did take Lorentz seriously was H. A. Lorentz, a Dutch physicist. Although much younger than Lorentz, it is interesting that in the same year, 1881 [5], these two men published papers in the same journal and on the same topic, the relation between refractive index and density in materials [6, 7]. However, Lorentz went on to publish on many subjects in mathematical physics. Some to which his name became attached include the Lorentz force law, Lorentz contraction, Lorentz invariance, and the Lorentz transformation. But starting in 1892 [8], one year after the death of Lorentz (1853-1893), his many papers supporting the concept of the retarded potential and his clear derivation of Equation (1) strongly identified his name with the gauge. It is interesting that Lorentz's work is not referenced in Lorentz's seminal paper [8], or in his later book [9], except concerning the 1880 paper on refractive index and density.

It didn't take long after the publication of Lorentz's 1892 paper, and his several subsequent articles, for the error to creep into the literature. Witness the following excerpts [3, 10]:
1927, “comes from Lorentz,” H. Thirring [15]

H. A. Lorentz (1853–1928), almost surely knew of some of these attributions, but, if he ever disputed them, or if in fact he even knew that he didn’t invent the gauge, it wasn’t reported in readily available literature. It is also probably true that many of these men had a copy of Maxwell’s treatise on their shelf, as do many of us today in the Dover edition, but evidently they didn’t read it either.

To the credit of many authors in the first half of the twentieth century, Equation (1) was simply referred to as “the condition” or “the continuity relation for the electromagnetic field.” in analogy to the continuity condition for current and charge [16]. However, in spite of many articles in the 1890s and two books [3, 10] on the history of electromagnetic theory (one [3] revised as late as 1951), in which Lorentz was recognized as the inventor of the gauge, by the end of the 1950s, the term “Lorentz gauge” was in common use by just about everyone. That is, everyone but the Danes. In his address to the 1962 Symposium on Electromagnetic Theory and Antennas in Copenhagen [17], Maugans Phil, of the University of Copenhagen, in the usual polite Danish way, pointed out that Ludwig Lorentz invented the gauge.

References


Introducing the Feature Article Authors

Robert Noyels received the BSEE degree from the University of Kentucky, the MSE degree from Georgia Tech, and the PhD from the University of Mississippi. He joined Texas A&M University in 1978 as an Assistant Professor, became Full Professor in 1993, and is currently Assistant Department Head. During the summer of 1992, he was a Visiting Professor at the Institute for Light Sources, Fudan University, Shanghai, China. Dr. Noyels served as Associate Editor of the IEEE Transactions on Antennas and Propagation from 1986-1989, and of the Wiley four-volume book series, Handbook of Microwave and Optical Components, from 1990-1992. He is currently Associate Editor of the Periodical Microwave and Optical Technology Letters. He has received nine teaching awards, most recently, the university-level Faculty Distinguished Achievement Award in Teaching. Dr. Noyels is currently a member of the IEEE Antennas and Propagation Society ACDOM, Chair of the IEEE-AP Man and Radiation Committee, and General Chair of the 2002 IEEE International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting. He is a member ofEta Kappa Nu, Sigma Xi, Commission E of USNC/URSI, the Electromagnetics Academy, and the American Scientific Affiliation. Dr. Noyels’ interests are in analytical and numerical techniques for electromagnetic scattering, antennas, and microwave-device design.