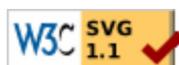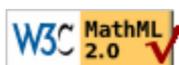**Multimedia textbook**

Electromagnetic waves

Microwave technique

Zbyněk **RAIDA**, Dušan **ČERNOHORSKÝ**, Dalimil **GALA**, Stanislav **GOŇA**, Zdeněk **NOVÁČEK**, Viktor **OTEVŘEL**, Václav **MICHÁLEK**, Vlastimil **NAVRÁTIL**, Tomáš **URBANEC**, Zbyněk **ŠKVOR**, Petr **POMĚNKA**, Jiří **ŠEBESTA**, Geert **VANDERSTEGEN**, Bart **VANDIJCK**, Bert **SOORS**, Jeroen **SCHEVERNELS**, Javier **MARTÍN DEL VALLE**, Martin **ŠTUMPF**, Vladimír **ŠEDĚNKA**, Peter **KOVÁCS**, Jaroslav **LÁČÍK**, Jana **JILKOVÁ**, Zbyněk **LUKEŠ**, Michal **POKORNÝ**.

Chapter 1: Multimedia textbook

Introduction

The life of today's society requires transmitting and receiving more and more data. Since the capacity of today's communication channels is not sufficient, new channels in higher and higher frequency bands have to be built.

Design of electronic circuits, antennas and other systems, which are required to operate at frequencies of tens to hundreds gigahertz, is rather difficult. This is given by the fact that phenomena at the above-specified frequencies are of wave nature, and therefore, all the analysis and design have to be based on Maxwell equations in differential or integral form. Understanding phenomena, which are described by differential or integral equations, makes usually troubles to students.

A relative complexity and abstractness of the mathematical description is the first cause of troubles (how can students imagine curl or divergence operators, special integrals, etc.). Handling complex vector operations, one can hardly imagine their concrete sense and practical signification.

A relatively complicated and abstract nature of wave phenomena is the second cause of troubles (how can students imagine EM wave propagation along the microstrip transmission line). Properties of studied phenomena and systems and their mutual relations cannot be observed in a direct way, but only by the indirect one.

The electronic textbook we have in our hands now, tries to solve the above-described problems. The textbook is going to explain, using simple examples, the practical signification of complicated mathematical operators. The textbook is going to visualize the studied phenomena in order to build a right notion about their matter and their mutual relations. The textbook is going the ways of exploiting examined phenomena in the engineering practice.

The textbook is conceived as a set of selected topics, which are harder understandable.

The electronic textbook is not a classical textbook describing the studied matter in a complex way. Our textbook assumes the reader, who is equipped by the basic knowledge of electromagnetic field, who would like to improve this knowledge and apply it in the solution of practical problems.

The content of the textbook is divided into six main topics if this introduction is not considered. Each topic is subdivided to subchapters. Within each subchapter, the following layers can be found:

The layer A is conceived as a continuous text containing primarily the verbal description of the studies phenomenon. The text is free of complex mathematical expressions and hardly-understandable derivations. The layer is aimed to reveal the matter of the phenomenon as simply as possible. The layer is created for bachelor students dominantly. If the reader is interested in the deeper description, he can follow the link to the second level of the textbook (the layer B).

The layer B describes all the mathematical derivations and their detailed discussions. Even this second layer is conceived as a continuous text, which expects an experienced reader (typically the master student). The chapters, which are thought by authors as relatively simple, contain all the descriptions in the layer A; the layer B presents the same descriptions in English. The reader, who does not understand the English text well, can jump to the parallel Czech version in order to verify his correct understanding of the English text.

The layer C offers the download of the numerical model of the studied phenomenon in the form of m-files of MATLAB. The layer contains a user's description of the numerical model. Thanks to the standard Windows user interface of numerical models, the reader does not need to work with the source code of the model.

The layer D contains the programmer's description of the numerical model. Students can use m-files containing the computational kernel (the code without the user's interface), modify and extend them.

The layer E contains the numerical models of studied phenomena in the form of JAVA applets. The applets can be viewed directly on the webpage without the necessity to buy MATLAB.

Fig. 1.1 *Navigation menu*

The layer F enables readers to test the proper understanding of studied phenomena. In the layer, pentads of questions are available, and the reader is asked to select the correct answer. After answering the pentad of questions, the answers are evaluated. If the question is answered incorrectly (or is not answered at all), the correct answer is displayed. The test is concluded by the complex evaluation.
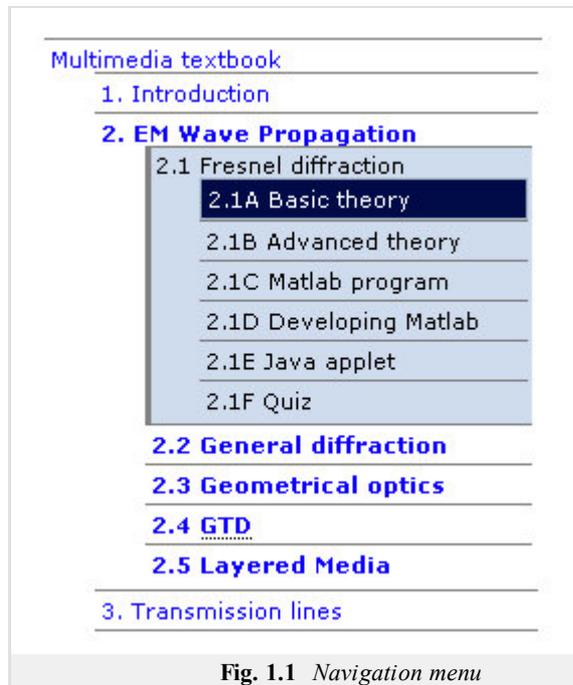
The authors of the electronic textbook of electromagnetic waves and microwave techniques hope that the textbook becomes a useful tool helping students to understand well principles of examined phenomena. This aim is conditioned by the continuous improvement and completion of its contents, which are based on the recommendations of readers. Therefore, we kindly ask all the readers for their opinion on the textbook,

for their comments of the form and contents and for their error warnings.

We thank the readers for their help and wish them a pleasant time with the textbook.

On behalf of the authors,

Zbynek Raida
raida@feec.vutbr.cz

1.1 Before you start studying...

... we would like to give you some practical information, which can make reading the textbook simpler and more pleasant to you.

The textbook uses following markup languages: XHTML for text and MathML for equations. There is special SVG layer included as well. It is used for better image description. Your browser should support at least XHTML+MathML, SVG support isn't necessary.

The textbook was carefully tested in web browsers Internet Explorer (8), Mozilla Firefox (3.5), Opera (v. 9.64), Google Chrome (2.0) and Safari (4.0).

Tab. 1.1.1 *Support in browsers*					
Support	**MSIE 8.0**	**Firefox 3.5**	**Opera 9.64**	**Chrome 2.0**	**Safari 4.0**
XHTML	yes	yes	yes	yes	yes
MathML	yes*	yes**	yes	no	no
SVG	no	yes	yes	partial	yes
* with MathPlayer plug-in					
** you can get better support by installing STIX fonts					

Reading the textbook, you might be refereed to another layer of the textbook (e.g., if you read the layer **A** and you would like to get more detailed information provided in the layer **B**, you simply click to a respective link). Clicking the button "Back" of your browser or using the navigation menu, you return back to the layer **A**.

The index of the textbook can work in two modes. The first mode is activated by clicking the highlighted text (the indexed term). That way, a web page explaining the indexed term is opened. The reader can return back from the index by clicking the link "Back". We recommend to prefer the button "Back" of the browser. If the reader is interested in a rather systematic reading of the contents of the index, he can enter the index via the link Register in the menu item "Appendices".

Simpler terms can be explained by the bubble help. The bubble is displayed when the reader puts the cursor to the highlighted indexed term and waits for a while. The functionality of the bubble help is dominantly influenced of the properties of the used browser (the bubble can disappear after a certain time period, or the bubble can contain a part of the explanation text only). The Opera browser is recommended because the whole text is displayed for the unlimited time. Firefox users can use "No Tooltip Timeout" add-on.

References of the recommended literature are based on the same principles. The systematic reading of references is available via the link References in the menu item "Appendices".

Simulation programs, which are added to most chapters, are zipped. You have to save them on your disk and unzip them. That way, a separate folder is created, which contains all the files of the program. Then, you have to start MATLAB and run the main m-file.

If you find some problems, do not hesitate and send us an email, please.

We wish you many successful clicks...

Chapter 2: Propagation of electromagnetic waves

Introduction

When studying dynamic electromagnetic fields, we usually start with the propagation of waves in an infinite homogeneous medium, which is supposed being linear and isotropic. In this situation, a single wave is propagating, which can be attenuated only is lossy medium is assumed.

Electromagnetic waves, which propagate in infinite homogeneous medium, can be classified according to their equiphase surface as planar, cylindrical and spherical.

Plane waves can be observed in a long distance from the transmitting antenna. Plane wave propagation can be mathematically described by a scalar differential equation in the Cartesian coordinate system.

Cylindrical wave propagates from an infinitely long direct wire, which is flown by a high-frequency current. Cylindrical wave propagation is described by a scalar differential equation in the cylindrical coordinate system.

Spherical wave propagates from a point source. Spherical wave propagation is described by a vector differential equation in the spherical coordinate system.

Propagation of the above-described waves is relatively well-understandable and well-imaginable. Even the mathematical relations describing propagation of those waves are relatively simple, and moreover, the final equations are of a closed form (i.e., we get relatively simple formulae which can be easily used for practical calculations). Therefore, those topics are not discussed in our electronic textbook.

Theoretical description of wave propagation in homogeneous medium is given in **[1]** and **[2]**. Practical relations for the solution of given engineering tasks are in **[3]** at your disposal.

Examination of electromagnetic wave propagation complicates in the situation when some inhomogeneity appears in the medium. The inhomogeneity can cause wave reflection, wave scattering or wave diffraction. Reflected (scattered, diffracted) waves interfere with incident waves. In the surrounding of the inhomogeneity, regions with higher intensity (compared to the case without inhomogeneity) can appear (*primary wave* and *secondary one* are of the same phase). Next, regions with very low intensity can be found (*primary wave* and *secondary one* are of the opposite phase).

Therefore, computation of the wave propagation in an inhomogeneous medium is much more complicated than is a homogeneous one, and even understanding the mathematical description and building a proper notion is much harder. Therefore, we are going to study the related phenomena more in detail here.

In an inhomogeneous medium, its electric and magnetic parameters (permittivity, conductivity, permeability) change from a region to another region. Moreover, variations of parameters can be continuous or discontinuous.

The discontinuous variation of parameters is characteristic for the surface of objects, i.e. for an medium containing conductive or dielectric (or ferromagnetic) objects. In a real situation, buildings, trees, hills, cars, people or raindrops play the role of those objects. In our textbook, we start to examine those phenomena in an example of a thin planar sheet, which neither reflects nor transmits the incident wave. We are speaking about Fresnel diffraction and we describe it in chapter 2.1.

In the practical life, more realistic objects have to be considered. Including those considerations to our computations is enabled by the general theory of diffraction, which is described in chapter 2.2. In this chapter, the diffraction on an infinitely long perfectly conducting circular cylinder is discussed.

In chapter 2.3, geometric optics is introduced. Geometric optics is an efficient method for the computation of wave phenomena in more complicated media. This method adopted the conception of wave propagation along *beams*. Comparing to the classical geometric optics, we can compute the intensity variations and wave polarization for wave propagation in the medium with continuously changing parameters.

Chapter 2.4 deals with the geometric theory of diffraction (GTD). GDT extends the abilities of the geometric optics to the media with discontinuous inhomogeneities. GTD solves interaction of beams and objects, and laws of geometric optics are modified in order to eliminate obvious errors (intensity discontinuities on the border between the reflection and shade) on one hand, and to preserve advantages of the geometric optics on the other hand (conception of beams).

All the section is closed by the examination of *layered media* (media consisting of several layers of different permittivity and different thickness). Borders of layers are usually planar and parallel. Hence, the layered media consists of a certain number of plan-parallel layers. More information is given in chapter 2.5.

2.1 Diffraction on a planar absorbing object

Basic theory

We are going to study the wave diffraction in the following situation. Between the wave source V and the observation point P, a thin planar object, which neither reflects nor transmits electromagnetic waves, (so called absorbing object) is placed. In the object, a small window is cut. Through the window, the wave can propagate from the source to the observation point. The object can have a form of a half-plane as depicted in fig. **2.1A.1** (the free upper half-plane plays the role of the window here). Computing the intensity in the observation point P is our task.

The described situation approximates a certain category of practical problems (wave propagation behind a chine, e.g).

For the first time, A. J. Fresnel (1788 - 1827) solved the above-formulated task. He elected a relatively simple approach we are going to explain now. The Fresnel solution is usually called the Fresnel diffraction.

Fig. 2.1A.1 *Diffraction on a half plane*

The plane of the object is signed S. The plane is assumed being perpendicular to the line VP. No matter the problem is solved for the half-plane (fig. **2.1A.1**), the approach can be extended even to differently shaped *windows*.

First, electric-field intensity $E^{(S)}$ is computed in the plane of the object S the same way as the object is absent. More, the spherical wave is assumed to propagate from the source V (fig. **2.1A.1**), and therefore

$$E^{(S)} = C \frac{e^{-jkr_1}}{r_1}. \qquad (\text{ 2.1A.1 })$$

Here, C is a constant dependent on the radiated power and k denotes wave number.

Now, Huygens principle is applied. The plane S is left-illuminated and every its element dS is considered to be a radiation source for the right-hand half-space. Illumination intensity $E^{(S)}$ is given by eqn. (2.1A.1) for the region lying above the edge of the object (half-plane S_1), but $E^{(S)} = 0$ below the edge due to the absorption nature of the object.

Each element of the free part of the plane S can be therefore considered as Huygens source illuminated by the wave of the intensity (2.1A.1). Now, we have to perform summation of field intensities created by those sources in the point P. Using the relation for the field intensity created by the Huygens source, we get

$$E^{(P)} = \frac{j}{\lambda} \int_{S_1} E^{(S)} \cos(n, r_2) \frac{e^{-jkr_2}}{r_2} dS. \qquad (\text{ 2.1A.2 })$$

Integration is performed on the free (transmitting) part of the plane S (in the integral, we formally sign that as S_1). From the physical point of view, the problem is solved out.

In order to solve the integral (2.1A.2), Fresnel had to accept some simplifications. All these simplifications are based on the assumption that the most important contribution to the intensity in P is caused by the Huygens sources lying near to the upper edge of the object (i.e., even relatively near to the origin O as depicted in fig. **2.1A.1**). For those source, the angle between the normal n to the object plane and the radius vector r_2 to the observation point is small, and therefore, we can assume $\cos(n, r_2) = 1$. Next, the coordinates x, y of single Huygens sources can be assumed to be (numerically) small comparing to distances d_1 and d_2. Exploiting rules for small-numbers computation, we get

$$r_1 = \sqrt{d_1^2 + x^2 + y^2} = d_1 \sqrt{1 + \left(\frac{x}{d_1}\right)^2 + \left(\frac{y}{d_1}\right)^2} \cong d_1 \left[1 + \frac{1}{2}\left(\frac{x}{d_1}\right)^2 + \frac{1}{2}\left(\frac{y}{d_1}\right)^2\right]. \qquad (\text{ 2.1A.3 })$$

Finally, we assume $r_1 = d_1$ and $r_2 = d_2$ in the denominators of the fractions (2.1A.1) a (2.1A.2) because these assumptions negligibly influence the amplitude value of the total electric field intensity only. Performing these simplifications, the integral (2.1A.2) becomes solvable and the result can be expressed (for simpler shapes of the *window* S_1) in closed, relatively simple form. The details are given in the layer B of the textbook.

Using numerical methods, the integral (2.1A.2) can be solved without simplifications. Nevertheless, numerical methods are usually exploited in the case of complicated shapes of the window only. For the half-plane diffraction, the Fresnel (analytical) solution is simple, well understandable and its accuracy is sufficient for most applications.

Let us remind the reader that the Fresnel diffraction considers:

1. a planar absorbing object,

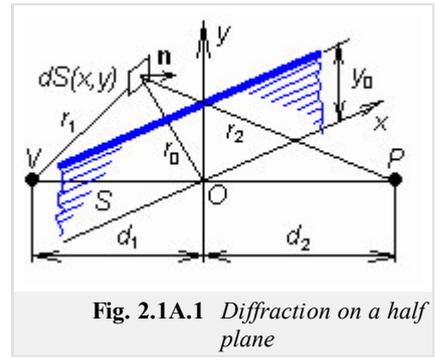2. an application of Huygens principle,

3. the above-described simplifications used when analytically solving an integral.

The transmitting *window* can be of an arbitrary shape.

In the following, we consider a half-plane object (fig. **2.1A.1**).

Exploiting the above-described approach, the field intensity in the observation point *P* can be computed depending on the distances d_1 and d_2, on the upper-edge height y_0 (which can be even negative if the edge is below the line *VP*), on the wavelength and on the source properties *V*. In order to reduce the number of input quantities, field intensity is usually related to the field intensity in the same point in the situation when the object is removed; we therefore compute the ratio *E* (behind the object) / *E* (in free space). Moreover, the upper-edge coordinate y_0 is not given in meters but in Fresnel zone radii r_{0n}. Then the result is independent on values of d_1, d_2 and λ. In the object plane, Fresnel zones are of the shape of annuli (the border between the neighboring zones are circles with the common center in the intersect of the line *VP* and the object plane (*O*). Fresnel zone radii can be computed according to
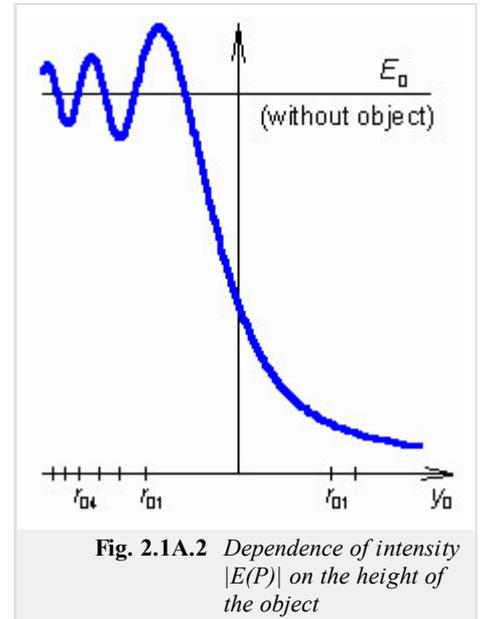

Fig. **2.1A.2** *Dependence of intensity |E(P)| on the height of the object*

$$r_{0n} \cong \sqrt{n}\sqrt{\lambda \frac{d_1 d_2}{d_1 + d_2}}, \qquad (2.1A.4)$$

which accurate enough if the computed radius is small with respect to the distances d_1 and d_2. More detailed information about Fresnel zones is given in the layer B.

Relative field-intensity (E/E_{vp}) dependency on the edge height y_0 is depicted in fig. **2.1A.2**. We can see that the half-plane influences field-intensity in *P* even if y_0 is negative and the line *VP* is not interrupted by the object. If y_0 goes from negative values to zero, field intensity rises and drops down in cyclic way, and the variation amplitude increases up to nearly 3 dB. If $y_0 = r_{01}/2$ (approximately one half of the first Fresnel zone is free between the line *VP* and the edge), magnitude of the intensity is the same as in free space without objects. If the edge is nearer to the line *VP* or even above it (y_0 is positive, the line *VP* is interrupted by the object), the object shades and the field intensity monotonously decreases. For $y_0 = 0$, $E/E_{vp} = 1/2$.

A cyclic variation of field intensity for $y_0 < 0$ is a rather interesting phenomenon, which can be observed even experimentally both on radio frequencies and on optical ones. In literature, a cyclic covering and uncovering Fresnel zones explain the phenomenon. Unfortunately, this explanation is not accurate and well understandable. Exact explanation is presented in the layer B. Graphical representation is provided by a program, which is introduced in the layer C.

Finally, a small note. In the surrounding of a well conducting object, which does not transmit and absorb electromagnetic waves but reflect them, the diffraction phenomena are of different nature. This situation is described in the chapter 2.2.

2.1 Diffraction on a planar absorbing object

Advanced theory

Assume a planar object, which is perpendicular to the line *VP* (wave source - observation point) and which is partially covered by a perfectly absorbing matter. The rest of the plane is free, and therefore, waves can past through this area to the half-space containing the observation point *P*. Computing field intensity in *P* is our task. Except of the marginal case (all the plane is free), we deal with the diffraction task.

A classical solution of the above-formulated task was provided by Fresnel (*Fresnel diffraction*). Fresnel assumed the planar obstacle (field intensity of its reverse side) as a wave source for the half-plane containing the observation point *P*. Hence, Fresnel understood the planar obstacle as a large planar antenna. Next, Fresnel expected the intensity on the reverse side being zero in points covered by the absorbing matter and being non-zero (uninfluenced by the obstacle) in points above the obstacle edge. Both the assumptions are inaccurate but the reality usually approaches

Fig. 2.1B.1 *Half-plane diffraction*

them. Thanks to this conception, the diffraction task is well solvable as two independent tasks: wave propagation from the source to the obstacle plane and wave propagation from free regions of the plane to the observation point. That way, the obstacle was removed pro computations and the result is influenced by the shape of free regions only.

In the basic theory, Fresnel solution for the half-plane obstacle (fig. **2.1B.1**) was briefly presented.

Since spherical wave propagates from the source *V*, field intensity on the obstacle plain is

$$E^{(S)} = C \frac{e^{-jkr_1}}{r_1}, \qquad (2.1B.1)$$

where *k* is wave number, *C* is a constant dependent on a radiated power and a directivity factor of the transmitting antenna, and where the distance r_1 is given by $r_1 = \sqrt{d_1^2 + x^2 + y^2}$. Symbols *x*, *y* denote coordinates of a point on the obstacle plane. Every point (*x*, *y*), or every element *dS* = *dxdy* is left-illuminated by the intensity (2.1B.1) and it becomes (according to Huygens principle) the wave source for the right-hand half-space. It contributes to the intensity in *P* by $dE^{(P)} = \frac{j}{\lambda} E^{(S)} \cos(n, r) \frac{e^{-jkr}}{r} dS$.

Integrating all the contributions of the free part of the obstacle plane (i.e. $-\infty < x < +\infty$, $y_0 < y < +\infty$) we yield the final intensity in *P*

$$E^{(P)} = \frac{j}{\lambda} \int_{y_0}^{+\infty} \int_{-\infty}^{+\infty} E^{(S)} \cos(n, r_2) \frac{e^{-jkr_2}}{r_2} \, dydx. \qquad (2.1B.2)$$

Let us note that changing the shape of the *window* in the obstacle plane changes integration limits only.

In order to finish the computations, Fresnel proposed some simplifications, which are characteristic to his approach. They are mostly based on the assumption that the strongest contribution to the field intensity in *P* is given by elementary sources *dxdy* lying near to the line *VP*, i.e. sources lying on coordinates specified by inequalities $x \ll d_{1,2}$ and $y \ll d_{1,2}$. Then, we can assume $\cos(n,r_2) = 1$, and in denominators $r_1 = d_1$, $r_2 = d_2$. In phase terms, distances r_1 and r_2 have to be expressed more accuratelly

$$r_1 = \sqrt{d_1^2 + x^2 + y^2} = d_1 \sqrt{1 + \left(\frac{x}{d_1}\right)^2 + \left(\frac{y}{d_1}\right)^2} \cong d_1 \left[1 + \frac{1}{2}\left(\frac{x}{d_1}\right)^2 + \frac{1}{2}\left(\frac{y}{d_1}\right)^2\right]. \qquad (2.1B.3)$$
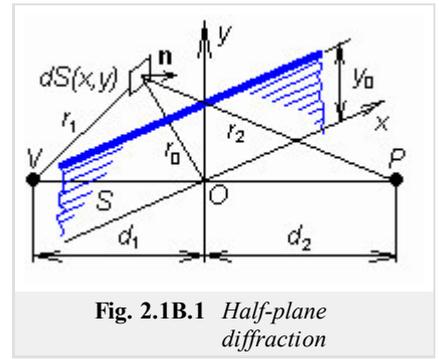
Performing the described rearrangements, we get

$$E^{(P)} = C \frac{j}{\lambda} \int_{y_0}^{\infty} \int_{-\infty}^{\infty} \exp\left(-jax^2\right)\exp\left(-jay^2\right)dxdy \qquad (2.1B.4)$$

where

$$a = \frac{k}{2} \frac{d_1 + d_2}{d_1 d_2}. \qquad (2.1B.5)$$

Since in the integrand of (2.1B.4), every term is a function of a single variable, knowledge of the solution of the integral $\int \exp\left(-jau^2\right)du$, $u = x, y$ is sufficient. Substituting

$$u = v\sqrt{\frac{\pi}{2a}} \qquad\qquad (\text{2.1B.6})$$

we transform the above integral to Fresnel integrals

$$\int\limits_0^v \exp\left(-j\frac{\pi}{2}v^2\right)dv = \int\limits_0^v \cos\left(\frac{\pi}{2}v^2\right)dv - j\int\limits_0^v \sin\left(\frac{\pi}{2}v^2\right)dv = C(v) - jS(v). \qquad\qquad (\text{2.1B.7})$$

Values of Fresnel integrals $C(v)$ and $S(v)$ are tabularized, are programmed and are graphically expressed as coordinates of the curve depicted in fig. **2.1B.2**. The curve is called klothoidy. Numbers on the curve are argument values v and respective values of integrals $C(v)$ and $S(v)$ are read on axes. The right (upper) branch ends at the point $C(v=\infty) = 0.5$, $S(v=\infty) = 0.5$ and the left branch is symmetric with respect to the origin.

The final solution of the task is of the form

$$E^{(P)} = C\,\frac{\exp\left[-jk(d_1+d_2)\right]}{d_1+d_2}\,\frac{\sqrt{2}}{4}\exp\left(\frac{j\pi}{4}\right)\left\{\left[1 - 2C(v_0)\right] - j\left[1 - 2S(v_0)\right]\right\}, \qquad v_0 = y_0\sqrt{\frac{2a}{\pi}}. \qquad (\text{2.1B.8})$$

Dependency of the field intensity in P (behind the obstacle) on the obstacle height y_0 was discussed in the basic theory. We remind here that the obstacle influences the wave propagation even if the direct line VP is not interrupted and that the obstacle can cause even a small increase of intensity (comparing to free-space propagation). If the distance between the obstacle edge and the line VP becomes smaller than one half of the first Fresnel zone then field intensity starts to decrease below the level given by free-space propagation.

Since in the integrand of the integral (2.1B.4), every term is a function of a single variable, the integral can be rewritten as a product of two independent integrals: the first one according to dx, the second one according to dy. The integral according to dx is of infinite limits, its value is independent on the obstacle height and is constant for the given situation. I.e., field intensity in P is proportional to the value of the integral according to y, and therefore, using the substitution (2.1B.6)

$$\int\limits_{v_0}^{\infty} \exp\left(-j\frac{\pi v^2}{2}\right)dv = \left[C(\infty) - jS(\infty)\right] - \left[C(v_0) - jS(v_0)\right], \qquad v_0 = y_0\sqrt{\frac{2a}{\pi}}. \qquad (\text{2.1B.9})$$

The right-hand side of (2.1B.9) is a difference of two complex numbers in a complex plane $C(v)$, $jS(v)$, i.e. in a klothoidy plane in the fig. **2.1B.2**, where the vertical axis plays the role of the imaginary one.

Difference of coordinates of two points in a complex plane equals to the abscissa length from the first point to the second one. Finally, the field intensity in P is proportional to the abscissa length from $[C(v_0), S(v_0)]$ to $[C(\infty), S(\infty)]$ on the klothoidy. The point $[C(\infty), S(\infty)]$ is a point of coordinates 0.5, 0.5 and is fixed. If the upper edge of the obstacle rises to the line VP, the point $[C(v_0), S(v_0)]$ being on the left (down) klothoidy. Klothoidy branch travels along the curve towards the origin and the abscissa length rises and decreases in cyclic way. When $[C(v_0), S(v_0)]$ leaves the last spire of the left branch, oscillations stop and both the abscissa length and the field-intensity monotonously decrease. These phenomena are demonstrated by a computer program, which is described in the layer C.
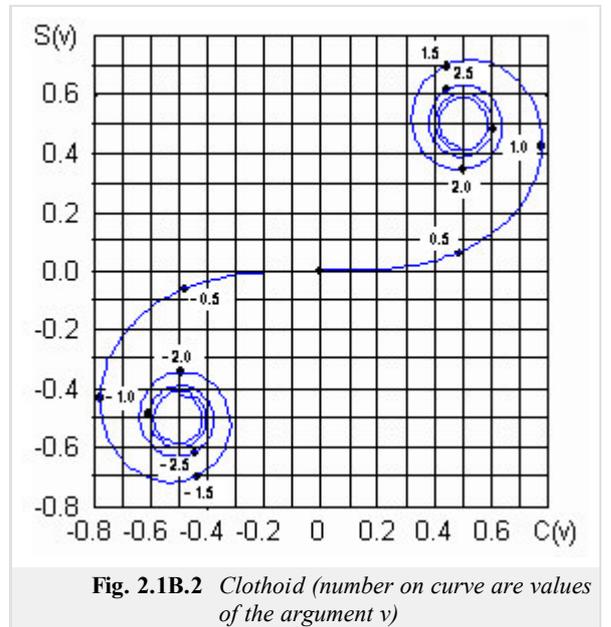
In order to illustrate the influence of Fresnel zones to the field intensity behind the obstacle, we consider again the general result (2.1B.2). Instead of coordinates x, y of a facet dS, we introduce a radial distance $r_0 = \sqrt{x^2 + y^2}$ of the facet from the origin. Then,

Fig. 2.1B.2 *Clothoid (number on curve are values of the argument v)*

$$E^{(P)} = C\,\frac{\exp\left[-jk(d_1+d_2)\right]}{d_1+d_2}\,\frac{j}{\lambda}\int\limits_S \exp\left(-jar_0^2\right)dS. \qquad (\text{2.1B.10})$$

We can see that contributions of facets in the different distance from the origin are of the different phase ($-ar_0^2$). In the distances r_{01}, r_{02}, ... r_{0n} given by

$$ar_{0n}^2 = n\,\pi \qquad\qquad (\text{2.1B.11})$$

the phase is retarded for 180°. This is caused by enlarging the pointed line V - dS - P for 1/2. The radii r_{0n} in eqn. (2.1B.11) are the radii of the

external border of given Fresnel zones. Substituting a from (2.1B.5) to (2.1B.11), we get a result, which is identical with (2.1A.4).

The fact that the phase of contributions of neighboring Fresnel zones (on the plane S) differs for 180° has got interesting causes. Let us imagine that the absorbing obstacle cover all the plane S. To this covered plane, a circular *window* with the center in O and with the arbitrary radius r_0 is cut. Though this *window*, wave propagates to the point P. Computing intensity $E^{(P)}$, we use eqn. (2.1B.10) and thanks to the rational symmetry of the window, we perform integration on S (over *window*) in polar coordinates: $dS = r_0\, dj dr_0$:

$$E^{(P)} = C\,\frac{\exp\left[-jk(d_1+d_2)\right]}{d_1 d_2}\,\frac{j}{\lambda}\int\limits_{0}^{r_0}\int\limits_{0}^{2\pi} r_0 \exp\left(-jar_0^2\right)\mathrm{d}\varphi\mathrm{d}r_0.$$

(2.1B.12)

The integral is easily solvable using the substitution $r_0^2 = r$. Inscribing by

$$E_0 = C\,\frac{\exp\left[-jk(d_1+d_2)\right]}{d_1 d_2}$$

(2.1B.13)

the free-space field intensity in P, following results are obtained for different *window* radii r_0:

Tab. 2.1B.1 *Intensity values for different window radios*		
Window radio r_0	**Zones**	**Intensity $E^{(P)}$**
∞	All the space free	E_0
$0.58\, r_{01}$	Apart of the 1$^{\text{st}}$ Fresnel zone free	E_0
r_{01}	All the 1$^{\text{st}}$ Fresnel zone free	$2E_0$
r_{02}	The 1$^{\text{st}}$ Fresnel zone and the 2$^{\text{nd}}$ one free	0

Observing results, we might conclude that field intensity E_0 we measure in P without the obstacle present, is created by contributions of a free part of the first Fresnel zone and that contributions of the rest of S mutually eliminate. Unfortunately, the conclusion is not correct (we cannot determine the part of Fresnel zone producing the contributions which are not eliminated). On the other hand, covering single zones can suppress the mutual elimination and field intensity can be increased in P that way. Covering all the even Fresnel zones (all the odd ones) yields $E^{(P)} = \infty$. Such coverage acts as a burning glass, which focuses the wave to point P.

2.1 Diffraction on a planar absorbing object

Matlab program

The program computes attenuation caused by a thin obstacle, and Fresnel zones radii.

First, the path of Matlab has to be set to the folder **Fresnel**. The program is run using the m-file `fresdif`. Passing the introductory window of the program, the reader is asked for input parameters (frequency, distance of the obstacle from the source and the observation point). Choosing the button Display manually , the height of the obstacle can be changed manually by the scrollbar, and the changing transmition can be observed. Choosing the button Display automatically , transmition changes without user interaction. If one of the above-described two buttons is chosen, a new window is opened. In the upper-left corner, a clothoid is displayed. In the upper-right corner, the dependency of the attenuation on the ratio H/r_1 (distance between the obstacle edge and the line *VP* over the first Fresnel zone radius) is depicted. In the down left corner, a sequential covering of Fresnel zones is shown, and in the down right corner, important parameters of the task are listed. Moreover, we can find here the button Finish (which ends the program) and the button Again (which enables to input new data and perform new computation).

2.1 Diffraction on a planar absorbing object

Developing Matlab

The program is based on the principle of the Fresnel diffraction. Therefore, Fresnel integrals have to be evaluated, and then, the attenuation of the wave propagation can be computed. The Fresnel integrals are computed the following way:

```
sig=sign(v);
v=abs(v);
fv=(1+.926*v)/(2+1.792*v+3.104*v^2);
gv=1/(2+4.142*v+3.492*v^2+6.67*v^3);
    if v==0 v=1e-8; end
av=pi/2*v^2;
cx(n)=(.5+fv*sin(av)-gv*cos(av))*sig;
sx(n)=(.5-fv*cos(av)-gv*sin(av))*sig;
v=v*sig;
```

Values of `cx` and `sx` are later used for computing the attenuation:

```
b=sqrt((cx(n)-0.5)^2+(sx(n)-0.5)^2)/sqrt(2)*100;
h=sv/r1;
db=8.68*log(b/100);
```

In the program, radii of the first four Fresnel zones are computed.

```
r1=round(10*sqrt(la*h_prijimac*h_vysilac/(h_prijimac+h_vysilac)*1000))/10;
r2=round((r1*14.14))/10;
r3=round((r1*17.32))/10;
r4=round((r1*20))/10;
```

2.1 Diffraction on a planar absorbing object

Java applet

2.1 Diffraction on a planar absorbing object

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**10** of **10**:

Question 1

Fresnel diffraction describes a wave phenomenon caused by …

Possible answers for question 1:
- ○ … planar perfectly electrically conductive obstacle between transmitter and receiver.
- ○ … planar perfectly absorbing obstacle between transmitter and receiver.
- ○ … cylindrical dielectric obstacle between transmitter and receiver.
- ◉ I won't answer this question

Question 2

Applying Huygens principle to the solution of Fresnel diffraction …

Possible answers for question 2:
- ○ … surface currents on the obstacle, which is induced by the incident wave, is computed as the source of the secondary wave.
- ○ … does not produce the solution of the problem.
- ○ … each elementary surface over the obstacle, which is illuminated by the incident wave,becomes the source of the secondary wave.
- ◉ I won't answer this question

Question 3

Inserting an obstacle between the transmitter and the receiver …

Possible answers for question 3:
- ○ … field intensity at the receiver can never be higher compared to the situation without any obstacle.
- ○ … field intensity at the receiver can be even higher compared to the situation without any obstacle.
- ○ … does not significantly influences wave propagation.
- ◉ I won't answer this question

Question 4

Field intensity at the receiver depends on …

Possible answers for question 4:
- ○ … the distance between the top of the obstacle and the join between the transmitter and the receiver.
- ○ … electromagnetic parameters of the obstacle.
- ○ … frequency of the electromagnetic wave.
- ◉ I won't answer this question

Question 5

In the practical life, Fresnel diffraction …

Question 6

When the top of the obstacle touches the join between the transmitter and the receiver …

Possible answers for question 6:

- ⊙ … field intensity at the receiver is the same compared to the intensity without any obstacle.
- ⊙ … field intensity at the receiver is higher compared to the intensity without any obstacle.
- ⊙ … field intensity at the receiver is one half compared to the intensity without any obstacle.
- ⦿ I won't answer this question

Question 7

The first Fresnel zone is given by the circle on the plane, which is perpendicular to the join between the transmitter and the receiver.

Possible answers for question 7:

- ⊙ The distance transmitter - circumference of the circle - receiver is half wavelength longer compared to the distance transmitter - receiver.
- ⊙ Radius of the circle equals to the distance between the transmitter and the receiver.
- ⊙ The distance transmitter - circumference of the circle - receiver is one wavelength longer compared to the distance transmitter - receiver.
- ⦿ I won't answer this question

Question 8

When all the even (odd) Fresnel zones are covered …

Possible answers for question 8:

- ⊙ … zero field intensity can appear at the receiver.
- ⊙ … theoretically infinite field intensity can appear at the receiver.
- ⊙ … nothing can be said about the field intensity at the receiver.
- ⦿ I won't answer this question

Question 9

Fresnel integrals can be evaluated …

Possible answers for question 9:

- ⊙ … analytically.
- ⊙ … cannot be evaluated due strong singularities.
- ⊙ … numerically.
- ⦿ I won't answer this question

Question 10

When explaining Fresnel diffraction, we assume the propagation of …

see result

2.2 General Theory of diffraction

Basic theory

Fresnel diffraction gives correct results for planar, thin and absorbing obstacles only. Diffraction phenomena on three-dimensional objects have to be therefore computed a different way. Hence, we have to study the basic steps of the general solution of diffraction tasks.

An objects of a known shape and of known electrical properties (permittivity, conductivity, etc.) is illuminated by a wave, which intensity is known in every point of the environment. This is so called primary wave. Reacting to the *primary wave*, the object is polarized (if the object is of dielectric nature) or currents are induced on the surface of the object (if the object is conductive). As a cause of polarization or current induction, the object becomes a source of a new wave (*secondary wave*). The secondary wave propagates in all the directions from the object. The total field intensity in the object surrounding (even behind the object) is then given by the summation of the primary wave and secondary one. Mathematical description of the secondary wave can be obtained by solving homogeneous wave equation. Magnitude of the secondary wave can be determined from the requirement that the total field intensity on the object surface (summation of the primary wave and the secondary one) meets the boundary condition. The algorithm of the solution is shown on an example. We are going to compute plane-wave diffraction on an infinitely long perfectly conducting circular cylinder.
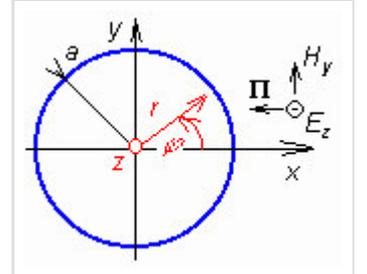
Fig. 2.2A.1 *Diffraction on cylinder*

The cross-section of the cylinder is depicted in figure **2.2A.1**.

The cylinder axis identical with the *z* axis of the Cartesian coordinate system, the cylinder radius is denoted as *a*. The cylinder is illuminated by the plane wave propagating in the direction -*x* and consisting of the component:

E_z: $E_{z\ prim} = E_0 \exp(jkx)$ only. Since the situation is identical in all the planes parallel to the plane *xy*, we solve the task in two dimensions. In the next step, we transform the problem from the Cartesian coordinate system *x*, *y* to the polar one *r*, *φ* (fig. **2.2A.1**). Therefore, we consider the substitution $x = r\ cos(\varphi)$. Field intensity of the primary wave is then

$$E_{z\ prim} = E_0 e^{jkr\cos\varphi}, \qquad \frac{\partial E_{z\ prim}}{\partial z} = 0. \qquad\qquad (\ 2.2A.1\ )$$

This wave excites currents on the surface of the cylinder, which act as sources of the secondary wave. Dealing with induced currents, we do not have any information about them now. We know only that the secondary wave has to meet the homogeneous wave equation . Its general solution is of the form

$$E_{z\ sek} = \sum_{m=0}^{\infty} A_m H_m^{(2)}(kr)\cos(m\varphi). \qquad\qquad (\ 2.2A.2\ )$$

Here, *m* is a separation constant, A_m are integration constants and $H_m^{(2)}(kr)$ is Hankel function of order *m* of second kind and of the argument (*kr*). After the rotation for 360 degrees in the direction *φ*, values of the field intensity have to repeat, and therefore, the constant *m* is integer number and summation (2.2A.2) is performed for all the integer *m*. Derivation of (2.2A.2) is given in the layer B.

Finally, values of the integration constants A_m have to be found. To reach that, the boundary condition has to be applied: tangential component of the *total* field intensity on the surface of perfectly conducting cylinder has to be zero. Since E_z is tangential, we get

$$E_{z\ prim} + E_{z\ sek} = 0 \text{ on condition } r = a. \qquad\qquad (\ 2.2A.3\ )$$

Substituting (2.2A.1) and (2.2A.2) to (2.2A.3) and performing several mathematical re-arrangements, we get relations for computing integration constants A_m:

$$A_0 = -E_0 \frac{J_0(ka)}{H_0^{(2)}(ka)}, \; A_m = -2j^m E_0 \frac{J_m(ka)}{H_m^{(2)}(ka)}. \qquad\qquad (\ 2.2A.4\ )$$
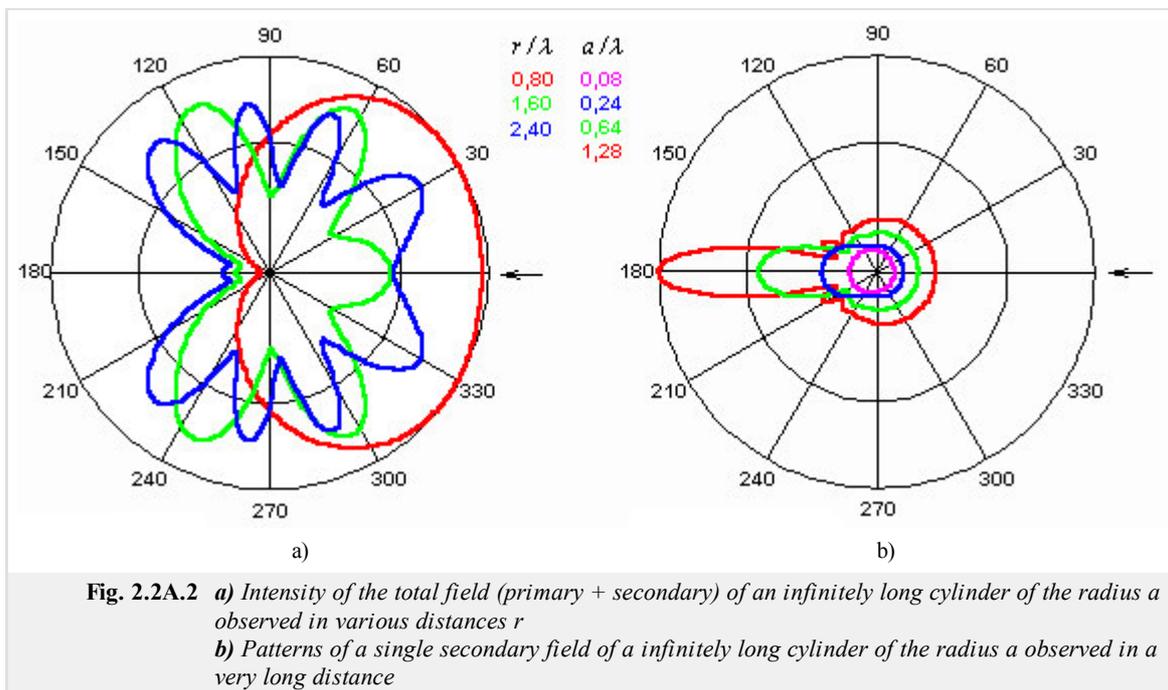
In the above-given relation, E_0 denotes the primary wave amplitude, *k* is wave-number of the environment containing the cylinder, *a* is radius of the cylinder, $J_m(ka)$ is Bessel function of order *m* and of the argument *ka*, $H_m^{(2)}(kr)$ is Hankel function of order *m* of second kind and of the argument (*ka*) and *j* denotes imaginary unit.

Derivation is given in the layer B.

The total intensity in the surrounding of the cylinder equals to the summation of the primary wave intensity and the secondary one. The result is obtained in the form of an infinite series - the secondary wave (2.2A.2) summed with the primary one (2.2A.1). That way, the solution is finished. In applications, even the speeding-up the convergence of the final series has to be solved.

Now, turn our attention to the structure of the wave in the surrounding of the cylinder. The structure is complicated. In the radial direction, cylindrical waves given by (2.2A.2) propagate. These waves interfere with the primary wave, and therefore, both the traveling wave, and the

standing one exist in the surrounding of the cylinder. Since each radial direction contains a different angle with the propagation direction of the primary wave, the standing-wave wavelength is different in every direction. As a result, *directivity patterns* in different distances are different as shown in fig. **2.2A.2**. In this figure, the propagation direction of an incident wave is indicated by an arrow. In the left, directivity patterns of the total wave is depicted, in the right, directivity patterns of only secondary wave are shown. Surprisingly, the maximum of the secondary wave is in the direction behind the cylinder.

Fig. 2.2A.2 *a) Intensity of the total field (primary + secondary) of an infinitely long cylinder of the radius a observed in various distances r*
b) Patterns of a single secondary field of a infinitely long cylinder of the radius a observed in a very long distance

The wave structure near the conducting cylinder can be observed using matlab programs. The first program displays *directivity patterns* of both the single secondary wave and the total wave, which are observed in various distances from the cylinder axis. The second program shows the total (standing) wave in various radial directions. Programs are described in the layer C.

Today, solution of the diffraction task is known for various geometrically simple objects. From the practical point of view, the solution for a general ellipsoid is of great importance because the proper choice of the half-axis length enables to approximate technically useful shapes. For $a = b = c$ the ellipsoid becomes a sphere. For $a = b \ll c$ the ellipsoid approximates a cylindrical conductor of a finite length. For $a = b \gg c$ the ellipsoid approaches a circular slab.

Finally, a small note. During the solution of the diffraction on the cylinder, the primary wave was assumed to exist *everywhere in the space, i.e. even behind the cylinder*. I.e. even behind the cylinder on the reverse side, the boundary condition (2.2A.3) has to be fulfilled. This fact enabled to solve many diffraction problems at the beginning of the 20th century, which was very important for the development of radio electronics.

2.2 General Theory of diffraction

Advanced theory

Now, we are going to become familiar with the theory, which enables us to solve diffraction tasks on objects of more general shape in an analytical way. As demonstrated, generality of the solution is related to the physical approach only. From the formal (mathematical) point of view, the described approach can be practically applied to computing the diffraction on objects of a geometrically simple shape.

As described in the layer A, the physical principle is as follows. In the surrounding of the object, a known primary wave, which illuminates the object, exists. As an effect of illumination (i.e., due to the induced currents), the object becomes the source of *own*, secondary wave, which propagates from the object in all the directions. The total intensity in the surrounding of the object is given by the summation of field intensities of the primary wave and the secondary one. Since the intensity of the primary wave is known, we have to compute the secondary-wave intensity only. The following two conditions can help us for that:

Fig. 2.2B.1 *Diffraction on cylinder*

1. the secondary wave has to be the solution of the wave equation;

2. summation of the field intensity of the primary wave and the secondary one has to meet the boundary condition on the object surface.
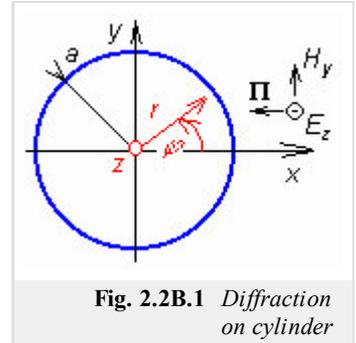
Up to now, everything seems being clear. Unfortunately, factual computations bring difficulties. Therefore, we describe the solution of the plane wave diffraction on an infinitely long perfectly conducting circular cylinder without significant mathematical skips here.

The cross section of the cylinder is depicted in fig. **2.2B.1**. The cylinder axis is identical with the *z* axis of the Cartesian coordinate system, and the radius of the cylinder is *a*. The cylinder is illuminated by the plane wave propagating in the direction -*x* and the electric-intensity vector **E** has got a single component E_z i.e. is parallel with respect to the cylinder surface). Since in all cuts lying in parallel with the plane *xy* the situation is the same, the task can be solved in two dimensions. Due to the axial symmetry, the task is solved in the cylindrical coordinate system. In the planar cut, the solution is found in polar coordinates *r*, *φ* (see fig. **2.2B.1**). Obviously, $x = r \cos(\varphi)$.

Considering the above description, filed intensity of the primary wave is

$$E_{z\,prim} = E_0 e^{jkx} = E_0 e^{jkr\cos\varphi}, \qquad \frac{\partial E_{z\,prim}}{\partial z} = 0. \tag{2.2B.1}$$

Field intensity of the secondary wave has to meet the wave equation

$$\nabla^2 E_{z\,sek} + k^2 E_{z\,sek} = 0. \tag{2.2B.2}$$

Rewriting operator ∇^2 for the cylindrical coordinate system **[1]** and substituting $d/dz = 0$, we obtain

$$\frac{\partial^2 E_{z\,sek}}{\partial r^2} + \frac{1}{r}\frac{\partial E_{z\,sek}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_{z\,sek}}{\partial \varphi^2} + k^2 E_{z\,sek} = 0. \tag{2.2B.3}$$

The above equation is solved by the method of variables separation. We assume $E_z = R(r)\,\Phi(\varphi)$ and substitute to (2.2B.3). Deriving by *φ* is *R* constant and vice versa. Then, the equation is divided by *RΦ* and the term containing *φ* and *Φ* is simply separated. The rest of terms contain *r* and *R* only then. The equation is separated.

If the equation should be valid for any combination of variables *r* and *φ*, each of both separated parts of the equation has to be constant. The term containing *φ* and *Φ* only is put being equal to $-m^2$, where *m* is so called separation constant. That way, the following equation is obtained:

$$\frac{d^2\Phi}{d\varphi^2} + m^2\Phi = 0. \tag{2.2B.4}$$

Eqn. (2.2B.4) is the second-order differential equation with constant coefficients (*m*). Solving it via characteristic equation, the result can be expressed by exponential functions or goniometric ones. Since the result, i.e. the function $\Phi(\varphi)$, describes the field-intensity dependency on the coordinate *φ*, i.e. around the cylinder axis, and since no traveling wave is expected in this direction, we choose goniometric functions:

$$\Phi = A\cos(m\varphi) + B\sin(m\varphi) = C\cos(m\varphi + \varphi_0). \tag{2.2B.5}$$

A and *B*, or *C* and φ_0 denote integration constants. Here, we prefer the last form of the solution, where the constant φ_0 depends only on the direction from which the angle *φ* is measured.

Thanks to the axial symmetry, the direction $\varphi = 0$ can be chosen arbitrarily, and therefore, we put $\varphi_0 = 0$. The final relation is formally simplified that way. Considering eqn. (2.2B.5), another important conclusion can be done. Changing the angle *φ* for 2π, we reach the same point of the space, i.e. we come back to the same field intensity, and consequently to the same functional value of *Φ*. This is conditioned by the

requirement that the separation constant m is integer number.

Now, turn our attention to the second part of the eqn. (2.2B.3), which contains variables r and R. This second part is of the form

$$\frac{r^2}{R}\frac{d^2R}{dR^2} + \frac{r}{R}\frac{dR}{dr} - m^2 + k^2 r^2 = 0. \qquad (2.2B.6)$$

Substituting $\rho = kr$ and performing simple rearrangements, we get Bessel equation:

$$\frac{d^2R}{d\rho^2} + \frac{1}{\rho}\frac{dR}{d\rho} + \left(1 - \frac{m^2}{\rho^2}\right)R = 0. \qquad (2.2B.7)$$

Since the constant m is integer number, the solution of Bessel equation is a linear combination of Bessel and Neumann functions or linear combination of Hankel functions of the first kind or of the second kind. Since the function R describes field-intensity dependency on the radial coordinate r (the direction outwards the cylinder), and in that direction *the wave propagates*, we choose Hankel functions. Since Hankel functions of the first kind describe the wave from the infinity *to its source* and Hankel function of the second kind describes the wave propagating outwards the source, we choose the second-kind function only. Combining the solution of (2.2B.4) and (2.2B.7), we yield the general integral of the equation (2.2B.3):

$$E_{zsek} = AH_m^{(2)}(kr)\cos(m\varphi). \qquad (2.2B.8)$$

Eqn. (2.2B.3) is fulfilled by the solution of (2.2B.8) for each integer m. Since the right value of m unknown this time, we have to admit all the possible values of the separation constant m and even the arbitrary linear combination of those solutions. The final result of solving (2.2B.3) can be expressed in the form of an infinite series

$$E_{zsek} = \sum_{m=0}^{\infty} A_m H_m^{(2)}(kr)\cos(m\varphi). \qquad (2.2B.9)$$

The result (2.2B.9) can be understood as a set of the infinite number of waves propagating outwards the cylinder, which mutually interfere into the secondary wave. Each summand of the series (2.2B.9) represents one of these partial waves. These waves differ in their amplitudes (unknown coefficient A_m), in their dependency on the radial distance r (the coefficient $H_m^{(2)}(kr)$ depends on m) and also in their dependency on φ, i.e. on the direction outwards the cylinder. The wave with $m = 0$ propagates in all the directions with the same amplitude $\cos(m\varphi) = \cos(0) = 1 = \text{const.}$, the wave with $m = 1$ exhibits *eight-like* directivity pattern (maximum for $j = 0$, π; zero amplitude for $\varphi = \pm\pi/2$), etc.

At this moment, the solution seems to become complicated. But, thanks to the infinite series (2.2B.9), the solution can be successfully finished. Properly choosing coefficients A_m we can *enforce* the series (2.2B.9) to meet in its whole those boundary conditions, which have to be met on the surface of the object (the situation is conformable to Fourier series: a proper choice of Fourier coefficients enables us to express an *arbitrary* time course of the signal).

If perfectly conducting and infinitely long cylinder is assumed, than the boundary condition is formulated in a simple way: the tangential component of the total-field intensity has to be zero on the surface. Since our wave consists of the component E_z only, which is parallel to the surface, we get

$$E_{z\,prim} + E_{z\,sek} = 0 \text{ on condition } r = a. \qquad (2.2B.10)$$

Substituting from (2.2B.1) and (2.2B.9), we get

$$E_0\exp(jka\cos\varphi) + \sum_{m=0}^{\infty} A_m H_m^{(2)}(ka)\cos(m\varphi) = 0. \qquad (2.2B.11)$$

From the above equation, unknown coefficients A_m have to be computed. Unfortunately, a single equation of an infinite number of coefficients is at our disposal. In similar situations, the method of inexplicit coefficients ([1], [4]) can help us. This method requires all the addends (both in our case) in eqn. (2.2B.11) being expressed by a series of the same type (in our case, by a series containing terms $\cos(m\varphi)$. If this requirement is satisfied then $m = 0, 1, 2, \ldots \infty$ are sequentially substituted to the equation, and that way, the single equation comes to the set of the infinite number of equations for coefficients A_m.

Fortunately, we know that

$$\exp(jka\cos\varphi) = J_0(ka) + \sum_{m=1}^{\infty} 2j^m J_m(ka)\cos(m\varphi), \qquad (2.2B.12)$$

which enables us to expand the primary wave to the series consisting of term of the type $\cos(m\varphi)$. The series is substituted to (2.2B.11), and coefficients at $\cos(\varphi)$, $\cos(2\varphi)$, $\cos(3\varphi)$,... are sequentially compared. That way, following relations are obtained:

$$E_0 J_0(ka) = -A_0 H_0^{(2)}(ka), \quad 2j^m E_0 J_m(ka) = -A_m H_m^{(2)}(ka) \qquad (2.2B.13)$$

and consequently

$$A_0 = -E_0 \frac{J_0(ka)}{H_0^{(2)}(ka)}, \quad A_m = -2j^m E_0 \frac{J_m(ka)}{H_m^{(2)}(ka)}.$$

(2.2B.14)

That way, the analytical solution of our task is finished. First, coefficients A_m are evaluated, and then, the secondary wave is given by the series (2.2B.9). The total-field intensity in the cylinder surrounding is obtained by adding field intensities of the secondary wave (2.2B.9) and the primary one (2.2B.1).

Finally, let us remind conditions of the successful operation of the method. First, the object shape has to be simple; its surface has to be identical with a plane in such coordinate system, where the wave equation (2.2B.2) can be solved. Second, the primary wave has to be expanded into the series of the same type, which is given by the solution of the wave equation. Finally, the secondary-wave series has to exhibit the satisfactory convergence. The convergence is rather problematic in situations, when object dimensions are several-orders larger than the wavelength. E.g., if computing the diffraction of very short waves on the surface of the globe, then the satisfactorily accurate expression of the secondary wave requires addition of tens and hundreds millions of addends in some regions.

Today, the solution of the diffraction task is known for various simple objects: sphere, cylinder (spherical elliptical, parabolic) and for general ellipsoid; the proper choice of the half-axis length enables to approximate technically useful shapes. For $a = b = c$ the ellipsoid becomes a sphere. For $a = b << c$ the ellipsoid approximates a cylindrical conductor of a finite length. For $a = b >> c$ the ellipsoid approaches a circular slab.

The structure of the wave was briefly commented in the layer A. In the surrounding of the cylinder, both the traveling wave and the standing one exist. The standing wave is created by the interference of the primary wave and the secondary one. Since both the waves are of the different propagation direction and since event the phase velocities of both the waves are different, the wavelengths of the standing wave are in different directions different too. Surprisingly, the secondary wave exhibits the highest intensity in the direction behind the cylinder. If the radius of the cylinder is smaller than approx. 1/10 of the wavelength, the the influence of the cylinder to the primary wave is negligible. In detail, the structure of waves can be studied using computer programs (see the layer C). The first program displays field intensity of the secondary wave and of the total one in polar coordinates in the form of directivity patterns. Field intensities in various directions in a constant distance from the cylinder axis can be observed. The second program shows the distribution of field intensities along radial lines in different directions. A more detailed description is given in the layer C.

Finally, a brief note on the solution of the diffraction task in the situation, when the conductivity of the cylinder is not infinite or if a dielectric cylinder is analyzed. In that case, the wave equation (2.2B.2) has to be solved even inside the cylinder. There, waves do not propagate and the standing wave appears there. Therefore, we do not use Hankel functions the solution of he wave equation, but Bessel and Neumann functions are exploited. Since for $r = 0$ (on the cylinder axis) the value of Neumann function approaches infinity whereas the field intensity has to be finite, we have to eliminate this function. In analogy to (2.2B.9), we can write for the region inside the cylinder:

$$E_{sekuvnit\check{r}te\check{c}} = \sum_{m=0}^{\infty} B_m J_m(kr)\cos(m\varphi).$$

(2.2B.15)

The coefficients can be determined using the boundary condition

$$E_{sekuvnit\check{r}te\check{c}} = E_{sekvn\check{e}te\check{c}} + E_{primte\check{c}}.$$

(2.2B.16)

Even here all the intensities have to be expressed in the form of series of the same type, and the method of inexplicit coefficients [4] has to be applied. Since two infinite sets of unknown coefficients (A_m, B_m) appear here, the only condition (2.2B.16) is insufficient. The solution has to be repeated even for magnetic field including the proper boundary condition.

2.2 General Theory of diffraction

Matlab program

The program serves for computing and displaying the secondary (diffraction) field and the total field in the surrounding of the circular conducting cylinder, which is illuminated by a plane wave.

First, a path of matlab has to be directed to the folder **Gen_diffraction**. The program runs using the m-file `valec`. After the introductory window disappears, the program asks for the input of data for computation. We have to give the relative cylinder radius a/λ and the distance of the observation point d/λ. Then, the computation is started (the computation takes longer time, and therefore, an indicator of the computation progress is used). When the computation is over, distribution of the primary field and the secondary one in the surrounding of the cylinder is displayed in the plane, which is perpendicular to the cylinder axis.

Then, buttons for finishing the program | Finish | and for starting a new run of the program with new input data | Again | appear. In five directions, dependency of the secondary wave and the total one on the distance from the cylinder axis are shown. Field intensity $E=f(r)$ is displayed for a direction, which is opposite to the direction of the incident wave propagation, and then by steps 45° until the direction of the propagation of the incident wave is met.

2.2 General Theory of diffraction

Developing Matlab

The program is based on the general theory of diffraction. We suppose that the object (cylinder) is illuminated by a wave of known intensity (primary wave). Thanks to this wave, the cylinder becomes the source of other waves (secondary wave). The total wave is created by the interference of the primary wave and the secondary one.

User has to introduce two parameters to the input of the program: `a` (relative radius of the cylinder a/λ) and `r` (relative distance from the cylinder axis r/λ). Field intensity of the primary wave is then programmed by the following way:

```
Ezprim(theta) = cos(2*pi*r*cos((theta-1)*pi/180))+...
+j*sin(2*pi*r*cos((theta-1 )*pi/180));
```

Here `r` is a constant, `a` equals to a/λ, `theta` varies from 0° to 180°. `theta` is used even as an index of an array, its values have to be from the interval from `1` to `181`. The result is symmetrical with respect to the horizontal axis, and therefore, there is no need of computing to 360°.

Next, the secondary wave intensity is computed according to

$$E_{zsek} = \sum_{m=0}^{\infty} A_m H_m^{(2)}(kr)\cos(m\varphi).$$

(2.2D.1)

In our situation, we limit the value of `m` by 150.

First, the term for `m` = 0 is computed:

```
Ezsekk = A0*besselh(0,2,2*pi*r)*cos(0);
```

where

```
A0 = -besselj(0,2*pi*a)/besselh(0,2,2*pi*a)
```

A cycle for computing `Ezsekk` is of the form:

```
for m=1:60+x
    Am(theta) = -2*j^m*besselj(m,2*pi*a)/besselh(m,2,2*pi*a);
    Ezsek(theta) = Ezsekk+Am(theta)*besselh(m,2,2*pi*r)*...*cos(m*(theta-
    1)*pi/180);
end
```

Summing `Ezprim` and `Ezsek`, the intensity `Ezvyst` is obtained, which is displayed together with `Ezsek` in charts `Ezsek=f(theta)` and `Ezvysl=f(theta)`.

The second part of the program computes the same dependencies, but *cuts* of the distribution of the secondary wave and the total one on the distance from the cylinder axis in given directions `theta` are displayed. For given direction, `theta` is a constant and r/λ is changed. The computation algorithm stays the same and only variables are changed. As a result, five charts for `theta` = (0, 45, 90, 135, 180)° containing `Ezsek` and `Ezvyslare` displayed.

Note: Programs are called in the following order:

`Valec.m – Inform.m – Zadanivalec.m – Vypocetvalec.m – Vypocetrezy.m`

2.2 General Theory of diffraction

Java applet

Theta = 0°	Theta = 45°	Theta = 90°	Theta = 135°	Theta = 180°

Input / Vstup dat	Ez result / Ez výsledná	Ez secondary / Ez sekundární

Relative diameter of cylinder a / lambda [-]:

Relativní poloměr válce a / lambda [-]:

a / lambda [-]:

1

Relative distance from axe of cylinder r / lambda [-]:

Relativní vzdálenost od osy válce r / lambda [-]:

r / lambda [-]:

4

Calculate / Vypočti

Help / Nápověda

2.2 General Theory of diffraction

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

In case a dielectric cylinder is illuminated by a primary electromagnetic wave …

Possible answers for question 1:

- ⊙ … conductive currents are induced on the surface of the cylinder, and a secondary wave is radiated.
- ⊙ … the dielectrics is polarized and a secondary wave is radiated.
- ⊙ … the cylinder absorbs the primary wave.
- ⦿ I won't answer this question

Question 2

The secondary wave that is radiated by currents induced on a perfectly electrically conductive cylinder is described by …

Possible answers for question 2:

- ⊙ … a homogeneous wave equation in Cartesian coordinates.
- ⊙ … the first Maxwell equation.
- ⊙ … a homogeneous wave equation expressed in terms of Hankel functions.
- ⦿ I won't answer this question

Question 3

In the surrounding of the cylinder …

Possible answers for question 3:

- ⊙ … both the traveling waves and the standing ones appear.
- ⊙ … only traveling waves appear.
- ⊙ … only standing waves appear.
- ⦿ I won't answer this question

Question 4

In the analysis, boundary conditions …

Possible answers for question 4:

- ⊙ … are considered in order to determine integration constants in the general solution.
- ⊙ … are not considered.
- ⊙ … are considered in order to determine the amplitude of the primary wave.
- ⦿ I won't answer this question

Question 5

When evaluating integration constants in the analytical solution …

Possible answers for question 5:

- ○ … the method of variables separation is applied.
- ○ … the method of inexplicit coefficients is used.
- ○ … the method per partes is applied.
- ● I won't answer this question

see result

2.3 Geometrical optics

Basic theory

Fresnel theory of diffraction is simple, but using it, we can analyze thin planar obstacles only. General theory, which was described in chapter 2.2, is formally rather complicated, and only geometrically simple objects can be handled with. Therefore, alternative ways of the analysis were sought out. *Geometric theory of diffraction* (GTD) belongs to those ways: GTD numerically computes even rather complicated situations. Before explaining the matter of GTD, the basic terms of *geometrical optics* (GO) are introduced to the reader.

Today's geometrical optics is an efficient tool for solving wave phenomena (wave propagation) in complex media. GO is not limited to the range of optical frequencies, and it can be used even for radio waves. From the classical geometrical optics, the idea of wave propagation along beams was adopted. Moreover, GO is able to compute not only wave trajectories but too changes of field intensities and polarization of waves during propagation. The theory of GO is based on the following two assumptions:

1. Wavelength is small, and therefore, wave number k is high.

2. The wave is observed far away from the source. Whereas the wave amplitude changes slowly in the propagation direction, phase varies quickly. The sense of this requirement can be perceived using the following illustration example.
   We are interested in the propagation of the spherical wave in the distance of 10 wavelengths from the source. If the distance is increased for one half of the wavelength, i.e. for 5 %, the intensity amplitude decreases for 5 % too, but the phase changes for π radians (a significant change).

We start the explanation of geometrical optics by modifying the relation for the intensity of electromagnetic field. Instead of $E = E_m \exp(-jkr)$, we write

$$E = E_m \exp\left[-jk_0 L(x, y, z)\right]. \tag{2.3A.1}$$

In the exponent, we have in all the situations $k_0 = \omega \, (\varepsilon_0 \, \mu_0)^{1/2}$ and the parameters of the medium are included in the function L. We simply understand that $L(x, y, z) = const$ is equation of equiphase surface (wave surface) and that the vector grad L is of the direction, which is perpendicular to equiphase surface, i.e. of the propagation direction.

The relation (2.3A.1) is substituted to Maxwell equations. Assuming that the wave number k is high, relatively complicated rearrangements yield

$$\left|grad\, L\right|^2 = n^2, \tag{2.3A.2}$$

where

$$n = k / k_0 = \sqrt{\varepsilon_{rel}\mu_{rel}} \tag{2.3A.3}$$

denotes the refractive index of the medium.

Eqn. (2.3A.2) is called *the basic equation of geometrical optics*. The function $L(x,y,z)$ is called the eiconale. It is the scalar function of coordinates. The vector grad L is of the direction of spherical wave propagation in every point. The curve, which tangent is of the direction of grad L is every point, is called the beam. The beam is of the direction of the steepest change of phase in every point, and it is of the direction of Poynting vector too (i.e. of the direction of the energy flow). In an inhomogeneous medium, beams can be curved and eqn. (2.3A.2) is the differential equation of beams.

For practical computations of beams, the form of (2.3A.2) is not suitable. Therefore, the following relations are used for computing beam trajectories:

$$\frac{\partial}{\partial s}\left(n\frac{\partial x}{\partial s}\right) = \frac{\partial n}{\partial x}, \quad \frac{\partial}{\partial s}\left(n\frac{\partial y}{\partial s}\right) = \frac{\partial n}{\partial y}, \quad \frac{\partial}{\partial s}\left(n\frac{\partial z}{\partial s}\right) = \frac{\partial n}{\partial z}, \tag{2.3A.4}$$

$$\left(\frac{\partial x}{\partial s}\right)^2 + \left(\frac{\partial y}{\partial s}\right)^2 + \left(\frac{\partial z}{\partial s}\right)^2 = 1. \tag{2.3A.5}$$

The variable s is curvilinear coordinate *along* the beam. Details are given in the layer B including the derivation and an illustrative example.

Geometrical optics enables to compute not only beam trajectories but too the variations of amplitude and phase of field intensity along the beam:

In the starting point (A e.g.) a (infinitely) facet dS_1 is chosen of the wave surface and a beam is led through every point of the edge of this facet. That way, a *beam tube* is obtained. On some of the following equiphase surfaces (B), the beam tube is of the different cross section dS_2 (fig. **2.3A.1**). Since the energy propagates *along* the beams, it cannot leave the tube through the side walls. In the lossless medium, the power passing facets dS_1 and dS_2 is identical. Since $P = \Pi\, S = (E^2/Z_0)\, S$ and $Z_0 = (\mu/\varepsilon)^{1/2}$, we can simply derive the relation between intensities on both the facets:

$$\sqrt{\frac{\varepsilon_1}{\mu_1}}|E_1|^2 dS_1 = \sqrt{\frac{\varepsilon_2}{\mu_2}}|E_2|^2 dS_2. \qquad (2.3\text{A.}6)$$

Phase of field intensity in B can be computed using eiconale, resp. using eqns. (2.3A.1) or (2.3A.2). If the eiconale is of the value L_A at the beginning of the trajectory A, then in B (which has to be located at the same beam)

$$L_B = L_A + \int_A^B n \, ds. \qquad (2.3\text{A.}7)$$

Integration is done along the beam.

Eqn. (2.3A.6) is not valid in regions, where the beams cut (infinitely high field intensity would be obtained). Such situation can be met in the focus and on the surface called *caustics* (see layer B).

In more complicated cases, beams in different transversal planes are of different curvature radii of their wave surfaces. In such situations, (2.3A.6) is not valid. Nevertheless, the intensity can be computed (see layer B).
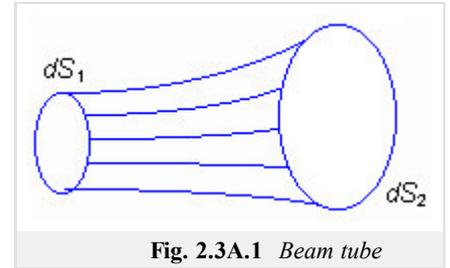
Fig. 2.3A.1 *Beam tube*

2.3 Geometrical optics

Advanced theory

Today&s geometrical optics (GO) enables to examine the propagation of electromagnetic waves even in complex inhomogeneous media of a given type. As the classical geometrical optics, GO is based on the notion that waves propagate along beams, which can be even curvilinear now. GO is based on the exact mathematical theory coming out Maxwell equations and enabling to compute not only wave trajectories but too variation of magnitude and phase of field intensity and wave polarization. Moreover, the method is well-understandable thanks to the simple notion of beams.

As already described in the layer A, theory of geometrical optics is based on the following two assumptions:

1. Wave number k is high (so that some terms containing the higher power of k in the denominator can be neglected during mathematical rearrangements). This assumption predetermines geometrical optics to the applications in the area of optical frequencies. Since the wave number is not needed being extremely high, the method can be used even for radio frequencies.

2. The wave is examined in a relatively long distance from the source. In such distance, wave amplitude varies slowly whereas the wave phase changes quickly (for 2π for every wavelength).

In the following, our attention is focused in the media exhibiting a continuous variation permittivity and permeability. The media are assumed to be lossless. Refractive index of the medium

$$n = \sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}} = \sqrt{\varepsilon_r\mu_r} \qquad (2.3B.1)$$

is a real number and is a function of the position (x, y, z). Symbol s denotes the coordinate (curvilinear in general), which is measured along the direction of propagation.

Investigating the wave propagation in a homogeneous medium, wave number $k = k_0(\varepsilon_r\mu_r)^{1/2} = k_0 n$ and the propagation direction are usually known. E.g., the following relation is valid for the field intensity of a plane wave

$$E = E_0 \exp(-jks) \ (s \text{ is in the propagation direction}) \qquad (2.3B.2a)$$

On the part of the trajectory ds, the phase changes for

$$d\varphi = kds = k_0 nds. \qquad (2.3B.2b)$$

In the coordinate system x, y, z, the propagation direction need not to be identical with the direction of a coordinate axis. We usually solve this situation by considering the wave number k as a vector oriented to the propagation direction. Its components k_x, k_y, k_z, to the directions of coordinate axes are wave numbers in the directions x, y, z. In two arbitrarily located points A, B, which coordinates differ for dx, dy, dz, the phases differ for

$$d\varphi = k_x dx + k_y dy + k_z dz. \qquad (2.3B.3)$$

Geometrical optics does not suppose any a priori knowledge of the direction of wave propagation, because it is different in different points (in the medium of a variable refractive index). In parallel with (2.3B.2), we use

$$E = E_m \exp[-jk_0 L(x, \ y, \ z)] \qquad (2.3B.4a)$$

$$d\varphi = k_0 dL, \qquad (2.3B.4b)$$

where the distance is not explicitly present. The quantity L is a scalar function of coordinates x, y, z, it is called *eiconale* and it determines the phase of field intensity in every point. Comparing (2.3A.4b) and (2.3A.2b) shows that on the infinitely short path ds (where $n = const$), is

$$dL = nds. \qquad (2.3B.5)$$

Considering (2.3A.4), equation $L = const$ is the equation of a plane, where wave is of the same phase, i.e. the equation of a wave surface. Vector grad L is of the direction of the steepest variation of the eiconale (i.e., even of the phase), and this is the direction of wave propagation. For example:

$$\frac{dL}{ds} = |grad \, L|$$

and according to (2.3A.5) is

$$\frac{dL}{ds} = n,$$

and therefore

$$|grad\,L| = n \qquad (2.3B.6)$$

or

$$k_0\,\frac{dL}{dx} = k_0\,\frac{dL}{ds}\,\frac{ds}{dx} = k_0\,\frac{dL}{ds}\cos(x,\,s) = k_0 n \cos(x,\,s) = k_x. \qquad (2.3B.7)$$

As already said in the layer A, the equation

$$|grad\,L| = \sqrt{\left(\frac{\partial L}{\partial x}\right)^2 + \left(\frac{\partial L}{\partial y}\right)^2 + \left(\frac{\partial L}{\partial z}\right)^2} = n \qquad (2.3B.8)$$

is considered as the basic equation of the geometrical optics. The vector grad L is of the direction of wave propagation in every point, and the curve, which tangent is of the direction grad L in every point, is called beam. Beams can be curvilinear and eqn. (2.3B.8) is the differential equation of beams.

For practical calculations, the form (2.3B.8) is inconvenient. Instead, following equations are used:

$$\frac{\partial}{\partial s}\left(n\frac{\partial x}{\partial s}\right) = \frac{\partial n}{\partial x}, \quad \frac{\partial}{\partial s}\left(n\frac{\partial y}{\partial s}\right) = \frac{\partial n}{\partial y}, \quad \frac{\partial}{\partial s}\left(n\frac{\partial z}{\partial s}\right) = \frac{\partial n}{\partial z}, \qquad (2.3B.9)$$

$$\left(\frac{\partial x}{\partial s}\right)^2 + \left(\frac{\partial y}{\partial s}\right)^2 + \left(\frac{\partial z}{\partial s}\right)^2 = 1. \qquad (2.3B.10)$$

As already said, s is the coordinate *along* the beam. Now, the way of deriving the first of relations (2.3B.9) is demonstrated.

In a homogeneous medium, we exploit in computations the wave number, and we understand it as a vector, which is oriented to the propagation direction. Components of k are k_x and k_y (in two dimensions) and we can imagine that wave propagates from any point along the pointed trajectory going in parallel to x (wave number k_x) and then, going in parallel to y (wave number k_y). n an inhomogeneous medium, the same approach can be used within a very small region dx, dy only, where the refractive index can be considered constant. In the frame of the whole medium, k_x and k_y vary, and those variations have to be examined first.
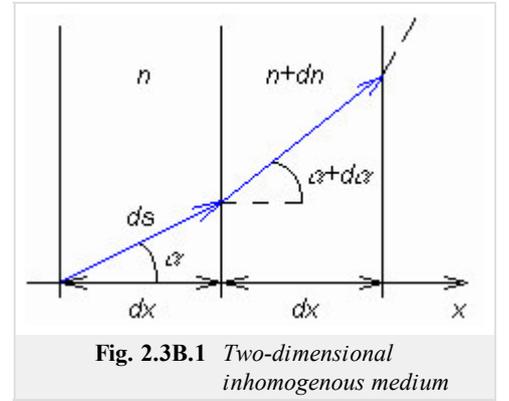
In an inhomogeneous medium, computing with refractive index is preferred to computing with wave number. Since $k = k_0\,n$ and wave number is a vector, even the right-hand side of the equation has to be a vector. Hence, the refractive index n assumed being vector. Its magnitude is n and its direction is identical with the wave propagation (in general, different in different points).

Fig. 2.3B.1 *Two-dimensional inhomogenous medium*

Components n_x and n_y correspond to respective wave numbers, e.g. $k_x = k_0\,n_x$. This formalism converted the inhomogeneous medium to the anisotropic medium, which is not truth from the microscopic point of view. From the macroscopic point of view, this might be considered as a truth: if wave enters the medium in various directions, the trajectory is incurvated different way every time. In the homogeneous medium, the situation is different.

Assume a two-dimensional medium, where refractive index varies in the direction x only. In fig. **2.3B.1**, two elementary layers of this medium are depicted, including respective quantities. Let us examine variations of components of refractive index $n_x = n\cos(\alpha)$ and $n_y = n\sin(\alpha)$ in the direction x:

$$dn_x = d(n\cos\alpha) = (n + dn)\cos(\alpha + d\alpha) - n\cos\alpha,$$
$$dn_y = d(n\sin\alpha) = (n + dn)\sin(\alpha + d\alpha) - n\sin\alpha$$

Rewriting goniometric functions, multiplying, and substituting *sin (dα) = dα, cos(dα) = 1* yield

$$dn_x = dn\cos(\alpha) - n\sin(\alpha)d\alpha, \qquad (2.3B.11a)$$

$$dn_y = dn\sin(\alpha) - n\cos(\alpha)d\alpha. \qquad (2.3B.11b)$$

Next, the variation of dn_y in the direction x, i.e. dn_y/dx, has to *equal zero*. This is caused by the fact that two waves propagating *along* the boundary at different sides near each other have to be of the same wave number, although they propagate in different media. In the opposite case, a continuously rising phase shift between these two waves would exists, and on the boundary, a phase jump would appear, which is impossible.

In (2.3B.11b), we put $dn_y = 0$, express $d\alpha$ and substitute to (2.3B.11a). We get:

$$dn_x = \frac{dn}{\cos\alpha}. \qquad (2.3B.12)$$

The above relation is rather interesting. It shows that the variation dn_x is *higher* than the variation of the refractive index dn at the same point.

The next steps are relatively simple. Eqn. (2.3B.12) is divided by dx and rearranged: $dn_x/dx = (dn_x/ds)(ds/dx) = (dn_x/ds)(1/\cos(\alpha))$. The term $\cos(\alpha)$ vanishes and we get the relation (2.3B.9a).

Exploitation of (2.3B.9) and (2.3B.10) is demonstrated by the following example.

Assume a two-dimensional medium (x, y), where the permittivity ε, and consequently the refractive index n, continuously vary along the coordinate y in accordance with the function $n(y)$; along the coordinate x, the refractive index is constant. Since $\partial n/\partial x = 0$, eqn. (2.3B.9a) results to:

$$\frac{\partial x}{\partial s} = \frac{C}{n}. \tag{2.3B.13}$$

Here, C is an integration constant. Substituting to (2.3B.10), the term containing z is deleted because the task is two-dimensional, and we get

$$\frac{\partial y}{\partial s} = \sqrt{1 - \left(\frac{C}{n}\right)^2}. \tag{2.3B.14}$$

Dividing both equations, the variable s is eliminated and the differential equation of the beam is obtained:

$$\frac{\partial x}{\partial y} = \frac{C}{n}\frac{1}{\sqrt{1 - \left(\frac{C}{n}\right)^2}}. \tag{2.3B.15}$$

We substitute a given function $n(y)$ and integrated (analytically or numerically). The equation of the beam $y = f(x)$ is the result. The integration constant C is determined using a boundary condition, a known beam direction in a given point. If the value of the *eiconale* is known in this point L_A (A e.g.), we can compute its value in another different point (B e.g.) on the same beam. Using (2.3B.5), we get

$$L_B = L_A + \int_A^B n\,ds, \tag{2.3B.16}$$

if the integration path follows the beam. In our task, the differential ds is replaced by the differential dy according to (2.3B.14), n is replaced by $n(y)$ and integration is performed according to y from y_A to y_B . [End of example]

In layer A, we have found that geometrical optics enables to compute intensity changes during the propagation. This ability is conditioned by the knowledge of computing the shape of beams, i.e. the trajectory of wave. Computation of the intensity is based on the fact that wave energy propagates *along* beams. This means, that no electromagnetic energy leaves or enters the walls of a tube, which appears when displaying many neighboring beams and which transversal cut gives a closed curve. Energy propagates inside the tube along beams only. On the basis of this idea, the following relation was derived in the layer A:

$$\sqrt{\frac{\varepsilon_1}{\mu_1}}|E_1|^2 dS_1 = \sqrt{\frac{\varepsilon_2}{\mu_2}}|E_2|^2 dS_2. \tag{2.3B.17}$$

Eqn. (2.3B.17) is not valid for regions, where beams cut (which would cause an infinite field intensity). Such situation appears in a focus and on a caustics. The term caustics is explained thereinafter.

Sometimes, beams can exhibit different curvature radii of wave surface in different planes (e.g., *astigmatic beams*). A part of such a tube is depicted in fig. **2.3B.2**.

The curvature center of the equiphase surface in the horizontal plane is denoted as O_h, and in the vertical one as O_v. The distances of facet dS_1 and O_h (O_v) is r_1 (r_2), r is the distance between facets dS_1 and dS_2.

Considering fig. **2.3B.2**, we get:

Fig. 2.3B.2 *Astigmatic beam*

$$dS_1 = r_1\,d\alpha\,r_2\,d\beta, \quad dS_2 = (r_1 + r)\,d\alpha\,(r_2 + r)\,d\beta.$$

Dividing relations, we express the radius of both the facets, and using (2.3B.17), we compute the ratio of intensities. For a homogeneous medium, when respecting phase variation, we get:

$$E_2 = E_1 \sqrt{\frac{r_1 r_2}{(r_1 + r)(r_2 + r)}} \exp(-jkr). \tag{2.3B.18}$$

The described method of the geometric optics can be used to the examination of the electromagnetic wave propagation through the medium which structure corresponds to the idealized ionospheric layer (if the scaling factor is not considered). Wave radiated by a point source propagates through the medium with the unitary dielectric constant first, and then continues through the layer which dielectric constant decreases down to a certain minimum, and then rises to 1 again. For the given dependency of the dielectric constant in the layer and for the given position of a source, the wave trajectory, the equiphase surface and the eiconale increase can be computed by numeric integration.
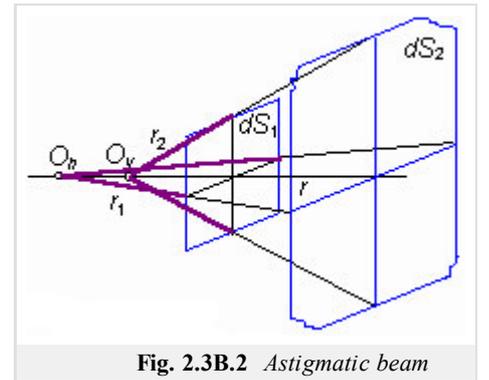
Below the layer, where several waves are propagating (due to the trajectory curvature), the direction of the propagation, field intensity and eiconale (phase) can be computed for each wave. Computations show that beams return back by nice curls in a certain part of space. Tops of the curls can be connected by an imaginary spline, which is not crossed by any beam, and in a close vicinity of which (below) two neighboring beams intersect. This spline (or a surface in space) is called caustics.

2.3 Geometrical optics

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

Geometrical optics is a method for wave propagation modeling that …

Possible answers for question 1:

- ○ … is applicable at optical frequencies only
- ○ … can be used both at optical frequencies and at radio frequencies.
- ○ … works properly at any frequency without limitations
- ◉ I won't answer this question

Question 2

Geometrical optics enables to compute …

Possible answers for question 2:

- ○ … an angle of incidence and an angle of reflection of a beam only.
- ○ … wave propagation in case the phase of the wave does not need to be considered.
- ○ … wave trajectories, intensity variations, and polarization properties.
- ◉ I won't answer this question

Question 3

The basic equation of geometrical optics tells us that …

Possible answers for question 3:

- ○ … the squared refractivity index equals to the squared gradient of the eiconale.
- ○ … the refractivity index equals to the square root of permittivity and permeability product.
- ○ … the angle of incidence equals to the angle of reflection.
- ◉ I won't answer this question

Question 4

The beam tube was introduced in order to …

Possible answers for question 4:

- ○ … compute intensity variations of the propagating wave.
- ○ … define the finite space where the beam energy is concentrated.
- ○ … model polarization properties of propagating waves.
- ◉ I won't answer this question

Question 5

Computations of intensity variations in geometrical optics are conditioned by …

see result

2.4 Geometric theory of diffraction

Basic theory

Geometrical optics is a proper method for computing wave propagation in homogeneous media or in the environment exhibiting a continuous variation of parameters. If the environment contains objects, the method fails.

Look at the fig. **2.4A.1**. We can observe the situation in the surrounding of an illuminated *reflecting* half-plane from the point of view of geometrical optics. The space around the half-plane can be subdivided into three sections: into the section *a*, where both the incident wave and the reflected one are present, into the section *b*, which contains the incident wave only, and to the section *l*, where no wave (from the point of view of geometrical optics) propagates. Moreover, exactly defined planes H_0 (*border of reflection*) and H_S (*border of shade*) separate the sections. From the point of view of geometrical optics, a discontinuity of the field distribution appears on those planes because on different sides of H_0 and H_S, intensity is computed in a different way. In real world, the field distribution is continuous, and in the section of shade, intensity is non-zero. In the sixties of 20th century, *Keller* proposed a specific correction of geometrical optics in order to eliminate field discontinuities in the surrounding of objects. That way, the basics of the *geometric theory of diffraction* were built.
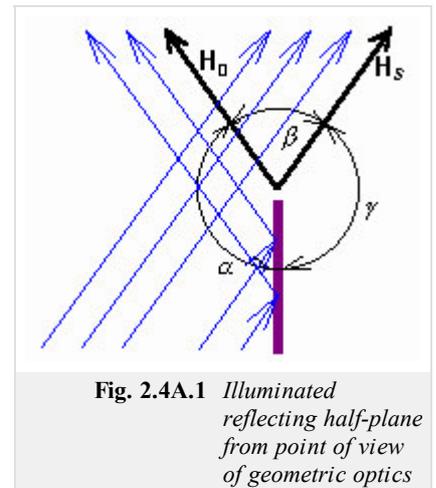
Fig. 2.4A.1 *Illuminated reflecting half-plane from point of view of geometric optics*

Geometric theory of diffraction (GTD) is an extension of geometrical optics to inhomogeneous media. GTD solves interaction of beams and objects according to the laws of geometrical optics, which is completed that way to eliminate obvious incorrectness (intensity discontinuities at the border of the reflection section and the shade one) and to preserve the main advantages of geometrical optics (conception of beams). The incorrectness of geometric optics is eliminated following the below-described postulates:

1. Beams of the incident wave excite new waves on the illuminated object. Those waves propagate away *from the object* (not from the primary source). Those waves are called *diffraction waves*. Diffraction waves can be described (computed) using beams. Diffraction waves (diffraction beams are of such amplitudes and phases to eliminate field discontinuities on boundaries.
   Diffraction waves are excited only by the beams, which
   - "fall to the edges of the object, to the boundaries of different curvature radii of the surface, etc.
   - "touch the surface of the object.

2. Amplitude of diffraction waves is proportional to the amplitude of the incident wave. Proportionality coefficient is called *diffraction factor*.

3. Diffraction constant is given by *local* (!) properties of the object surface.

In fig. **2.4A.2a**, arrows indicate diffraction waves excited by an incident wave illuminating the *edge* of the half-plane. Diffraction waves create the set of cylindrical waves, which common source is the half-plane edge. Adding incident and reflected wave together, continuous intensity transitions on the border of the reflection region and the shade one is ensured.

In fig. **2.4A.2b**, beams of diffraction waves excited by the incident wave beam are depicted for the situation when the beam touches the object surface. In the moment when the beam of the diffraction wave leaves the object surface, it *starts to behave* according to the geometric optics rules.

Finally, the postulate about the local parameters of objects is very important. In the situation of fig. **2.4A.2b**, the diffraction factor is given by parameters *e*, *m* of the object surface and by the curvature radius of the surface in the region, where the beam *slides* (over the surface. Parameters of the rest of the object do not play any role. Thanks to this property, even complicated diffraction situations appearing in a real terrain can be solved (fig. **2.4A.2c**).

Fig. 2.4A.2 *Geometric theory of diffraction. Diffraction waves beams **a)** on the half-plane edge, **b)** on the spherical surface, **c)** on obstacles in the real terrain*

Thanks to the above-mentioned postulate, values of diffraction factors can be computed. E.g., *all* the oblong objects, which have the same curvature radius of a certain part of their surface, have the same diffraction factor. The factor has to be the same as for the cylinder of the same radius. And the diffraction on a cylinder can be solved exactly. We only have to rearrange the results of the exact theory (see the chapter 2.2) to the form applicable to the theory of geometric optics. This is solved by so called *canonical problems*.

GTD is together with geometrical optics an efficient tool for solving diffraction problems. Preserving the conception of beams and computing with local object properties are the main advantages of the method. The method has to be implemented using computer.

2.4 Geometric theory of diffraction

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

GTD …

┌─Possible answers for question 1:──
  ◎ … explains the diffraction phenomena using a simple Euclidian geometry.

  ◎ … is an original interpretation of Huygens' principle.

  ◎ … enables to apply the principles of geometrical optics to wave propagation modeling in an inhomogeneous environment.

  ◉ I won't answer this question
└───

Question 2

In GTD, diffraction waves are assumed to be …

┌─Possible answers for question 2:──
  ◎ … radiated by the currents on the obstacle to eliminate field discontinuities at boundaries.

  ◎ … negligible – otherwise the wave theory of diffraction has to be applied.

  ◎ … dominant – other waves can be therefore neglected.

  ◉ I won't answer this question
└───

Question 3

Diffraction factor is dominantly determined …

┌─Possible answers for question 3:──
  ◎ … by the contrast between the permittivity of the obstacle and surrounding.

  ◎ … by the local properties of the obstacle.

  ◎ … by the angle of incidence.

  ◉ I won't answer this question
└───

Question 4

In the practical life, GTD …

┌─Possible answers for question 4:──
  ◎ … enables to model the wave diffraction on hills, e.g..

  ◎ … is not applicable.

  ◎ … can be used at very low frequencies only.

  ◉ I won't answer this question
└───

Question 5

Diffraction waves are excited …

see result

2.5 Wave propagation in layered media
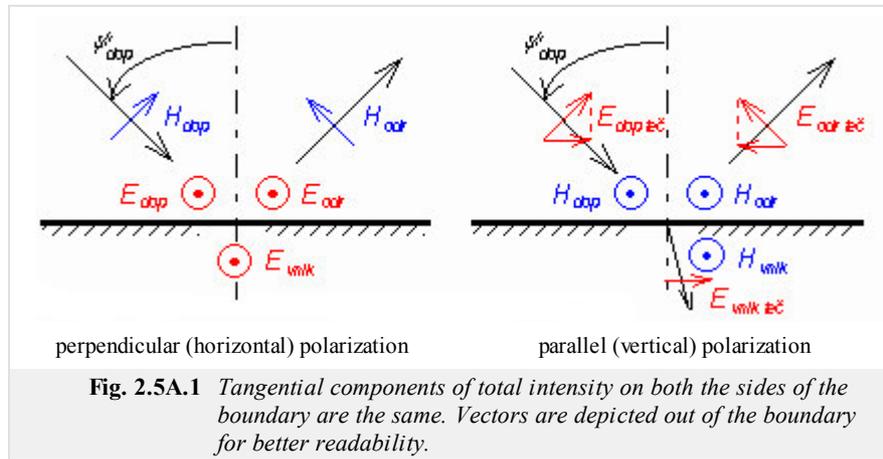
Basic theory

In several technical equipments, media consisting of several layers, which differ in their permittivity, permeability and thickness, are used. Hence, the layered medium is composed of a certain number of plan-parallel dielectric layers (or slabs).

The layered media (or *structures*) enable to build reflection-free walls. Those layers absorb the wave incident to a surface, which is covered by several dielectric layers of properly chosen parameters, without exhibiting significant reflections (reflection coefficient approaches zero). Applications can be found in anechoic chambers or at military equipment, e.g.

The layered structures can provide a reflection-free transition of a wave from one media to another one (from air to water, e.g.) in the sense of impedance matching. As a *matching circuitry*, a set of several dielectric layers, inserted between the media, can be considered. To this set of applications, dielectric radomes of antennas and HF devices belong. The radome has to preserve the device mechanically, but it must not influence the propagation of the electromagnetic wave.

In the following, we consider a normal incidence of waves with respect to the boundary among layers. The explanation is divided into two parts. In the first part (layer A), phenomena inside the layered medium are observed (changes of intensities on boundaries and inside layers). In the second part (layer B), the phenomena are investigated from the global point of view and parameters of the medium are computed in its whole.

In the layered medium, multiple reflections and transmissions of waves on single boundaries can be observed. Therefore, important terms related to the reflection and to the transmission are reminded.

perpendicular (horizontal) polarization   parallel (vertical) polarization

Fig. 2.5A.1 *Tangential components of total intensity on both the sides of the boundary are the same. Vectors are depicted out of the boundary for better readability.*

Plane of boundary is a plane separating two media. Plane of incidence is perpendicular to the plane of boundary and is parallel to the direction of the wave propagation. If the wave is of the perpendicular (horizontal) polarization, the vector E is perpendicular to the plane of incidence. If the wave is of the parallel (vertical) polarization, the vector E is parallel to the plane of incidence. Orientation of vectors for both the polarizations depicted in fig. **2.5A.1**. Plane of boundary is perpendicular to the plane of the screen. Plane of incidence is identical with the plane of the screen.

Each point on the boundary has to be considered as two points in an infinitely short distance. The first point is in the first medium, the second point belongs to the second medium. In each point in the plane of boundary, two various field intensities are therefore considered: E_1, E_2 and H_1, H_2 - each one in another medium. However, every time the tangential components of electric field intensity and the normal components of electric displacement density have to equal:

$$E_{1te\check{c}} = E_{2te\check{c}}, \quad \varepsilon_1 E_{1norm} = \varepsilon_2 E_{2norm} \quad \text{(on plane of boundary)} \qquad (\,2.5A.1\,)$$

Similar relations are valid even for vector H if no currents flow on the boundary. From the basic boundary condition (2.5A.1a), important relations among intensities, which are depicted in fig. **2.5A.1**, can be derived:

For perpendicular polarization:

$$E_{vnik} = E_{dop} + E_{odr} = E_{dop}\left(1 + \rho_{\perp}\right). \qquad (\,2.5A.2a\,)$$

For parallel one:

$$E_{vnik}\cos\psi_{vnik} = E_{dop}\cos\psi_{dop} - E_{odr}\cos\psi_{odr} = E_{dop}\left(1 - \rho_{\parallel}\right)\cos\psi_{dop}. \qquad (\,2.5A.2b\,)$$

The $\rho_{\perp}$ and $\rho_{\parallel}$ are reflection coefficients for the perpendicular polarization and the parallel one.

Generally,

$$\rho = \frac{E_{odr}}{E_{dop}}.$$

As already mentioned, we are going to deal with normal wave incidence to the boundary only. Reflection coefficients depend then on wave impedances Z_{01} and Z_{02} of both media only

$$\rho_\perp = \frac{Z_{02} - Z_{01}}{Z_{02} + Z_{01}}, \quad \rho_\parallel = \frac{Z_{01} - Z_{02}}{Z_{01} + Z_{02}} = -\rho_\perp. \tag{2.5A.3}$$

Considering the normal incidence, the difference between polarizations disappears. Reflection coefficients differ in the sign, but the positive direction of E_{odr} differs in the orientation on the other hand (see fig. **2.5A.1**), and therefore, the result for both polarizations is the same.

During the wave propagation in a layered medium, the same phenomenon repeats many times: the wave transmits the boundary and then, it freely propagates through the dielectric layer towards the next boundary. From the physical and the mathematical point of view, only two phenomena are needed to be managed: the boundary transition and the layer propagation.
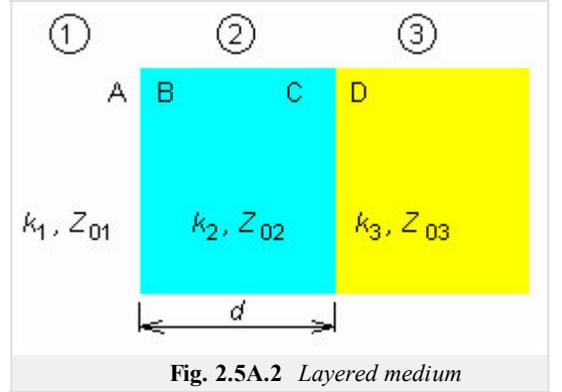
The explanation will be given for the situation, which is depicted in fig. **2.5A.2**. Here, three media 1 to 3 are shown. Their wave impedances $Z_{0i} = (\mu_i / \varepsilon_i)^{1/2}$ and wave

Fig. 2.5A.2 *Layered medium*

numbers $k_i = \omega (\mu_i \varepsilon_i)^{1/2}$ are indexed the same way. The media are separated by two boundaries, which consist of two planes (A, B; C, D): one of them is expected to be in the first medium and the second of them in the second medium. The wave propagates from left, transits the boundary perpendicularly, and is perpendicularly polarized; vectors E and H are tangential to the boundary. In the general case, both the forward wave and the backward one propagate in all the media, and therefore, the standing wave appears there. Intensities (E and H) of the forward wave are denoted by the index p, intensities of the backward wave by the index z. The other index specifies the position where the intensity is measured. Total intensities do not contain any index (neither p nor z)

$$E = E^{(p)} + E^{(z)}, \quad H = H^{(p)} - H^{(z)} \tag{2.5A.4}$$

(minus in (2.5A.4b) corresponds with fig. **2.5A.1**). Considering the above-introduced notation, E_B is total intensity on the boundary 1-2 in the medium 2, $E_A^{(z)}$ is intensity of reflected wave on the same boundary but in the medium 1.

Analysis of the wave propagation in the layered medium can be performed using two approaches. The difference between approaches is formal, but in given situations, it can influence the way of computing. No matter which approach is used, the analysis is done from the end of the medium to its beginning (in the contra-direction of the forward-wave propagation).

The first approach is based on the boundary condition (2.5A.1). Total intensities on both sides of the boundary have to be the same. E.g., for the boundary 2-3, we get:

$$E_C^{(p)} + E_C^{(z)} = E_D^{(p)} + E_D^{(z)}, \tag{2.5A.5a}$$

$$H_C^{(p)} - H_C^{(z)} = H_D^{(p)} - H_D^{(z)}. \tag{2.5A.5b}$$

In the second equation, we eliminate H substituting $H^{(p)} = E^{(p)} / Z_0$, $H^{(z)} = E^{(z)} / Z_0$ and we obtain two equations for $E_C^{(p)}$ and $E_C^{(z)}$, when quantities in the third medium are considered as known (the analysis is performed from the end). These equations yield:

$$E_C^{(p)} = \frac{1}{2}\left[E_D^{(p)}\left(1 + \frac{Z_{02}}{Z_{03}}\right) + E_D^{(z)}\left(1 - \frac{Z_{02}}{Z_{03}}\right) \right] = \frac{1}{2}\left[\left(E_D^{(p)} + E_D^{(z)}\right) + \left(E_D^{(p)} - E_D^{(z)}\right)\frac{Z_{02}}{Z_{03}} \right], \tag{2.5A.6a}$$

$$E_C^{(z)} = \frac{1}{2}\left[E_D^{(p)}\left(1 - \frac{Z_{02}}{Z_{03}}\right) + E_D^{(z)}\left(1 + \frac{Z_{02}}{Z_{03}}\right) \right] = \frac{1}{2}\left[\left(E_D^{(p)} + E_D^{(z)}\right) - \left(E_D^{(p)} - E_D^{(z)}\right)\frac{Z_{02}}{Z_{03}} \right]. \tag{2.5A.6b}$$

That way, the transition through the boundary is solved out. Dealing with the wave propagation inside a layer, well-known relations can be used:

$$E_B^{(p)} = E_C^{(p)} \exp\left(+jk_2 d\right), \tag{2.5A.7a}$$

$$E_B^{(z)} = E_C^{(z)} \exp\left(-jk_2 d\right). \tag{2.5A.7b}$$

Here, d is thickness of the layer (in the first equation, we move in the contra-direction of the wave propagation, and therefore, the exponent contains the sign 'plus' in the second equation, the situation is opposite). Intensities $E_2^{(p)}$ and $E_2^{(z)}$ play the role of input quantities for the foregoing boundary 1-2. During the analysis of the wave propagation in the layered medium, equations (2.5A.6) and (2.5A.7) are repetitively applied (when indexes are properly changed).

Whenever *inside* the layer, total field intensities can be computed substituting the coordinate of the observation point instead of the thickness d to (2.5A.7). Subsequently, (2.5A.4) is used. Reflection coefficient and standing wave ratio (in an arbitrary point) can be computed as:

$$\rho = \frac{E^{(z)}}{E^{(p)}}, \quad PSV = \frac{1+\rho}{1-\rho}. \tag{2.5A.8}$$

The question is, why the analysis is performed from the end of the medium to its beginning when the field intensity is usually known at its beginning. The reason is given by the physical matter of the fact. Relations in the medium depend on the ending of the medium. Eqns. (2.5A.6) and (2.5A.7) formally enables us to perform the analysis from the beginning of the medium to its end, but in a real situation, even if the total field intensity is known at the beginning, we are not able to decompose it to the forward-wave intensity and the backward-wave one. And *both* these intensities play their roles in eqns. (2.5A.6) and (2.5A.7). On the contrary, the situation at the end of the medium is quite different. Eliminating special situations, the structure is usually ended by free space or perfectly conducting plane. In the case of free space, there is no reflected wave at the end ($E^{(z)} = 0$) and a *single* intensity $E^{(p)}$ has to be known for the computations. If the medium is ended by a conducting wall, we know that $E^{(z)} = -E^{(p)}$, and again, a single intensity has to be known. This intensity can elected as unitary one and the computation is performed towards the known intensity at the input. Then, all the intensities are recomputed by a proper ration (all the relations are linear).

The second approach to the analysis of layered media is based on the transformation of reflection coefficient. This approach is well known from the analysis of the cascade of transmission lines. In both the cases, the situation is quite the same: several lines (media), which are arranged into the cascade, are of their own characteristic (wave) impedance Z_0 and their own length (thickness) d. Voltages and currents on the transmission line are related to field intensities E and H in an dielectric medium.

The solution consists of the following steps. First, reflection coefficient on the last boundary in the last but one medium (see fig. **2.5A.2**) is computed:

$$\rho_C = \frac{Z_{03} - Z_{02}}{Z_{03} + Z_{02}}. \tag{2.5A.9}$$

Reflection coefficient is recomputed to the beginning of the last but one medium, to the plane B:

$$\rho_B = \rho_C \exp(-2jkd). \tag{2.5A.10}$$

In order to compute reflection coefficient on the plane A, impedance on the plane B has to be determined first:

$$Z_B = Z_{02} \frac{1+\rho_B}{1-\rho_B} \tag{2.5A.11}$$

and consequently, reflection coefficient on A can be obtained:

$$\rho_A = \frac{Z_B - Z_{01}}{Z_B + Z_{01}}. \tag{2.5A.12}$$

If the structure consists of more layers, then the described computations are repeated considering the proper exchange of indexes.

This part of the solution has to be again led from the end of the structure to its beginning. Field intensities can be computed then in an arbitrary direction from a known intensity to unknown ones. If the total electric field intensity is known on the plane A (E_A), e.g., we decompose it to the intensities of the forward wave and the backward one first. Since

$$E_A^{(p)} + E_A^{(z)} = E_A^{(p)} + \rho_A E_A^{(p)} = E_A,$$

we get

$$E_A^{(p)} = \frac{E_A}{1+\rho_A}, \quad E_A^{(z)} = E_A^{(p)} \rho_A. \tag{2.5A.13}$$

The same decomposition can be done in an arbitrary medium (layer) when properly changing indexes. In order to evaluate field intensities inside the structure, the boundary condition (2.5A.1) and the expansion (2.5A.13) are applied. E.g.

$$E_B = E_A, \quad E_B^{(p)} = \frac{E_B}{(1+\rho_B)}, \quad E_B^{(z)} = E_B^{(p)} \rho_B.$$

Using (2.5A.7), we compute $E_C^{(p)}$ and $E_C^{(z)}$ and consequently the final E_C, etc. The intensities can be evaluated using (after a small rearrangement) equations (2.5A.5) or (2.5A.6).

Another way of the analysis of the layered medium is described in the layer B. In this part of the textbook, even more detailed information about the application of layered structures is given.

2.5 Wave propagation in layered media

Advanced theory

In the layer A, we introduced two approaches, which enable to solve electromagnetic wave propagation in a layered medium. The first approach was based on the distribution of field intensities and on the basic boundary condition. The second approach was based on the transformation of reflection coefficient and was identical with the method, which is used for solving similar tasks on transmission lines, in a fact. In this layer, we are going to introduce another approach, which makes the solution more formal, and therefore, it is suitable for more complicated situations.

The explanation is done for three media only (two boundaries). The situation is depicted in fig. **2.5B.1**. We can simply imagine multiple reflections appearing in the structure. Therefore, there are an infinite number of waves propagating in each layer in both directions. All the waves are coherent, and therefore, all the waves propagating in a given direction interfere into a single harmonic wave. Hence, only two waves in each medium are considered as depicted in fig. **2.5B.1**. The media are denoted as 1, 2, and 3. Next, A is the surface of the boundary 1-2 in the medium 1; B is the surface of the boundary 1-2 in the medium 2, etc. Using indexes A, B, C, D, we denote the surface, where the intensity is computed, and using indexes in brackets (p), (z), we denote the direction of propagation (forward wave, backward wave). E.g., $E_B^{(z)}$ denotes the intensity of the backward wave on the surface B, i.e. on the boundary 1-2 in the medium 2. The layered medium consists of boundaries and layers; we examine separately the wave transition through the boundary and through the layer.

Fig. 2.5B.1 *Wave propagation in layered medium*
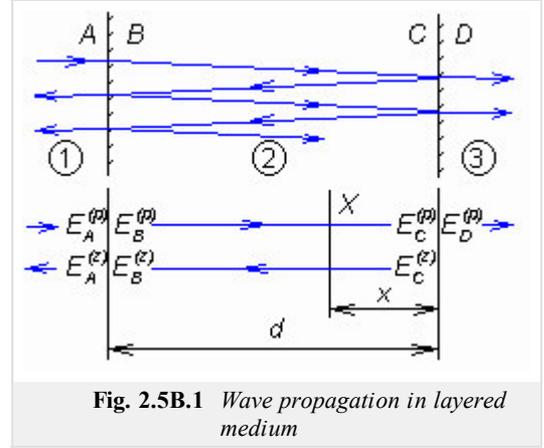
We consider the boundary 1-2 and express wave intensities, which travel outwards this boundary. The reflected wave of the intensity $E_A^{(z)}$ is created by the reflection of the incident wave $E_A^{(p)}$, and by the transition of the wave $E_B^{(z)}$ from the medium 2 to the medium 1. Similarly, the wave of the intensity $E_B^{(p)}$ is created by the transition of the wave $E_A^{(p)}$ from the medium 1 to the medium 2, and by the transition of the wave $E_B^{(z)}$. This is expressed by the equation:

$$E_A^{(z)} = \rho_{12} E_A^{(p)} + \tau_{21} E_B^{(z)}, \quad E_B^{(p)} = \tau_{12} E_A^{(p)} + \rho_{21} E_B^{(z)} \qquad (\,2.5B.1\,)$$

where ρ_{12} is reflection coefficient of the wave propagation in the medium 1 and reflecting from the boundary with the medium 2, and ρ_{21} is reflection coefficient of the wave propagating in the medium 2 and reflecting from the boundary with the medium 1. Similarly, coefficients of transmission are indexed as τ_{12} and τ_{21}. Eqn. (2.5B.1) can be rewritten to the matrix form:

$$\begin{bmatrix} E_A \\ E_B \end{bmatrix} = \begin{bmatrix} \rho_{12} & \tau_{21} \\ \tau_{12} & \rho_{21} \end{bmatrix} \begin{bmatrix} E_B \\ E_A \end{bmatrix}. \qquad (\,2.5B.2\,)$$

The right-hand side matrix, which is composed of coefficients of reflection and transmission, is called the scattering matrix of the boundary. The scattering matrix can be simply built, and moreover, their elements are of the obvious physical meaning (coefficients of reflection and transmission). Unfortunately, eqn. (2.5B.2) is not suitable for the solution of the problem because intensities of input and output waves are mixed in input and output column matrices. Eqn. (2.5B.2) is therefore solved for $E_A^{(p)}$ and $E_A^{(z)}$. Rearranging the relation, we consider $\rho_{21} = -\rho_{12}$ and $\tau_{12}\,\tau_{21} - r_{12}\,\rho_{21} = 1$. The result is expressed in the matrix form:

$$\begin{bmatrix} E_A^{(p)} \\ E_A^{(z)} \end{bmatrix} = \frac{1}{\tau_{12}} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} E_B^{(p)} \\ E_B^{(z)} \end{bmatrix}. \qquad (\,2.5B.3\,)$$

The right-hand matrix is called the *cascade boundary matrix*. The left-hand column matrix consists of input-wave intensities, and the right-hand column matrix consists of output-wave intensities. The cascade matrix is suitable for cascade connection of elements (two-ports), which are present even in our task. If an arbitrary number of elements are connected into the cascade, then the cascade matrices are simply mutually multiplied.

Now, the cascade matrix of the dielectric layer has to be derived. This aim can be met without any need for computing the scattering matrix. In the dielectric layer, there are two traveling waves propagating in opposite directions. Obviously,
$E_B^{(z)} = E_C^{(z)} exp(-jk_2d)$, $\quad E_B^{(p)} = E_C^{(p)} exp(+jk_2d)$, and therefore

$$\begin{bmatrix} E_B^{(p)} \\ E_B^{(z)} \end{bmatrix} = \begin{bmatrix} \exp(+jk_2 d) & 0 \\ 0 & \exp(-jk_2 d) \end{bmatrix} \begin{bmatrix} E_C^{(p)} \\ E_C^{(z)} \end{bmatrix}, \qquad (\,2.5B.4\,)$$

where k_2 is wave number in the second medium.

In the system of plan-parallel layers, single boundaries and layers are lined in parallel way. The final equation can be therefore obtained by multiplying respective matrices. For the situation from fig. **2.5B.1**, we get:

$$\begin{bmatrix} E_A^{(p)} \\ E_A^{(z)} \end{bmatrix} = \frac{1}{\tau_{12}} \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix} \begin{bmatrix} \exp(+jk_2d) & 0 \\ 0 & \exp(-jk_2d) \end{bmatrix} \frac{1}{\tau_{23}} \begin{bmatrix} 1 & \rho_{23} \\ \rho_{23} & 1 \end{bmatrix} \begin{bmatrix} E_D^{(p)} \\ 0 \end{bmatrix}. \qquad (2.5B.5)$$

The final result supposes a traveling wave in the third medium. In the opposite case, the intensity $E_D^{(z)}$ appears in the last column matrix. The way of extending the principle to an arbitrary number of layers is obvious.

Rewriting (2.5B.5), we get:

$$E_A^{(p)} = \frac{e^{jk_2d} + \rho_{12}\rho_{23}e^{-jk_2d}}{\tau_{12}\tau_{23}} E_D^{(p)}, \quad E_A^{(z)} = \frac{\rho_{12}e^{jk_2d} + \rho_{23}e^{-jk_2d}}{\tau_{12}\tau_{23}} E_D^{(p)}. \qquad (2.5B.6)$$

The ratio $E_A^{(z)}/E_A^{(p)}$ gives the real value of the reflection coefficient on the boundary 1-2. The ratio $E_D^{(p)}/E_A^{(p)}$ equals to the total transmission coefficient. Both the quantities can be computed using the last equations.

Let us note that in the case of the normal incidence of a perpendicularly polarized wave to a *single* boundary (no reflected wave can exists behind that), following relations are valid for the reflection coefficient and the transmission one:

$$\rho_\perp = \frac{Z_{02} - Z_{01}}{Z_{02} + Z_{01}}, \quad \tau_\perp = 1 + \rho_\perp = \frac{2Z_{02}}{Z_{02} + Z_{01}}. \qquad (2.5B.7)$$

The described method is suitable for more complicated structures. Common tasks can be numerically solved using Smith chart or directly considering boundary conditions on the boundary (see layer A.)

As already mentioned in the introduction, provision of the reflection-free transition of the electromagnetic wave from one medium to another one is an example of the application of layered media. Such requirement can appear when exciting waves in the ground or in the other media, when exciting waves in human body, at dielectric radomes of antennas, etc.; the wave is required to enter the medium and to exit it with the same intensity and the same equiphase surface. Covering the boundary between the media by one or several dielectric layers of exactly given permittivity and thickness can fulfill the requirement. The layers play the role of the matching circuitry (transformer) in analogy to the transmission line. In both the cases, electromagnetic wave propagates in the structure. On the transmission line, voltages and currents represent the wave. In the dielectric layer, the wave is represented by electric-field intensity and magnetic-field intensity. Voltage U corresponds to electric intensity E, and current I corresponds to magnetic intensity H. Characteristic impedance of the transmission line corresponds to wave impedance of the medium (layer). The difference between the above-described cases is hidden in the structure of circuits. On the transmission line, matching circuitry could consist not only of cascade-connected segments of the transmission line but too of parallel or serial shunts. Matching layers can be connected in cascade only. The reason for that is fundamental. In the part of the transmission line, where the parallel reactance is connected, a current jump appears (a part of current flows to the reactance). In the layered equivalent, a jump of magnetic-field intensity should appear in the respective part of the boundary. This is impossible from the physical point of view (it contradicts the basic boundary condition). In general, jump of magnetic-field intensity H can appear on the boundary if electric current is admitted. Then, the boundary condition is of the form $\mathbf{n} \times \mathbf{H}_1 = \mathbf{K} + \mathbf{n} \times \mathbf{H}_2$. Indexes distinguish between the first medium and the second one, vector products of the normal $\mathbf{n}$ result in tangential components and $\mathbf{K}$ [A m^{-1}] is surface density of electric current on the boundary. In practical life, this solution can be used. However, the elected technology has to admit the flow of electric currents on the surface of the boundary.

Routinely, we can design matching circuits consisting of a single one-quarter-wavelength thick dielectric layer or of a first dielectric layer of a given thickness plus of a second one, which is one-quarter-wavelength thick. No matter whether segments of the transmission line or dielectric layers are used. If efficient, Smith chart can be used.

The above-described matching circuits can be modeled by the computer program, which is described in the layer C. The program cannot be used for the design of those circuits, it analyses a circuitry of given parameters (permittivity, thickness of layers) and computes the distribution of E and H in layers, reflection coefficient and standing wave ratio in various points of the structure. Observing numerical values, the circuit operation can be investigated. The matching circuits can consists of up to four dielectric layers. The operation of the circuitry can be observed in a wider frequency band.

Frequency filters are another application area of layered media. As a basic element, the half-wavelength-thick dielectric layer is used. This layer (in analogy with the half-wavelength-long transmission line) transforms intensities and impedances in the ratio 1:1. Other words, the wave reflected at the beginning of the layer cancels out with the wave reflected from the end of layer. If the incident wave is of different frequency (the layer is not half-wavelength thick), the total cancellation does not appear. The layer reflects the part of energy, and therefore, the transmitted energy is lower. In front of the layer, the standing wave appears.

In the practical life, filters are composed of several half-wavelength-thick layers, which are mutually separated by layers of another permittivity

and thickness ($l/4$, e.g.).

Properties of layered structures as filters can be demonstrated by another program in the layer C. In this layer, even the more detailed description of the program is given.

2.5 Wave propagation in layered media

Matlab program

The program simulates the wave transition through several plan-parallel dielectric layers (through several boundaries).

First, the path of Matlab has to be set to the directory **Layers**. The program is run via the m-file `vrstvy`. Closing the introductory window, you are asked to provide input parameters: the number of media (incl. the input medium and the output one), the central frequency (where the distributions of the electric field intensity and the magnetic one are computed), and the bandwidth (where the frequency course of the reflection coefficient, of the standing wave ratio and of the power transmittance P_{out} / P_{in} are computed).

Confirming the number of layers, the menu for specifying parameters of the media is updated. The user is asked for the length and the dielectric constant of the layer. If a new layer is added, parameters of the previously specified layers are preserved. Pushing the button $\boxed{\text{Calculate}}$, the computation is started. In case some parameters are improperly specified, input lines to be corrected are shown in red. Results are displayed in a new window. The user can consequently modify the number of layers and the parameters of layers, and can compare newly computed results with previously calculated ones. Pushing the button $\boxed{\text{Close graphs}}$, the windows with results can be closed. The button $\boxed{\text{End}}$ ends the program.

Two charts in a single window graphically represent results. The upper chart displays the distribution of the electric field intensity (blue) and the magnetic one (green). In the last layer, a unitary value of electric field intensity is assumed, and the last layer is expected to be completed by a reflection-free termination (i.e., no reflected wave exists in the last layer on any frequency). The lower chart displays frequency course of reflection coefficient ρ (blue), of transmittance (red), and of standing wave ration (green, the scale on the next axis). If the graphical results are not needed more, the user can return back to the previous window, finish the program, insert new parameters or repeat the calculation.

2.5 Wave propagation in layered media

Developing Matlab

Subprograms of the main program `layers.m` read the input quantities and send them to the computational kernel `compute.m`. Here, the basic quantities (wavelength, e.g.) are computed first, using elementary relations:

```
lambda0 = 300/f*1000;
k0 = 2*pi/lambda0;
Z0 = 120*pi./sqrt(e);
K = k0*sqrt(e);
Lambda = 2*pi./k;
```

Second, reflection coefficient is computed for the central frequency in the m-file `compu_v.m`:

1. First, characteristic impedance is determined

   ```
   Z0 = 120*pi./sqrt(e);
   ```

2. Second, impedance at the beginning of all the layers is computed:

   ```
   for i=(I):-1:1
       Zvrch=Z0(i)*(cos(k(i)*l(i))+sqrt(-1)*Z0(i)/Z(i+1)*sin(k(i)*
       (l(i))));
       Zspod=(Z0(i)/Z(i+1)*cos(k(i)*l(i))+sqrt(-1)*sin(k(i)*(l(i))));
       Z(i)=Zvrch/Zspod;
   end
   ```

3. Next, reflection coefficient on the boundaries can be determined:

   ```
   for i=I:-1:1
       ro(i)=(Z(i+1)-Z0(i))/(Z(i+1)+Z0(i));
   end
   ```

4. Finally, standing wave ratio in layers is computed:

   ```
   for i=I:-1:1
       PSV(i)=(1+abs(ro(i)))/(1-abs(ro(i)));
   end
   ```

Inside the layers, intensities of forward and backward waves are computed the following way:

```
Edop(i)= Edop(i+1)*exp(-sqrt(-1)*k(cislo_vrstvy(i))*...
(pole_vzd(i)-pole_vzd(i+1)));
Eodr(i)= Eodr(i+1)*exp(sqrt(-1)*k(cislo_vrstvy(i))*...
(pole_vzd(i)-pole_vzd(i+1)));
```

The total electric-field intensity is computed by adding intensities of forward and backward waves:

```
E(i)=Edop(i)+Eodr(i);
```

Transmitting the boundary between two layers:

```
Edop(i)=E(i)/(1+ro(cislo_vrstvy(i)));
Eodr(i)=ro(cislo_vrstvy(i))/(1+ro(cislo_vrstvy(i)))*E(i);
E(i)=Edop(i)+Eodr(i);
```

Magnetic-field intensity can be computed as:

```
H(i)=(Edop(i)-Eodr(i))/Z0(cislo_vrstvy(i));
```

Transmittance is given then by the relation:

```
T=1-(abs(ro(1)))^2;
```

Frequency course of standing wave ratio, transmittance and reflection coefficient is computed using the above-given relations.

2.5 Wave propagation in layered media

Java applet

2.5 Wave propagation in layered media

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

If the vector of the intensity of electric field is perpendicular to the plane of incidence, then …

> Possible answers for question 1:
> - ⊙ … cannot say anything about the wave polarization.
> - ⊙ … we speak about the wave of parallel polarization.
> - ⊙ … we speak about the perpendicularly polarized wave.
> - ⦿ I won't answer this question

Question 2

Wave impedance of a layer Depends on …

> Possible answers for question 2:
> - ⊙ … permittivity and permeability of the layer.
> - ⊙ … permittivity and permeability of the layer, and is influenced by the parameters of neighboring layers.
> - ⊙ … permittivity, permeability and thickness of the layer.
> - ⦿ I won't answer this question

Question 3

In case of the perpendicular incidence of the electromagnetic wave …

> Possible answers for question 3:
> - ⊙ … the wave goes through layers without any reflection.
> - ⊙ … there is no difference between perpendicular and parallel polarization.
> - ⊙ … we have to distinguish between polarizations.
> - ⦿ I won't answer this question

Question 4

Designing the matching circuits …

> Possible answers for question 4:
> - ⊙ … we can apply layers as parts of transmission lines in a cascade.
> - ⊙ … we can apply layers the same way as parts of transmission lines (incl. shunts).
> - ⊙ … we cannot apply layers.
> - ⦿ I won't answer this question

Question 5

Analyzing wave propagation in a layered media, we have assumed the wave is …

Possible answers for question 5:

- ○ … planar.
- ○ … spherical.
- ○ … cylindrical.
- ● I won't answer this question

see result

Chapter 3: Waves on transmission lines

Introduction

In the previous chapter of the textbook, we studied EM wave propagation in free space, where various types of obstacles were present. In the case of Fresnel diffraction (par. 2.1), influence of thin, perfectly absorbing half-plane to the propagation of plane wave was studied. Considering general theory of diffraction (par. 2.2) the effect of infinitely long perfectly conductive cylinder, which is placed to the trajectory of a plane wave, was investigated. Next, propagation of plane wave in the layered medium was observed (par. 2.5). Analyzing more complicated obstacles, geometrical optics (par. 2.3) and geometric theory of diffraction (par. 2.4) were exploited in order to investigate influence of such obstacles to wave propagation.

In any of the above-described situations, the wave was *fixed* to a structure - the wave freely propagated through a medium and passed the above-described obstacles.

Now, another situation is going to be investigated - we are going to study a wave which does not freely propagate, but which is guided by a certain structure, a certain transmission line.

Among most popular transmission lines, a coaxial transmission line belongs. This type of the transmission line consists of the inner conductor and the outer one. The gap between conductors is filled in by dielectrics (fig. **3.1a**).

Fig. 3.1 *a) Coaxial transmission line, b) rectangular waveguide*

Electromagnetic wave propagates through the dielectrics in the direction of conductors. The wave does not propagate freely, but the wave propagation is *guided* by the transmission line.

Coaxial transmission line is not studied in our textbook. The notion about the wave propagation can be simply built here on the basis of currents induced to conductors and voltages between conductors. We concentrate on structures, which are exploited on very high frequencies (in microwave frequency band). On those frequencies, usual coaxial transmission lines are not exploited due to high attenuation.

As the first representative of microwave transmission lines, a waveguide (par. 3.1, fig. **3.1b**) is introduced. The waveguide is a hollow metallic cylinder of a given cross-section (usually of a rectangular or circular shape). In the waveguide, a wave propagates in a way, which can be understood as sequential reflections from the walls. Wave propagation is computed by an analytical approach and by a numerical one (we exploit finite-element method).

As the second representative of microwave transmission lines, a microstrip transmission line (par. 3.2). is introduced. A microstrip transmission line consists of a relatively narrow strip conductor appearing on the true side of a dielectric substrate. A continuous metallic ground plate of zero potential covers the backside. The whole transmission line is usually placed to a rectangular waveguide, which serves as a shielding. Since we are not able to analyze the described transmission line by purely analytical approach, we perform the analysis numerically - by the finite-element method.
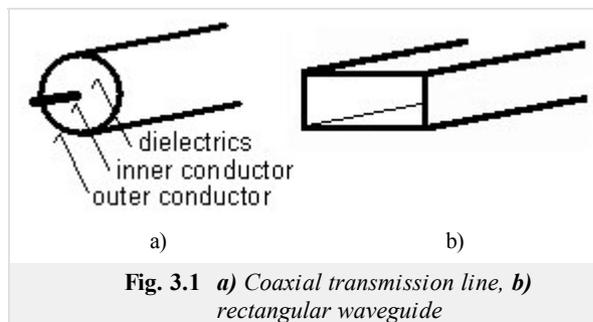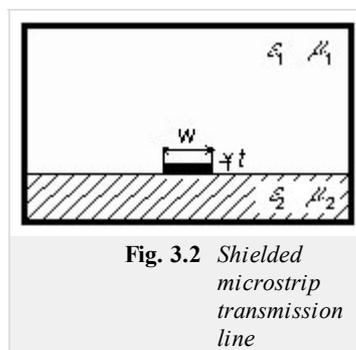

Fig. 3.2 *Shielded microstrip transmission line*

3.1 Waveguides

Basic theory

Our layer of this article brings basic information about the propagation of electromagnetic waves in waveguides. In the layer B, we present a Dutch translation of the article to the reader in order to help him (her) to become familiar with the terminology from the area of waveguides and microwave transmission lines.

Types of transmission lines we know from our everyday life (coaxial cables, etc.), can be used on microwave frequencies in limited number of applications due to unacceptable losses in the dielectrics of those lines. That is why waveguides are frequently used on higher frequencies.

The waveguide usually denotes a metallic cylinder which transversal dimensions are comparable to the wavelength. Inner walls of the waveguide are processed that way which minimizes losses in the metal (walls can be considered as a perfect electric conductor). The transversal profile of the waveguide is usually of the rectangular or circular shape. In special cases, the profile is of the shape of letters Π or H (fig. **3.1A.1**); such waveguides are of wider broad-band on one hand but they can transmit lower power on the other hand.
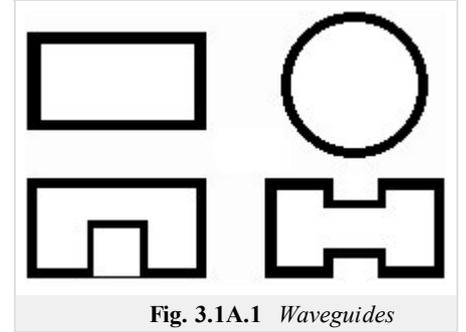
Fig. 3.1A.1 *Waveguides*

Waveguides can be met on gigahertz frequencies, because on lower frequencies, transversal dimensions of waveguides are too large. Waveguides are exploited in radars and in satellite communication systems. They serve here for the energy transmission from the generator to the antenna and vice versa.

Fig. 3.1A.2 *Longitudinally homogenous waveguide. u,v: transversal directions, z: longitudinal direction.*

Assume a longitudinally homogeneous metallic waveguide of an arbitrary transversal profile (fig. **3.1A.2**). Modeling electromagnetic field, which is produced by a source inside the waveguide, is the aim of our computations.

Assume that our computations are performed is large distance from sources. Then, computations can be based on the homogeneous wave equations for the longitudinal component of the electric Hertz vector Π_z^e and the magnetic one Π_z^m. Since the field distribution in the longitudinal direction z is not dependent on the field distribution in the transversal directions u, v, both Hertz vectors can be rewritten as the product of two functions where the first one is dependent on transversal coordinates u, v, and the second one depends on the longitudinal coordinate z, i.e. $\Pi_z = T_1(u,v) \times T_2(z)$. Exploiting the method of variables separation, we yield separation constants Γ and γ, which are related as

$$\gamma^2 + k^2 = \Gamma^2.$$ (3.1A.1)

In the above relation, Γ is the separation constant that is related to the field distribution in transversal directions, γ is related to the wave propagation in the longitudinal direction and k is wave-number for the free space (vacuum with permitivity ε_0 and permeability μ_0)

$$k = \omega \sqrt{\mu_0 \varepsilon_0}$$ (3.1A.2)

and ω denotes angular frequency.

The detailed mathematical description of the above approach can be found in **[1]**.

Solving the equation for the wave propagation in the longitudinal direction in the waveguide, following relation is obtained

$$T_2 = C_1 e^{\gamma z} + C_2 e^{-\gamma z}$$ (3.1A.3)

C_1 and C_2 are integration constants. First term describes the backward wave (it propagates in the contra-direction of the axis z), the second term is related to the forward wave (it propagates in the direction of the axis z). The separation constant γ is called the propagation constant and it can be rewritten as

$$\gamma = \beta + j\alpha.$$ (3.1A.4)

Considering (3.1A.4), β can be understood as the attenuation constant and α as the phase constant.

Since the wave-number k is real (a lossless medium is assumed inside the waveguide) and since the separation constant G is real too (as will be shown later), the propagation constant can be of the following values:

- $\gamma = \beta$ for $k < \Gamma$ in the longitudinal direction, evanescent wave propagates;
- $\gamma = \alpha$ for $k > \Gamma$ in the longitudinal direction, non-attenuated wave propagates;
- $\gamma = jk$ for $\Gamma = 0$ propagation properties are not influenced by the transversal profile.

Since the wave-number linearly depends on the frequency, an interesting conclusion can be done: whereas waves of lower frequency than

$$\omega_{krit} = \Gamma / \sqrt{\mu\varepsilon} \qquad (3.1A.5)$$

do not propagate in the waveguide, waves of higher frequency than ω_{krit} propagate without any attenuation. Frequency (3.1A.5) is called the critical frequency.
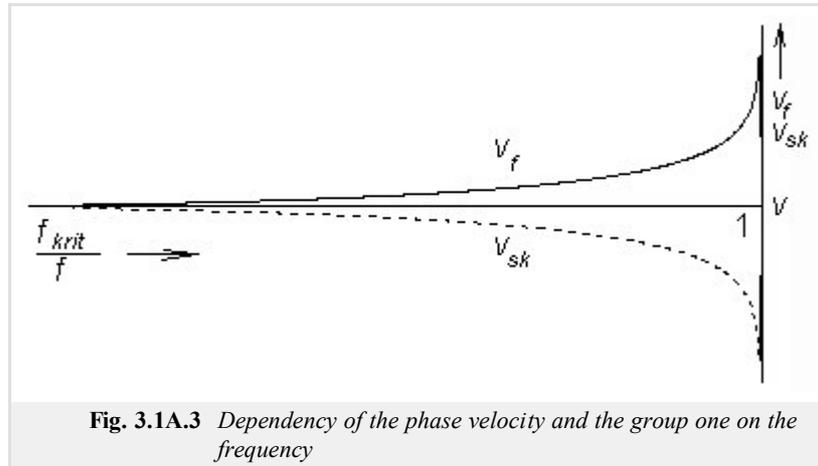
Fig. 3.1A.3 *Dependency of the phase velocity and the group one on the frequency*

Let us examine the phenomena appearing on frequencies higher than the critical one $f > f_{krit}$. Substituting $\gamma = j\alpha$ to (3.1A.1), we get

$$\alpha = \sqrt{k^2 - \Gamma^2} \qquad (3.1A.6)$$

Expressing the separation constant Γ from (3.1A.5)

$$\Gamma = \omega_{krit}\sqrt{\mu\varepsilon}$$

and using wave-number k from (3.1A.2) we get the relation for the phase constant in the longitudinal direction

$$\alpha = k\sqrt{1 - \left(f_{krit} / f\right)^2} \qquad (3.1A.7)$$

Substituting the phase constant (3.1A.7) to the relation for the phase velocity

$$v_f = \omega / \alpha,$$

we get dependency of the phase velocity in the waveguide on the frequency

$$v_f = \frac{v}{\sqrt{1 - (f_{krit}/f)^2}}. \qquad (3.1A.8)$$

Here, v denotes phase velocity of the wave in free space (vacuum in our case)

$$v = 1 / \sqrt{\mu\varepsilon}. \qquad (3.1A.9)$$

Using known phase velocity, relation between phase velocity and wavelength

$$\lambda_g = v_f / f \qquad (\ 3.1A.10\ )$$

in the longitudinal direction in the waveguide can be obtained

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (f_{krit}/f)^2}} \qquad (\ 3.1A.11\ )$$

Here, λ denotes wavelength in free space.

If we are interested in the velocity of energy propagation, the group velocity has to be computed. Since the product of the group velocity and the phase one has to equal to the square of the velocity of light, we can compute the group velocity using the following relation:

$$v_{sk} = v\sqrt{1 - (f_{krit}/f)^2}, \qquad (\ 3.1A.12\ )$$

where v is phase velocity in free space.

The above computations were performed for the propagation of harmonic waves. If the wave consists of more harmonics then every frequency component propagates with velocity, and therefore, the output signal differs from the input one (the signal is distorted). The described phenomenon is called dispersion (fig. **3.1A.4**).
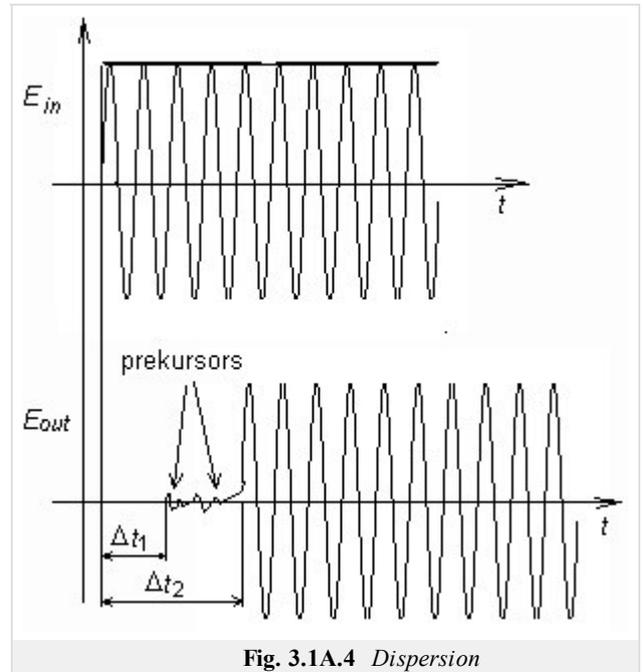
Fig. 3.1A.4 *Dispersion*

Up to now, we dealt with analysis in the longitudinal direction. Results of this analysis do not depend on the shape of the transversal profile, and therefore, they are valid for any homogeneous waveguide. In the transversal directions, the situation is totally different:

- In the transversal direction, no wave is propagating. Here, waves reflected from waveguide walls interfere, and hence, standing waves appear here.
- Since reflections from walls (and consecutively the standing waves) depend on the profile, the analysis has to be performed for the respective shape of the profile. In our description, we are going to concentrate on the rectangular shape

The analysis is done for two types of waves, which can propagate in the waveguide, for the transversally magnetic waves (TM, components of the magnetic intensity vector are non-zero in transversal directions only), and for the transversally electric ones (TE, components of the electric intensity vector are non-zero in transversal directions only). Following [1], relations for critical frequencies of waves of both types can be obtained:

$$\omega_{krit} = \frac{1}{\sqrt{\mu\varepsilon}}\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}. \qquad (\ 3.1A.13\ )$$

Here, permitivity and permeability are related to the medium inside the waveguide, a denotes the width of the waveguide and b is its height. Integral coefficients m and n are called mode numbers. Increasing mode numbers, critical frequency is increased (therefore, higher modes appear on higher frequencies).

Assume that a supplying generator is tuned from lower frequencies to higher ones. Reaching the critical frequency of the lowest mode, a single wave starts to propagate in the waveguide. If the critical frequency of the second mode is reached then two waves of two different modes propagate in the waveguide. Those waves can interfere, which can lead to many problems, Therefore, waveguides operate in the single-mode band usually. The low bound of this band is given by the critical frequency of the lowest mode, the high bound equals to the critical frequency of the second mode. The mode of the lowest critical frequency is called the dominant mode.

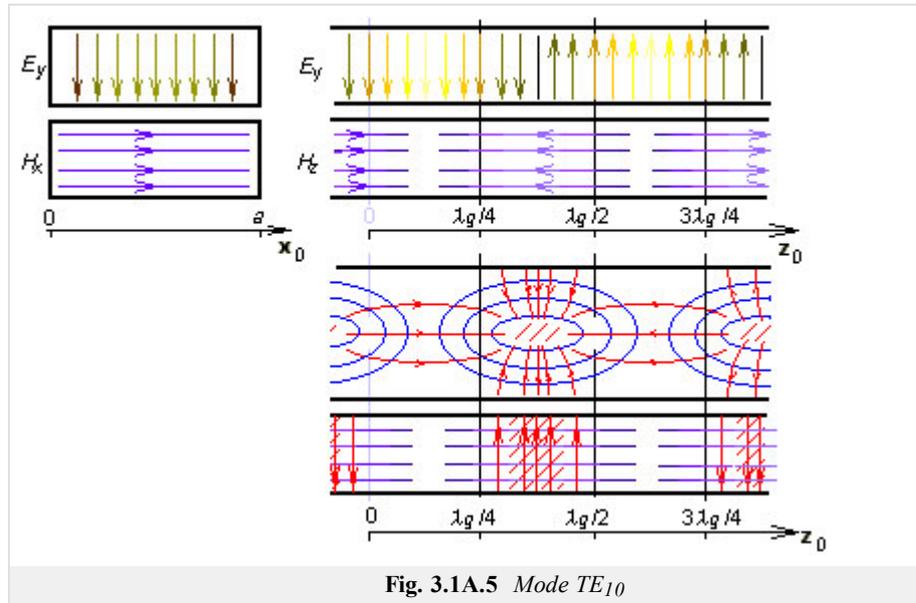Let us concentrate on the field distribution of the dominant mode TE_{10} in the moment $t = t_0$ (see animation).

Fig. 3.1A.5 *Mode TE$_{10}$*

Cutting the waveguide by a longitudinal wall that is perpendicular to the wider side, a harmonic course of the transversal component of electric intensity E_y can be *observed*. Maximal intensity in the figure is located to coordinates $z = \lambda_g/4$ and $z = 3\lambda_g/4$ (phase is opposite here). In $z = 0$ and $z = \lambda_g/2$, electric intensity is zero. In the points of maximal E_y, longitudinal component of magnetic intensity H_z is zero and transversal component of magnetic intensity H_x is maximal. In the transversal cut in $z = \lambda_g/4$, E_y is maximal in the center of the waveguide and zero on walls (boundary condition met). Transversal component of magnetic intensity H_x is in $z = \lambda_g/4$ constant.

If the cutting wall is perpendicular to the narrower side of the waveguide, magnetic intensities form ellipses (similarly as in the case of a wire flowing by a current). In case of waveguide, a current flowing from the low wall of the waveguide to the up one through the vacuum produces magnetic field. If the current reaches the wall, it flows in the form of the conducting current on the waveguide wall back.

Finally, we can state that the transversal electromagnetic field is distributed that way so that the boundary conditions on perfectly electrically conducting walls are met. This fact is illustrated by the Matlab program for the analysis of the electromagnetic field in the rectangular waveguide by the finite-element method. User's guide of the program can be found in the layer C. A brief introduction to the finite-element method and a brief description of the software implementation are given in the layer D.

3.1 Waveguides

Advanced theory

Deze pagina laag A bevat basisinformatie over de voortplanting van elektromagnetische golven in een golfgeleider. Op deze pagina (in laag B) vindt u een Nederlandse vertaling van dit artikel om zo de lezer vertrouwd te maken met termen in verband met elektromagnetische golven en transmissielijnen.

Verschillende soorten van transmissielijnen kennen we van uit het dagelijkse leven (coax kabel, enz.). Deze kunnen gebruikt worden voor microgolf frequenties in beperkte aantallen omwille van onaanvaardbare verliezen in de diëlectricums van die lijnen. Dit is waarom golfgeleiders regelmatig gebruikt worden op hogere frequenties.

De golfgeleider bestaat meestal uit een metalen cylinder waarvan de afmetingen vergelijkbaar zijn met deze van de desbetreffende golflengte. De binnenkant van de golfgeleider is zo gemaakt dat de verliezen in het metaal tot een minimum worden herleid (de behuizing kan gezien worden als een perfecte geleider). Het profiel van de golfgeleider is meestal cirkelvormig of rechthoekig. In speciale gevallen maakt men gebruik van het Π of H profiel (fig. **3.1B.1**); zulke golfgeleiders hebben een bredere bandbreedte maar kunnen minder vermogen uitzenden.
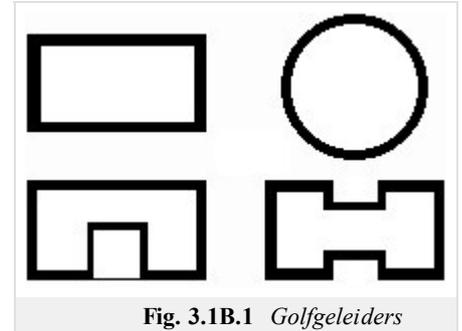
Fig. 3.1B.1 *Golfgeleiders*

Fig. 3.1B.2 *Longitudinally homogenous waveguide. u,v: transversal directions, z: longitudinal direction.*

Golfgeleiders worden meestal toegepast op gigahertz frequenties, omdat op lagere frequenties de afmetingen van de golfgeleider te groot worden. Golfgeleiders worden gebruikt in radars en sateliet communicatie systemen. Ze dienen hier voor energieverzending van de generator naar de antennen en omgekeerd.

Veronderstel een longitudinale homogene metalen golfgeleider van een willekeurig transversaal profiel (fig. **3.1B.2**). Het resulterende elektromagnetische veld, welk geproduceerd is door een inwendige bron in de golfgeleider, is het doel van onze berekeningen.

Veronderstel dat onze berekeningen uitgevoerd zijn op lange afstand van de bron. Dan kunnen berekeningen gebaseerd worden op de homogene golf regressievergelijking voor de longitudinale component van de elektrische Hertz vector Π_z^e en de magnetische vector Π_z^m. Sinds het distributieveld in de longitudinale richting z niet meer afhankelijk is van het distributieveld in de transversale richting u, v, beide Hertz vectors kunnen herschreven worden als het product van twee functies waarvan de eerste afhankelijk is van de transversale coördinaten u, v, en de tweede afhandelijk is van de longitudinale coördinaat z, vb. $\Pi_z = T_1(u,v) \times T_2(z)$. Wanneer we de scheidingsmethode van de variabelen toepassen, dan bekomen we een scheiding van de variabelen Γ en γ, welke zich onderling verhouden als

$$\gamma^2 + k^2 = \Gamma^2.$$ (3.1B.1)

In de bovenstaande vergelijking is Γ de constante die betrekking heeft op de velddistributie in transversale richtingen, γ betreft de golfverspreiding in de longitudinale richting en k is hetgolfnummer als men werkt in open lucht (vacuüm met permitiviteit ε_0 en permeabiliteit μ_0)

$$k = \omega \sqrt{\mu_0 \varepsilon_0}$$ (3.1B.2)

en ω staat voor de hoeksnelheid.

De gedetailleerde wiskundige beschrijving van de bovenstaande benadering kan men vinden in **[1]**.

Door het oplossen van de regressievergelijking van de golfverspreiding in de longitudinale richting in de golfgeleider bekomt men volgende vergelijking

$$T_2 = C_1 e^{\gamma z} + C_2 e^{-\gamma z}$$ (3.1B.3)

C_1 en C_2 zijn constanten. De eerste term beschrijft de terugwaardse golf (ze verspreidt zich in de tegengestelde richting van de *z-as*), de tweede term heeft betrekking op de voorwaardse golf (deze verspreidt zich in de richting van de *z-as*). De opsplitsing van de constante γ noemt men de verspreidings constante en deze kan herschreven worden als

$$\gamma = \beta + j\alpha.$$ (3.1B.4)

Veronderstellen we (3.1B.3) dan verstaan we onder β de verzwakkings constante en onder α de fase constante.

Omdat we rekening moeten houden met k (er wordt een min of meer verliesvrij medium verondersteld in de golfgeleider) en met de constante Γ (dit zal later worden aangetoond) zal de verspreidings constante de volgende waardes aannemen:

- $\gamma = \beta$ for $k < \Gamma$ in longitudinale richting;

- $\gamma = \alpha$ for $k > \Gamma$ in longitudinale richting, verliesvrije golfverspreiding;
- $\gamma = jk$ for $\Gamma = 0$ de verspreidingsproporties worden niet beinvloed door het transversale profiel.

Omdat het golfaantal recht evenredig is met de frequentie kunnen we hier een interessante conclusie maken: hoewel golven met een lagere frequentie

$$\omega_{krit} = \Gamma / \sqrt{\mu\varepsilon} \qquad (\,3.1B.5\,)$$

zich dan niet zullen verspreiden in de golfgeleiders, golven met een hogere frequentie dan ω_{krit} zullen zich verspreiden zonder verlies. Deze frequentie (3.1B.5) noemt men de kritische frequentie.
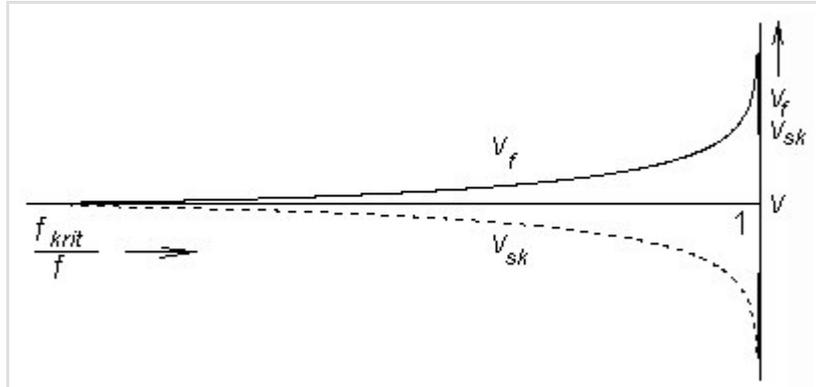
Fig. 3.1B.3 *Dependency of the phase velocity and the group one on the frequency*

Laten we ons verdiepen in het fenomeen dat zich voordoet bij frequenties hoger dan de kritische frequenties $f > f_{krit}$. Als we γ in de plaats stellen van $\gamma = j\alpha$ to (3.1B.1), krijgen we

$$\alpha = \sqrt{k^2 - \Gamma^2} \qquad (\,3.1B.6\,)$$

We bekomen de Γ door het omvormen van formule (3.1B.5) tot

$$\Gamma = \omega_{krit}\sqrt{\mu\varepsilon}$$

en gebruiken we golfnummer k van (3.1B.2) dan krijgen we de vergelijking voor de fase constante in de longitudinale richting

$$\alpha = k\sqrt{1 - \left(f_{krit}/f\right)^2} \qquad (\,3.1B.7\,)$$

Vervangen we de fase constante (3.1B.7) door de vergelijking van de fase snelheid

$$v_f = \omega / \alpha$$

dan krijgen we afhankelijk van de fase snelheid op de frequentie in de golfgeleider

$$v_f = \frac{v}{\sqrt{1 - \left(f_{krit}/f\right)^2}}. \qquad (\,3.1B.8\,)$$

In bovenstaande formule is v de fase snelheid van de golf in open lucht (vacuüm in ons geval)

$$v = 1 / \sqrt{\mu\varepsilon}. \qquad (\,3.1B.9\,)$$

Gebruiken we de gekende fase snelheid, dan staat de verhouding tussen de fasesnelheid en de golflengte

$$\lambda_g = v_f / f \qquad (3.1B.10)$$

in de longitudinale richting in de golfgeleider hier onder beschreven

$$\lambda_g = \frac{\lambda}{\sqrt{1-\left(f_{krit}/f\right)^2}} \qquad (3.1B.11)$$

Hier bedoelt men met λ de golflengte in open lucht.

Als we de snelheid van de energieverspreiding willen kennen, zullen we eerst de groep snelheid moeten berekenen. Als het product van de groep snelheid en de fase gelijk moet zijn aan het kwadraat van de snelheid van het licht, dan kunnen we de groep snelheid berekenen aan de hand van de volgende vergelijking:

$$v_{sk} = v\sqrt{1 - \left(f_{krit}/f\right)^2}, \qquad (3.1B.12)$$

hier is v de fase snelheid in open lucht.

De bovenstaande berekeningen werden verricht voor de verspreiding van harmonische golven. Men spreekt van spreiding (fig. **3.1B.4**) als iedere golf uit meerdere harmonische bestaat dan elke frequentie component verspreid én als het uitgangssignaal van het ingangssignaal (vervorming) verschilt..

Fig. 3.1B.4 *Dispersion*

Tot nu toe hebben we enkel geanalyseerd in de longitudinale richting. Deze resultaten hangen niet van de vorm van het transversale profiel af en zijn geldig voor elke homogene golfgeleider. In de transversale richtingen is de situatie totaal anders:

- In de transversale richting verspreidt de golf zich niet. De golf wordt gereflecteerd tegen de binnenkant van de golfgeleider, hierdoor ontstaan er staande golven.
- Omdat de reflecties afhangen van het profiel van de golfgeleider zal de analyse gedaan worden aan de respectievelijke vorm van het profiel. In onze beschrijving gaan we gebruik maken van een rechthoekige vorm.

De analyse is gedaan voor twee soorten golven, welke zich kunnen verspreiden in de golfgeleiders. Voor de transversale magnetische golf (TM, componenten van de magnetische intensiteitsvector zijn niet gelijk aan nul in de transversale richting), en voor de transversale elektrische golf (TE, componenten van de elektrische intensiteitsvector zijn niet gelijk aan nul in de transversale richtingen). Volgende **[1]**, vergelijking voor kritische golf frequenties van beide types kunnen herschreven worden als:

$$\omega_{krit} = \frac{1}{\sqrt{\mu\varepsilon}} \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}. \qquad (3.1B.13)$$

Hier zijn de permitiviteit en de permeabiliteit verwant aan het medium in de golfgeleider. a staat hier voor de breedte en b voor de hoogte. De integrale coëfficiënten m en n worden de mode nummers genoemd. Door het verhogen van de mode nummers zal ook de kritische frequentie verhogen (hoge modes komen voor bij hogere frequenties).

Veronderstel dat een voedende generator is afgesteld van lage naar hoge frequenties. Als de kritische frequentie van de laagste mode bereikt wordt, dan zal één enkele golf zich voortbewegen in de golfgeleider. Als de kritische frequentie van de tweede mode bereikt wordt zullen er zich twee golven (elk van een verschillende mode) verspreiden in de golfgeleider. Deze golven kunnen in elkaar overlopen wat tot vele problemen kan leiden. Daarom functioneren golfgeleiders meestal in de single-mode band. De ondergrens van deze band wordt gegeven door de kritische frequentie van de laagste mode, de bovengrens is gelijk aan de kritische frequentie van de tweede mode. De mode van de laagste kritische frequentie noemt men de dominante mode.

Laten we ons nu verdiepen in de veldverdeling van de dominante mode TE_{10} op het moment $t = t_0$ (zie animatie).
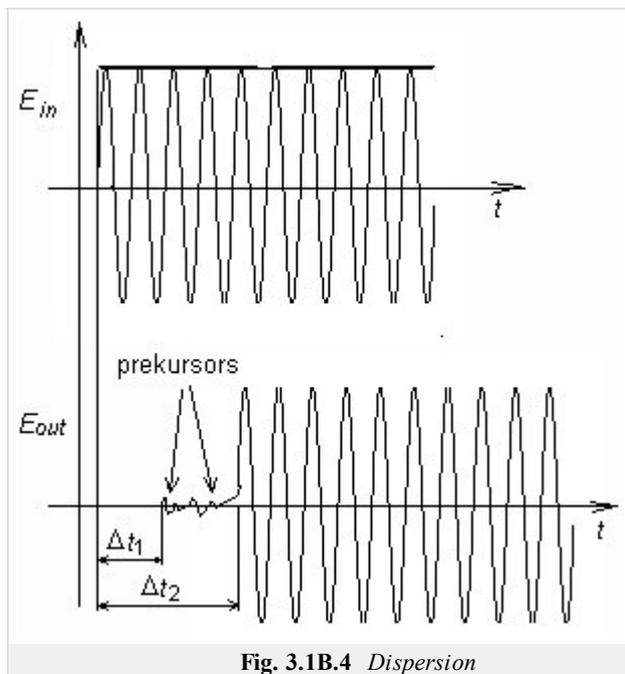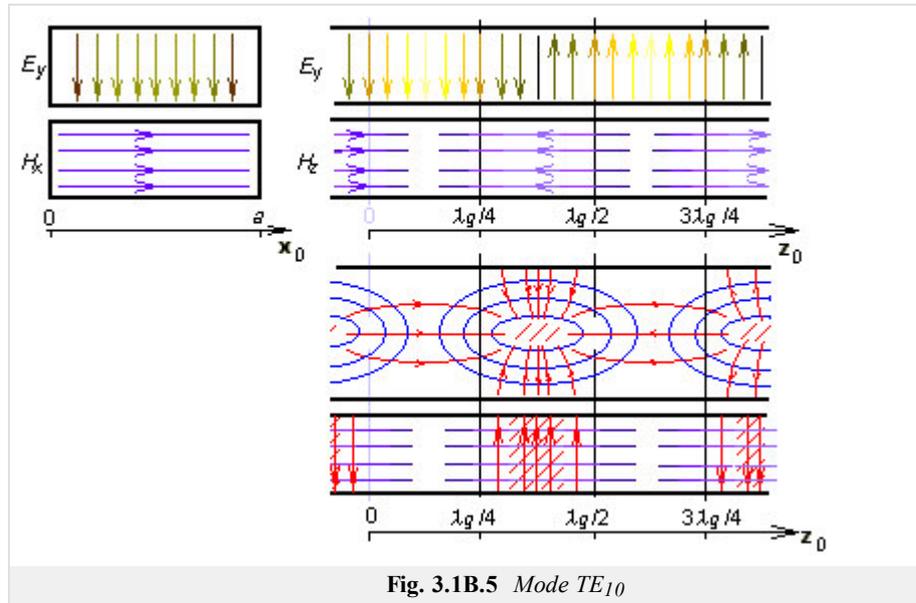
Fig. 3.1B.5 *Mode TE₁₀*

Snijden we de golfgeleider af door een longitudinale wand die loodrecht op de lange zijde staat, dan kunnen we een harmonische voortgang van de transversale component van de elektrische intensiteit E_y *observeren*. Maximale intensiteit in de figuur bevindt zich op de coördinaten $z = \lambda_g/4$ en $z = 3\lambda_g/4$ (dit bij omgekeerde fase). In $z = 0$ en $z = \lambda_g/2$ is de elektrische intensiteit nul. In de punten met maximale E_y, is de longitudinale component van de magnetische intensiteit H_z gelijk aan nul en de transversale component van de magnetische intensiteit H_x maximaal. In de transversale snijding in $z = \lambda_g/4$ is E_y maximaal in het centrum van de golfgeleider en nul aan de wonden. De transversale component van de magnetische intensiteit H_x is constant in $z = \lambda_g/4$.

Als de snijdingswand loodrecht op de korte zijde staat zijn de magnetische intensiteiten elipsvormig (zoals een draad zich voortbeweegt op het ritme van een golf). In het geval van de golfgeleider waar een stroom vloeit van de onderste wand naar de bovenliggende wand waarop het vacuüm een magnetisch veld produceert. Als de stroom de wand bereikt vloeit deze in de vorm van een leidende stroom weg van de wand.

Uiteindelijk kunnen we stellen dat het transversale elektromagnetische veld zo verdeeld is dat de grensvoorwaarden gelijk zijn aan een perfecte elektrische geleiding van de behuizing. Deze stelling is weergegeven door het matlab programma voor de analyse van het elektromagnetische veld in de rechthoekige golfgeleider door de finite-element methode. Een handleiding van het programa kun je vinden in laag C. Een introductie van de finite-element methode en een beschrijving van de software implementaties zijn gegeven in laag D.

3.1 Waveguides

Matlab program

A rectangular waveguide R100 (dimensions 22.86 mm x 10.16 mm) is analyzed by the Matlab program `waveguide.m`. This program (plus all the necessary additional m-files) is available here. We run Matlab and set the path to the folder **Waveguide**. Writing `waveguide` to the command window of Matlab, the program is started. Handling program can be subdivided to following steps:

1. Number of rectangular bi-elements, sampling the waveguide cross section in the direction x, is written to the input line number of elements in x, and number of bi-elements in y is written to the input line number of elements in y. If transversally electric modes are going to be computed, the button $\boxed{\text{Modes TE}}$ has to be pressed. If transversally magnetic modes are going to be computed, the button $\boxed{\text{Modes TM}}$ has to be pressed.

2. If the elected analysis is finished, a dialog containing the input line **Mode number** appears. The lowest mode is denoted as "1", higher modes are numbered by rising integer index (modes are sorted according to the value of the critical frequency). Pressing the button $\boxed{\text{Draw}}$, three graphical windows are opened: the first one displays distribution of the longitudinal component, the second one and the third one display transversal components of field intensities.

3. Program is finished by pressing the button $\boxed{\text{End}}$.

Program `waveguide` analyzes EM field distribution in a waveguide by the finite-element method. A brief description of this method is given in the layer D.

3.1 Waveguides

Developing Matlab

Finite-element method is a general numerical method for the solution of partial differential equations. Since Maxwell equations, which describe EM wave propagation in a waveguide, belong to the family of partial differential equations, finite-element method can be applied to the analysis of guiding structures.

The basic algorithm of computing field distribution in a waveguide can be expressed by following steps:

Fig. 3.1D.1 *Two-dimensional finite-element mesh. Blue numbers: local nodes. Red numbers: global nodes*

1. Waveguide cross-section is divided into finite elements. If longitudinally homogeneous structure is analyzed, the cross-section is broken to small triangles. Triangles are called finite elements, vertexes are called nodes.

2. Distribution of the longitudinal component of electric-field intensity (modes TM) and magnetic-field one (modes TE) is formally approximated over each finite element by linear, quadratic or cubic function. Formal approximation is based on unknown nodal intensities and known basis functions. Each basis function is unitary in a single node and is zero in the rest of nodes. Approximation is a single equation for N unknown nodal values.

3. Formal approximation is substituted to the solved equation. Due to the proximity, the solved equation is not met exactly. This fact is represented by adding a residual function to the solved equation. Decreasing residuum, accuracy of the approximation is increased.

4. Residuum is minimized by the method of weighted residua. The method consists in sequential multiplication of the residuum by proper weighting functions and in integrating products over the whole cross-section of the analyzed waveguide. If basis functions corresponding to the nodes of unknown nodal value are used as weighting functions, we get N equations for N unknown approximation coefficients. The described election of weighting functions is called Galerkin method.

5. We solve a matrix equation produced by the Galerkin method. That way, values of unknown nodal intensities are obtained. In the case of the waveguide, the matrix equation is of the form of generalized eigenvalue problem. As a result, we get a vector of eigenvalues (vector of squared critical wave numbers of respective modes) and a matrix of respective eigenvectors (nodal values of a longitudinal component of field intensity of respective modes).

6. Substituting nodal values to the formal approximation, an approximation function of the longitudinal component of field intensity over the cross-section of the waveguide is obtained. On the basis of the longitudinal component distribution, transversal components of electric field intensity and magnetic one can be computed.
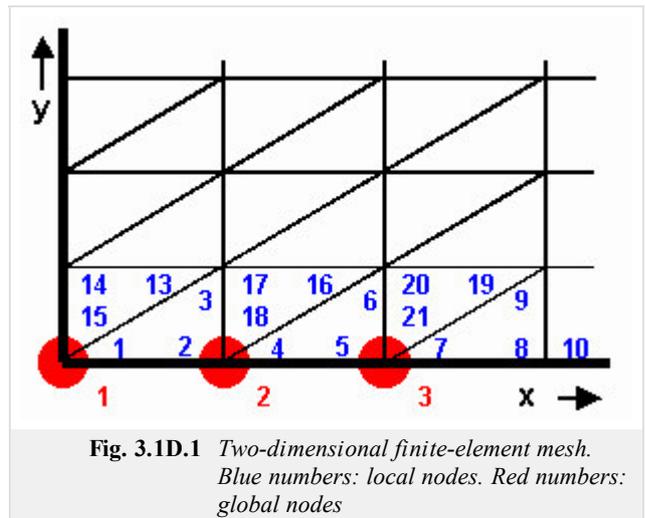
In practical computations by the finite-element method, the approach is slightly different. In references, already computed coefficient matrixes for normalized finite element are at our disposal. We adopt these matrixes, rearrange them to used finite elements and associate them together with finite elements to the net. The procedure can be expressed by the following steps:

1. Cross section of the waveguide is divided into finite elements, which is identical with the previous approach.

Fig. 3.1D.2 *Two-dimensional finite element for linear approximation*
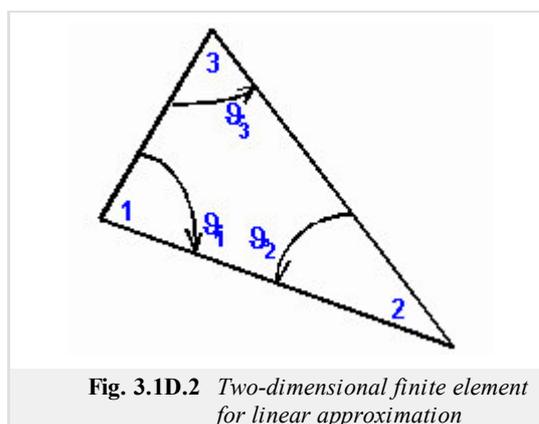
2. We find relations for computing coefficients of finite elements. Substituting surface of eth finite element $A^{(e)}$ and angles at vertexes of

eth element (fig. **2)** to these relations, we obtain coefficients for our finite elements. For linear approximation, we get:

$$\mathbf{S}^{(e)} = \sum_{n=1}^{3} \mathbf{Q}_n \cot g\,\vartheta_n^{(e)}, \quad \mathbf{T}^{(e)} = \frac{A^{(e)}}{12}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

$$\mathbf{Q}_1 = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & +1 & -1 \\ 0 & -1 & +1 \end{bmatrix}, \quad \mathbf{Q}_2 = \frac{1}{2}\begin{bmatrix} +1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & +1 \end{bmatrix}, \quad \mathbf{Q}_3 = \frac{1}{2}\begin{bmatrix} +1 & -1 & 0 \\ -1 & +1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(3.1D.1)

3. We build up diagonal matrices for isolated finite elements, i.e. for elements, which are not associated into the mesh (symbols 0 denote zero matrices of the size 3 x 3)

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}^{(1)} & & & \\ & \mathbf{S}^{(2)} & & \\ & & \mathbf{S}^{(3)} & \\ & & & \ldots \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{T}^{(1)} & & & \\ & \mathbf{T}^{(2)} & & \\ & & \mathbf{T}^{(3)} & \\ & & & \ldots \end{bmatrix}$$

(3.1D.2)

4. We build up an association matrix, which describes mutual relations among local nodes (blue number is fig. **1**) and global ones (red numbers in fig. **1**). Columns of the matrix correspond to global nodes, rows of the matrix to local ones. If a local node is asked to be associated with a global one, we put 1 to the row of the local node and to the column of global one. In the opposite case, the matrix element equals to zero. For nodes of fig. **1**, the association matrix is of the following form:

Fig. 3.1D.3 *Association matrix*

5. Independent isolated finite elements are associated to the finite-element mesh. From the mathematical point of view, the association is described by the following matrix operation:

$$\mathbf{S}_c = \mathbf{C}^{\mathrm{T}}\mathbf{S}\,\mathbf{C}, \quad \mathbf{T}_c = \mathbf{C}^{\mathrm{T}}\mathbf{T}\,\mathbf{C}$$

(3.1D.3)

6. Solving the final matrix equation (the same equation as obtained by the above-described approach)

$$\mathbf{S}\,\mathbf{E} + k^2\,\mathbf{T}\,\mathbf{E} = 0$$

(3.1D.4)

we get critical wave numbers of modes k and nodal values of the longitudinal component of electric-field intensity vector **E** of those

modes. Substituting nodal values to the formal approximation, a global approximation function of the electric-field intensity component is obtained for every point of the cross-section of the analyzed waveguide.

Our matlab program is based on the presented algorithm. A simplified list of the program follows. Since the source code is in detail commented, no further description is given.

```matlab
function kn = linearTE( Nx, Ny);

iNx = Nx + 1;
iNy = Ny + 1;

a = 22.86e-3; % width of waveguide
b = 10.16e-3; % height of waveguide

dx = ones(1,Nx) * (a/Nx); % vectors of dimensions of FE
dy = ones(1,Ny) * (b/Ny);

Q1 = [ 0 0 0; 0 1 -1; 0 -1 1] / 2;
Q3 = [ 1 -1 0; -1 1 0; 0 0 0] / 2;

Te = [ 2 1 1; 1 2 1; 1 1 2] /12;

N = 2 * Nx * Ny; % total number of FE

St = sparse( 3*N, 3*N); % matrices of isolated FE
Tt = sparse( 3*N, 3*N);

n = 0;

for ny=1:Ny
    for nx=1:Nx
        n = n + 1;
        lw = 3*n-2;
        hg = 3*n;
        St(lw:hg,lw:hg) = Q1 * dx(nx)/dy(ny) + Q3 * dy(ny)/dx(nx);
        Tt(lw:hg,lw:hg) = Te * dx(nx)*dy(ny)/2;
        St(lw+3*Nx:hg+3*Nx,lw+3*Nx:hg+3*Nx) = St(lw:hg,lw:hg);
        Tt(lw+3*Nx:hg+3*Nx,lw+3*Nx:hg+3*Nx) = Tt(lw:hg,lw:hg);
    end
    n = n + Nx;
end

C = get_c1( Nx, Ny, N)

S = C'*St*C; % associating isolated FE
T = C'*Tt*C;

clear C St Tt

[H,K] = eig( full(S), full(T)); % solving eigenvalue problem

kn = sqrt( diag( K)); % vector of critical wave numbers
```

3.1 Waveguides

Java applet

3.1 Waveguides

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**10** of **10**:

Question 1

The wave propagation inside a waveguide can be described by …

Possible answers for question 1:

- ○ … beams propagating in parallel to the walls of the structure.
- ○ … successive reflections of the propagating waves from the walls of the guiding structure.
- ○ … the diffusion of the electromagnetic energy deeper and deeper to the guiding structure.
- ◉ I won't answer this question

Question 2

Waveguides are used at higher frequencies because …

Possible answers for question 2:

- ○ … the length of the waveguide has to be related to the wavelength (at lower frequencies, the waveguide is too long).
- ○ … losses in metallic walls are too high at lower frequencies.
- ○ … the cross section of the waveguide has to be related to the wavelength (at lower frequencies, the waveguide cross section is very large).
- ◉ I won't answer this question

Question 3

Electromagnetic energy propagates through the waveguide …

Possible answers for question 3:

- ○ … if its frequency is higher than the critical one.
- ○ … anytime.
- ○ … if its frequency is lower than the critical one.
- ◉ I won't answer this question

Question 4

The propagation constant …

Possible answers for question 4:

- ○ … describes the changes of wave amplitude and phase when propagating in the waveguide.
- ○ … is not used in relation to the waveguides.
- ○ … is used to formulate excitation conditions for the waveguides.
- ◉ I won't answer this question

Question 5

Energy transmission on the waveguide is described by …

- ○ … the phase velocity.
- ○ … the group velocity.
- ○ … velocity of light.
- ◉ I won't answer this question

Question 6

In the waveguide, electromagnetic waves can propagate in different constellations (so called modes). When increasing operation frequency …

Possible answers for question 6:
- ○ … more and more modes appear in the waveguide – higher modes propagate at higher frequencies.
- ○ … a single mode propagates in the waveguide – if a higher mode is excited, a lower mode is suppressed.
- ○ … no wave propagates due to the enormous attenuation.
- ◉ I won't answer this question

Question 7

The dominant mode is the mode …

Possible answers for question 7:
- ○ … of the highest energy.
- ○ … of the strongest field intensity.
- ○ … of the lowest critical frequency.
- ◉ I won't answer this question

Question 8

In the waveguide …

Possible answers for question 8:
- ○ … transversally electromagnetic modes can propagate.
- ○ … transversally electric OR transversally magnetic modes can propagate.
- ○ … transversally electric AND transversally magnetic modes can propagate.
- ◉ I won't answer this question

Question 9

When analyzing the waveguide by finite elements …

Possible answers for question 9:
- ○ … partial differential equations are solved by replacing derivatives by differences.
- ○ … integral equations are solved by a rather complicated way.
- ○ … the error of the approximate solution is minimized by the Galerkin method.
- ◉ I won't answer this question

Question 10

Finite elements provide eigenvalues and eigenvectors:

Possible answers for question 10:
- ○ Eigenvalues are related to the critical wave numbers of modes, eigenvectors are not of any use.
- ○ Eigenvalues are related to the critical wave numbers of modes, and eigenvectors to the field distribution of modes.
- ○ Eigenvectors are related to the critical wave numbers of modes, and eigenvalues to the field distribution of modes.
- ◉ I won't answer this question

see result

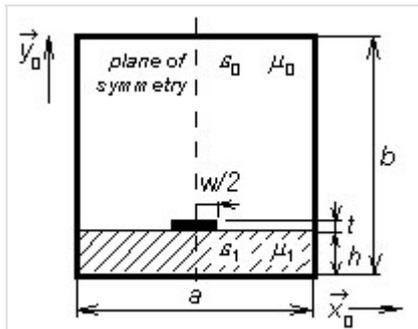# 3.2 Shielded microstrip tranmission lines

Basic theory

Shielded microstrip transmission line is a member of the family of planar microwave transmission lines. Open microstrip line (fig. **3.2A.1a**), shielded slot-line (fig. **3.2A.1b**) shielded fin-line (fig. **3.2A.1c**) or coplanar waveguide (fig. **3.2A.1d**) belong to the most common planar transmission lines.

Using segments of planar transmission lines, microwave circuits can be composed. Planar conductors can serve for connecting circuit components or can be used for creating passive circuit elements (capacitors and inductors, especially). Moreover, planar circuits can be relatively simply completed by active elements (transistors and diodes, e.g.). Therefore, the planar technology is of wide use these days and analysis of planar structures is of rising importance.

In this paragraph, we deal with the computation of the electromagnetic field distribution in shielded microstrip transmission line (fig. **3.2A.2**), which parameters are assumed to be constant in the longitudinal direction. Then, only a two-dimensional structure has to be analyzed (cross-section of the transmission line), which simplifies computation **[20]**.
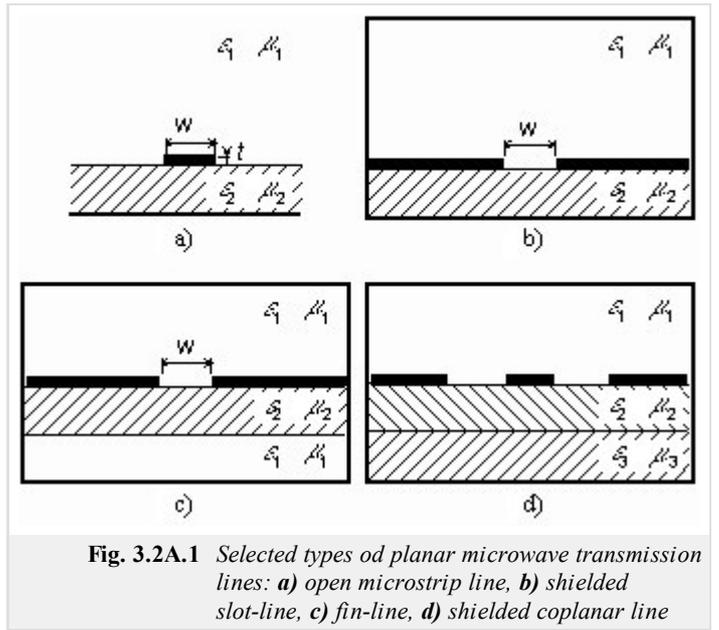
Fig. 3.2A.1 *Selected types od planar microwave transmission lines: a) open microstrip line, b) shielded slot-line, c) fin-line, d) shielded coplanar line*

Fig. 3.2A.2 *Shielded microstrip transmission line (longitudinally homogenous)*

Even if the above-described simplification is done, field distribution in the shielded microstrip transmission line cannot be computed analytically. Therefore, numerical methods or approximate ones have to be exploited. All the methods can be divided into two groups, to quasi-static methods and to full-wave methods **[20]**.

Quasi-static methods are based on the assumption that the dominant mode of the wave, which propagates along the transmission line, can be approximated (with good accuracy) by the transversal electromagnetic wave (TEM). Unfortunately, this assumption is valid on low microwave frequencies (typically by 5 GHz). If frequency is increased, then the value of longitudinal components of electromagnetic field rises, and hence, it cannot be neglected.

Working on higher microwave frequencies, full-wave methods have to be exploited for the analysis. Full-wave methods are based on the direct solution of Maxwell equations. Since the transmission line is going to be analyzed in the harmonic steady state, Maxwell equations are of the following form:

$$\nabla \times \mathbf{E} = -j\omega\mu_0\mu_r\,\mathbf{H}, \tag{3.2A.1a}$$

$$\nabla \times \mathbf{H} = +j\omega\varepsilon_0\varepsilon_r\,\mathbf{E} + \sigma\,\mathbf{E} + \mathbf{J}_S, \tag{3.2A.1b}$$

$$\nabla \cdot \left(\varepsilon_0\varepsilon_r\,\mathbf{E}\right) = \rho, \tag{3.2A.1c}$$

$$\nabla \cdot \left(\mu_0\mu_r\,\mathbf{H}\right) = 0. \tag{3.2A.1d}$$

Here, **E** denotes electric-field intensity vector, **H** is magnetic-field intensity vector, operator *nabla* is in Cartesian coordinate system of the form

,

$\mathbf{J}_S$ denotes density of currents, which are *imposed* to the analyzed system by sources, ω is angular frequency, ε_0 and μ_0 are permittivity and permeability of vacuum, ε_r and μ_r denote relative permittivity and permeability of dielectrics inside the analyzed structure, ρ is volume charge density in dielectrics and σ denotes electric conductivity of dielectrics.

Assume that sources of electromagnetic field are in a large distance from the area where computations are performed (then, imposed currents $\mathbf{J}_S$ are zero in this area). Next, zero charge density is supposed in the structure and the media are expected to be linear and isotropic (permittivity and permeability are scalar quantities, which do not depend on the value of respective field intensities). Further, dielectrics are assumed to be lossy (represented by electric conductivity σ) and all metallic parts (shielding waveguide, microstrip) are expected to be perfect electric conductors.

All the assumptions are substituted to (3.2A.1). Then, both sides of (3.2A.1a) are multiplied by the operator *nabla* from left (in vector way). That way, curl of the magnetic-field intensity vector is obtained on the right-hand side of (3.2A.1a), which is replaced by the right-hand side of

(3.2A1b). Finally, the following equation is obtained:

$$\nabla \times (\nabla \times \mathbf{E}) = -j\omega\mu_0\mu_r(\sigma + j\omega\varepsilon_0\varepsilon_r)\mathbf{E}. \qquad (3.2A.2)$$

If the square of wave number in vacuum is denoted as

$$k_0^2 = -j\omega\mu_0\, j\omega\varepsilon_0 \qquad (3.2A.3)$$

and complex relative permittivity of dielectrics as

$$\tilde{\varepsilon}_r = \frac{\sigma}{j\omega\varepsilon_0} + \varepsilon_r, \qquad (3.2A.4)$$

eqn. (3.2A.2) can be rewritten to

$$\nabla \times (\nabla \times \mathbf{E}) - \tilde{\varepsilon}_r\mu_r k_0^2 E = 0. \qquad (3.2A.5)$$

Analyzed microstrip transmission line is situated to the Cartesian coordinate system (coordinates x and y in transversal directions, coordinate z longitudinal). Hence, electromagnetic wave propagates along the axis z (along microstrip) and electric-field intensity vector depends on the longitudinal coordinate according to

$$\mathbf{E}(x,\, y,\, z) = \mathbf{E}(x,\, y)\exp(-\gamma z), \qquad (3.2A.6)$$

where γ is propagation constant

$$\gamma = \beta + j\alpha; \qquad (3.2A.7)$$

β is attenuation constant and α denotes phase constant.

Mathematical description of electromagnetic wave propagating in the direction z (3.2A.6), is substituted to vector equation (3.2A.5). Then, all the partial derivatives according to z can be evaluated (respective terms are multiplied by the propagation constant -γ). Rewriting all vectors in (3.2A.5) as a sum of the transversal component and the longitudinal one

$$\mathbf{E} = \mathbf{E}_t + z_0\, \mathbf{E}_z, \qquad (3.2A.8a)$$

$$\nabla = \nabla_t + \mathbf{z}_0\frac{\partial}{\partial z} = \nabla_t - \gamma\,\mathbf{z}_0, \qquad (3.2A.8b)$$

we get

$$\nabla_t \times (\nabla_t \times \mathbf{E}_t) - \gamma(\nabla_t E_z + \gamma\,\mathbf{E}_t) = k_0^2\mu_r\tilde{\varepsilon}_r\mathbf{E}_t, \qquad (3.2A.9a)$$

$$\nabla_t \times \big[(\nabla_t E_z + \gamma\,\mathbf{E}_t) \times \mathbf{z}_0\big] = k_0^2\mu_r\tilde{\varepsilon}_r E_z\,\mathbf{z}_0. \qquad (3.2A.9b)$$

Where as (3.2A.9a) is vector equation of transversal components, (3.2A.9b) is scalar equation of a longitudinal component.

The set of differential equations (3.2A.9) has to be completed by boundary conditions, which have to be met by the solution of (3.2A.9)

$$\left.\begin{array}{r} \mathbf{n}_0 \times \mathbf{E}_t = 0 \\ E_z = 0 \end{array}\right\}na\,\Gamma_1, \qquad (3.2A.10a)$$

$$\left.\begin{array}{r} \big[\nabla_t E_z + \gamma\,\mathbf{E}_t\big] \cdot \mathbf{n}_0 = 0 \\ \nabla_t \times \mathbf{E}_t = 0 \end{array}\right\}na\,\Gamma_2. \qquad (3.2A.10b)$$

Eqn. (3.2A.10a) describes the fact that components of electric-field intensity vector, which are tangential to the perfectly electrical-conductive surfaces Γ_1, have to be zero on those surfaces. Eqn. (3.2A.10b) expresses the fact that variation of components of electric-field intensity vector in the normal direction with respect to perfectly electric-conductive surface Γ_2 has to be zero on this surface.

Eqn. (3.2A.9) completed by boundary conditions (3.2A.10) are the initial relations to the full-wave analysis of microstrip transmission line. We have to keep in mind, that eqns. (3.2A.9), (3.2A.10) cover the first Maxwell equation and the second one only, and therefore, the solution has to be checked to meet the third Maxwell equation and the fourth one. In the opposite case, the solution would be physically non-existing (*spurious one*).

In the following paragraphs, we describe the way of implementing the full-wave analysis of a shielded microstrip line, which is based on (3.2A.9) and (3.2A.10), by finite-element method.

First, we briefly comment physical phenomena in the investigated structure because success of the analysis is conditioned by their correct modeling.

First, consider a hollow rectangular waveguide (inside, vacuum is assumed). In such waveguide, a transversal electric wave (longitudinal component of magnetic intensity is present) or a transversal magnetic one (longitudinal component of electric intensity is present) can propagate.

If a dielectric substrate is inserted into the waveguide (filling becomes inhomogeneous), longitudinal components of both the electric-field intensity and the magnetic-field one appear in the structure at the same time. In the opposite case, boundary conditions could not be met at the interface between dielectric and vacuum.

Wave propagation in a waveguide of inhomogeneous filling can be described by hybrid modes LSM (*Longitudinal Section Magnetic*) and LSE (*Longitudinal Section Electric*). Moreover, in inhomogeneous filling is completed by a metallic microstrip, transversal and longitudinal currents are induced there, which mutually couple LSE and LSM modes.

Numerical analysis of wave propagation in a metallic waveguide (vacuum inside) is relatively simple because homogeneous wave equation is needed to be solved for the longitudinal component of electric-field intensity (modes TM) or magnetic-field one (modes TE). If microstrip transmission line is going to be analyzed, attention has to be paid to two potential sources of spurious solutions.

First, production of spurious solutions can be caused by incorrect modeling of electromagnetic conditions on the boundary between the substrate and vacuum. Second, incorrectly modeled influence of metallic strips can cause appearance spurious solutions. Exploiting hybrid finite elements can eliminate both the causes.

Analyzing a shielded microstrip line by hybrid finite elements, all the components of electric-field intensity or magnetic-field one have to be included into the computations. Eqn. (3.2A.9) is the initial relation of the analysis.

Matter of hybrid finite elements consists in modeling a longitudinal intensity component using nodal approximation, and transversal intensity components using an approximation, which is based edge vectors.

Turn our attention to the general solution of (3.2A.9) under conditions (3.2A.10) by finite-element method. At every step of the general approach, we discuss the operation of the method in our situation.

First, an analyzed structure is subdivided into sub-spaces (finite elements), which do not overlap and which cover all the points of the structure. In the space of a finite element, parameters of the analyzed structure (permittivity, permeability, conductivity) have to be constant. There are no restrictions to size and shape of finite elements. Finite-element mesh can be denser in areas where details of the solution are of our interest, and can be relatively sparse in areas where details are out of our interest. Finite elements can be curvilinear, and therefore, an arbitrary geometry can be modeled.

Analyzing longitudinally homogeneous microstrip transmission line, two-dimensional problem is solved. Therefore, we turn our attention to two-dimensional finite elements. Leaving piecewise constant approximation, a linear function is the simplest approximation function. In two dimensions, the linear function is a plane over a finite element (the computed quantity can be imagined to be drawn above the element).
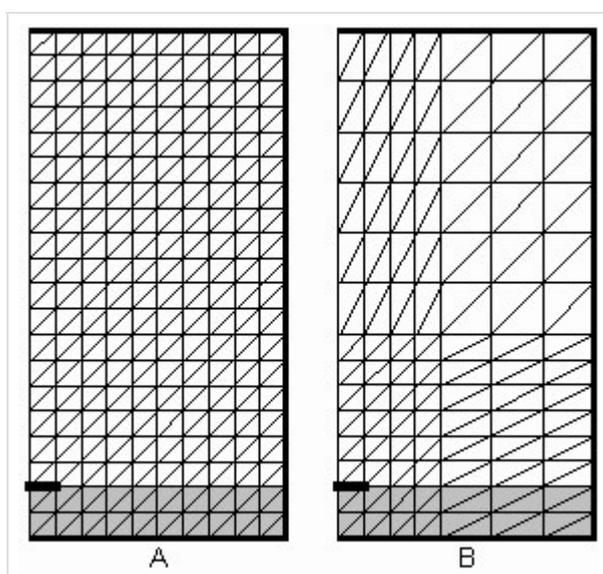
Fig. 3.2A.3 *Mesh examples of rectangular bi-elements for the analysis of a shielded microstrip transmission line*

If three points uniquely determine the approximation plane, the analyzed structure is subdivided to triangular finite elements. Coordinates of vertexes of triangular finite elements are independent variables of the two-dimensional linear approximation. Samples of sought field distribution are dependent functional values. Interlaying the approximation plane through the dependent functional values in triangle vertexes, a unique approximation plane is obtained.

The analyzed structure has to be divided to finite elements very carefully because both the final error and CPU-time demands of the analysis strongly depend on the mesh geometry.

In fig. **3.2A.3**, two examples of triangular finite-element meshes for the analysis of shielded microstrip line are depicted. For simplicity, two triangular finite elements create a rectangular bi-element together, which is the basic building block of our mesh.

Mesh A is homogeneous (consists of identical finite elements). Since electromagnetic field can be expected to quickly decrease with the distance from the microstrip, distant areas can be covered by a sparser mesh B. As shown later, the mesh B negligibly decreases accuracy of the analysis, but significantly reduces CPU-time demands (lower number of finite elements corresponds to lower number of mathematical operations).

In the second step of the solution, an approximation of a sought function over each finite element is expressed in a formal way. Usually, the unknown solution is approximated by a linear combination of elected approximation functions and unknown approximation coefficients. A linear function is the simplest approximation. As an example, we consider a general plane, which is uniquely determined by functional values of the approximated function in three points of a finite element.

This general plane can be composed of three sub-planes (a linear combination), each of which is unitary in a single points of a finite element and

which is zero in other two points of a finite element. Unknown approximation coefficients, which appear in the linear combination at partial planes, play the role of spatial samples of the spatial distribution of computed quantity in points, where partial functions are of unitary value (other functions are zero in this point, and therefore, they do not participate in sampling). For the longitudinal component of electric-field intensity, we get therefore

$$E_z^{(n)}(x, y) \approx \sum_{m=1}^{M_n} c_m N_m^{(n)}(x, y), \qquad (3.2A.11)$$

where $E_z^{(n)}$ is an approximation of a sought function over n^{th} finite element, c_m are unknown approximation coefficients and $N_m^{(n)}$ are elected partial approximation functions on n^{th} finite element, M_n denotes the number of partial approximation functions exploited for composition of the total approximation of sought function over n^{th} finite element.

Collecting approximations over all N finite elements, a global approximation over the whole analyzed structure is obtained

$$\tilde{E}_z(x, y) \approx \sum_{n=1}^{N} E_z^{(n)}(x, y). \qquad (3.2A.12)$$

Eqn. (3.2A.12) is a single equation for M unknown approximation coefficients c_m. If these coefficients are found, approximation of the sought quantity is obtained.

More information about above-described approximation can be found in the layer B.

In the third step, the formal approximation of the solution $\tilde{E_z}$ is substituted to the solved partial differential equation. Since the approximation differs from the exact solution, the initial equation is not met perfectly. This fact is respected by introducing a residual function, which equals to the difference between the exact solution and the approximation

$$R(x, y) = \tilde{E}(x, y) - E(x, y). \qquad (3.2A.13)$$

The approximation is as accurate as low values of the residual function are. Therefore, we are going to minimize the residual function over the whole analyzed structure. The residual function is minimized exploiting the method of weighted residuals.

The method of weighted residuals consists in multiplying the residual function $R(x, y)$ by a weighting function $W(x, y)$. The product is integrated over the whole analyzed space S and the result is set to equal zero

$$\int_S R(x, y) W(x, y) dS = 0, \qquad (3.2A.14)$$

where dS = dx dy.

If as many properly elected weighting functions W_n are used as many unknown approximation coefficients are computed, a set of linear equations (as many equations as unknown approximation coefficients) is obtained. Solving this set of equations, unknown approximation coefficients are obtained. If partial approximation functions of nodes, where nodal quantity is unknown, are elected as weighting functions the method of weighted residuals comes to Galerkin method.

For n^{th} finite element, the final matrix equation is of the form

where

$$\mathbf{T}_t^{(n)} \mathbf{E}_t^{(n)} = \sum_{i,j} \left\{ e_{t,ij}^{(n)} \iint_{S^{(n)}} \left[\mathbf{N}_{t,rs}^{(n)} \cdot \mathbf{N}_{t,ij}^{(n)} \right] dS \right\}, \qquad (3.2A.15a)$$

$$\mathbf{G}^{(n)} \mathbf{E}_z^{(n)} = \sum_{m} \left\{ e_{z,m}^{(n)} \iint_{S^{(n)}} \left[\mathbf{N}_{t,rs}^{(n)} \cdot \left(\nabla_t N_{z,m}^{(n)} \right) \right] dS \right\}, \qquad (3.2A.15b)$$

$$\mathbf{S}_z^{(n)} \mathbf{E}_z^{(n)} = \sum_{m} \left\{ e_{z,m}^{(n)} \iint_{S^{(n)}} \left[\left(\nabla_t N_{z,q}^{(n)} \right) \cdot \left(\nabla_t N_{z,m}^{(n)} \right) \right] dS \right\}, \qquad (3.2A.15c)$$

$$\mathbf{T}_z^{(n)} \mathbf{E}_z^{(n)} = \sum_{m} \left\{ e_{z,m}^{(n)} \iint_{S^{(n)}} \left[N_{z,q}^{(n)} N_{z,m}^{(n)} \right] dS \right\}, \qquad (3.2A.15d)$$

$$\mathbf{S}_t^{(n)} \mathbf{E}_t^{(n)} = \sum_{i,j} \left\{ e_{t,ij}^{(n)} \iint_{S^{(n)}} \left[\left(\nabla_t \times \mathbf{N}_{t,rs}^{(n)} \right) \cdot \left(\nabla_t \times \mathbf{N}_{t,ij}^{(n)} \right) \right] dS \right\}. \qquad (3.2A.15e)$$

In the above relations, $\mathbf{E}_t^{(n)}$ is unknown column vector of edge approximation coefficients (approximation of transversal components of electric-field intensity) over n^{th} finite element and $\mathbf{E}_z^{(n)}$ denotes column vector of three unknown nodal approximation coefficients (approximation of longitudinal component of electric-field intensity) over n^{th} finite element. Next, γ denotes complex propagation constant, k_0 is wave number in vacuum, $\mu_r^{(n)}$ is relative permeability of nth finite element and $\tilde{\varepsilon}_r^{(n)}$ is complex relative permittivity of the same element. Symbol dS denotes an elementary facet for the integration over n^{th} finite element and symbol $S^{(n)}$ give the total surface of nth finite element. Summation including index m symbolizes addition over all nodes of a finite element (i.e., $m = 0, 1, 2$) a summation including indexes i, j symbolized addition over all edged of an element (i.e., $i, j = 0\text{-}1, 1\text{-}2, 2\text{-}0$). Symbols $e_{t,ij}^{(n)}$ are edge approximation coefficients, symbols $e_{z,m}^{(n)}$ are nodal approximation coefficients.

Matrices $\mathbf{T}_t^{(n)}$, $\mathbf{G}^{(n)}$, $\mathbf{S}_z^{(n)}$, $\mathbf{T}_z^{(n)}$ and $\mathbf{S}_t^{(n)}$ matrices of coefficients of nth finite element of the size 3 x 3. Elements of the above matrices are computed by integration of the product of partial approximation functions and weighting functions (or their derivatives) over n^{th} finite element. Matrices can be evaluated using the following relations

$$\mathbf{S}_t^{(n)} = \frac{1}{A^{(n)}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \tag{3.2A.16a}$$

$$\mathbf{T}_t^{(n)} = \frac{1}{12} \sum_{i=0}^{2} \mathbf{Q}_i \, cotg\left[\theta_i^{(n)}\right], \tag{3.2A.16b}$$

$$\mathbf{G}^{(n)} = \frac{1}{6} \sum_{i=0}^{2} \mathbf{C}_i \, cotg\left[\theta_i^{(n)}\right], \tag{3.2A.16c}$$

$$\mathbf{S}_z^{(n)} = \frac{1}{2} \sum_{i=0}^{2} \mathbf{D}_i \, cotg\left[\theta_i^{(n)}\right], \tag{3.2A.16d}$$

$$\mathbf{T}_z^{(n)} = \frac{A^{(n)}}{12} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \tag{3.2A.16e}$$

where

$A^{(n)}$ is surface of nth finite element and $\theta_i^{(n)}$ is angle at i^{th} vertex of nth finite element. Given relations are valid for the following organization of nodes and edges

$$\mathbf{E}^{(n)} = \begin{bmatrix} e_{z,0}^{(n)} & e_{z,1}^{(n)} & e_{z,2}^{(n)} & e_{t,12}^{(n)} & e_{t,20}^{(n)} & e_{t,01}^{(n)} \end{bmatrix}^T. \tag{3.2A.16f}$$

If the matrix equation is solved out for the vector of unknown approximation coefficients, then solution of the problem is obtained. Substituting approximation coefficients to the formal approximation, a *real* approximation of a sought function in each point of nth finite element is obtained. Joining approximations over all finite elements, the global solution is found.

In the layer C, the reader can find a matlab program for the finite-element analysis of a shielded microstrip transmission line. Prom the programmer's point of view, the program is described in the layer D.

3.2 Shielded microstrip tranmission lines

Advanced theory

In this paragraph, we turn our attention to the computational details of the analysis of electromagnetic field distribution in a shielded microstrip transmission line (fig. **3.2B.1**) assuming constant parameters along the longitudinal axis. Then, only a two-dimensional structure has to be analyzed (cross-section of the transmission line), which simplifies computations [20]. For the analysis, a full-wave method, a finite-element method is exploited.

Fig. 3.2B.1 *Shielded microstrip transmission line (longitudinally homogenous)*

As already described in layer A, the analysis is based on Maxwell equations in differential form. Sources of electromagnetic waves are assumed to be in long distance from the region of analysis (then, imposed currents $\mathbf{J}_s$ are zero in this region). Charge density r in dielectrics, which surround the microstrip, is supposed to be zero. Dielectrics is expected to be isotropic and linear (permittivity and permeability are scalar quantities, which values do not depend on the value of respective intensity) and to exhibit electric losses (represented by electric conductivity s). All the metallic parts (shielding waveguide, microstrip) are assumed to be perfect electric conductors.

The analyzed microstrip transmission line is placed to Cartesian coordinate system (coordinates x and y in transversal directions, coordinate z in longitudinal one). Then, the wave can be said to propagate in the longitudinal direction z (along the microstrip), and the electric-field intensity vector depends on the longitudinal coordinate by the following way:

$$\mathbf{E}(x,\,y,\,z) = \mathbf{E}(x,\,y)\exp(-\gamma z). \qquad (\,3.2B.1\,)$$

Here, γ is the propagation constant.

Expressing all vectors as a sum of the transversal vector (index t) and the longitudinal one (index z), we get

$$\nabla_t \times (\nabla_t \times \mathbf{E}_t) - \gamma(\nabla_t E_z + \gamma \mathbf{E}_t) = k_0^2 \mu_r \tilde{\varepsilon}_r \mathbf{E}_t, \qquad (\,3.2B.2a\,)$$

$$\nabla_t \times \left[(\nabla_t E_z + \gamma \mathbf{E}_t) \times \mathbf{z}_0\right] = k_0^2 \mu_r \tilde{\varepsilon}_r E_z \, \mathbf{z}_0. \qquad (\,3.2B.2b\,)$$

Whereas (3.2B.2a) is vector equation for transversal components, (3.2B.2b) is scalar equation for longitudinal components. In those relations, ∇_t is a transversal operator *nabla*, $\mathbf{E}_t$ is transversal electric-field intensity vector, γ is propagation constant, E_z denotes longitudinal component of electric-field intensity, k_0 is wave number in vacuum, μ_r denotes relative permeability inside structure, $\tilde{\varepsilon}_r$ is complex relative permittivity inside structure and $\mathbf{z}_0$ denotes unitary vector in the longitudinal direction.
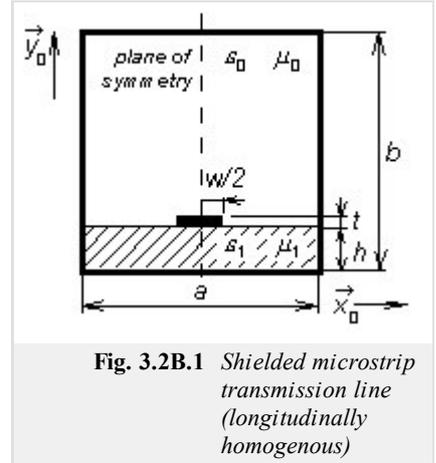
The set of differential equations (3.2B.2) has to be completed by boundary conditions, which have to be met by the solution of (3.2B.2)

$$\left.\begin{array}{r} \mathbf{n}_0 \times \mathbf{E}_t = 0 \\ E_z = 0 \end{array}\right\} na\ \Gamma_1, \qquad (3.2B.3a\,)$$

$$\left.\begin{array}{r} \left[\nabla_t E_z + \gamma \mathbf{E}_t\right] \cdot \mathbf{n}_0 = 0 \\ \nabla_t \times \mathbf{E}_t = 0 \end{array}\right\} na\ \Gamma_2. \qquad (\,3.2B.3b\,)$$

Eqns. (3.2B.2) completed by boundary conditions (3.2B.3) are initial relations for the full-wave analysis of the shielded microstrip transmission line. Unfortunately, the set (3.2B.2), (3.2B.3) includes the first Maxwell equation and the second one only. In order to meet the third Maxwell equation and the fourth one, the analysis has to be based on hybrid finite elements.

If the shielded microstrip transmission line is analyzed exploiting hybrid finite elements, all components of electric-field intensity vector or all components of magnetic-field one have to be included in computations. Eqn. (3.2B.2) stays the initial relation of the analysis.
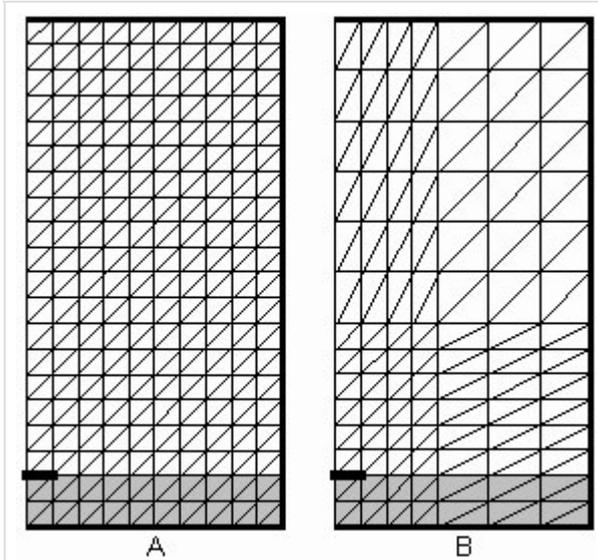
Fig. 3.2B.2 *Mesh examples of rectangular bi-elements for the analysis of a shielded microstrip transmission line*

The matter of hybrid finite elements consists in modeling the longitudinal component of field intensity exploiting nodal approximation and in modeling transversal components using edge vectors.

The first step of the finite-element method consists in dividing the analyzed structure (cross-section of the transmission line) to finite elements (non-overlapping sub-regions, which contain all the points of the analyzed structure). In the area of a finite element, parameters of the analyzed structure (permittivity, permeability, conductivity) have to be constant. There are no restrictions to shape and dimensions of finite elements. Examples of finite-element meshes are depicted in fig. **3.2B.2**.

In the second step of the solution, the distribution of a computed quantity is approximated over each finite element in a formal way. The approximation is expressed as a linear combination of elected partial approximation functions and unknown approximation coefficients.

Analyzing the shielded microstrip transmission line, a formal approximation of a scalar function $E_z = E_z(x, y)$ and a vector one $E_t = E_t(x, y)$ have to be expressed. Let us start with the scalar function.

The global approximation of the scalar function E_z over the whole cross-section of the transmission line is composed of local approximations over single finite elements.

Local approximation of the longitudinal component of electric-field intensity vector over a finite element is expressed as a linear combination of elected partial approximation functions and unknown approximation coefficients. Considering linear approximation, the approximation plane over a finite element is composed of three partial approximation planes. Each partial approximation plane is unitary in a unique vertex of the triangle and is zero in the other two vertexes (see fig. **3.2B.3**). Coefficients c_n at partial functions in the linear combination play the role of spatial samples of the computed function in vertexes of the finite element (fig. **3.2B.3**). Vertexes of the finite element are called nodes and respective functional values are called nodal values.
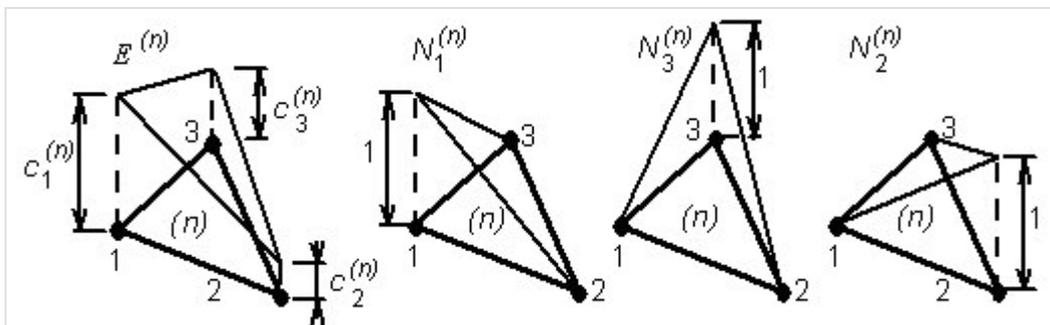
**Fig. 3.2B.3** *Linear approximation of E over finite element. Composed of three shape functions*

Partial approximation functions are called shape functions. All the shape functions, which are unitary in the same node (see fig. **3.2B.4**), compose together a basis function.
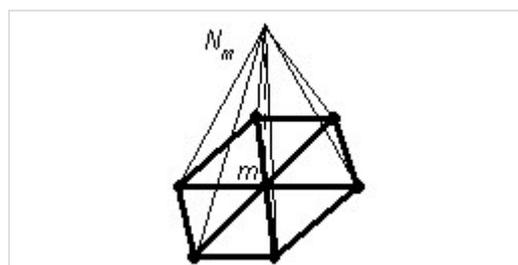
Fig. 3.2B.4 *Linear basis function related to m-th node*

In many cases, approximation functions of higher order are more suitable than the linear function. Although the triangular element has to contain more nodes (6 for quadratic approximation, 10 for cubic one, etc.), the same error comparing to linear approximation is reached even if significantly lower number of finite elements is exploited. Approximation functions of higher order are smoother, and therefore, they represent

better natural quantities.
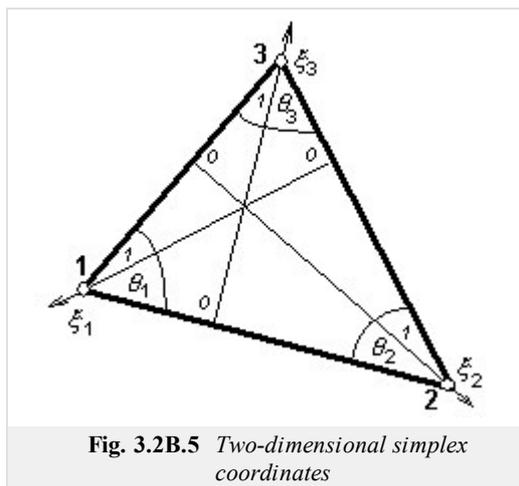
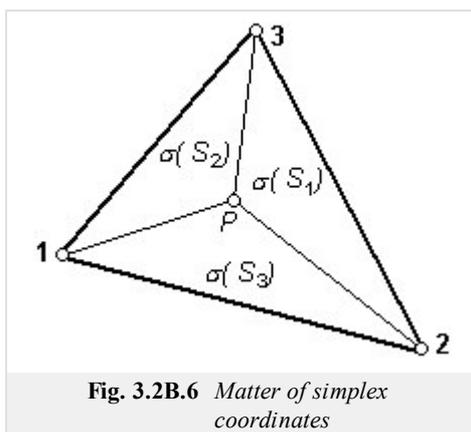**Fig. 3.2B.5** *Two-dimensional simplex coordinates*

Now, we are familiar with shape functions, and therefore, we can find their proper mathematical representation. For this purpose, Lagrange polynomials expressed in simplex coordinates **[21]** are usually used.

What does the term *simplex coordinates* mean? Considering triangular finite elements, simplex coordinate axes are of the direction of heights of the triangle. Simplex coordinates are unitary in the vertex and are zero on the opposite edge. Simplex coordinate do not depend neither on the shape nor on the dimensions of the finite element, and therefore, all the computations are sufficient to be performed once for a single finite element, and the results are recomputed for the other elements only.

Dealing with physical matter of simplex coordinates, a general point P inside a triangular finite element divides its surface to three partial triangles (fig. **3.2B.6**). The ratio of the surface of a triangle, which is positioned in front of the first node, to the surface of the whole finite element equals to the simplex coordinate of P on the first simplex axis

$$\xi_1 = \frac{\sigma(S_1)}{\sigma(S)}.$$ (3.2B.4)

For other simplex coordinate axes, the situation is similar. In eqn. (3.2B.4), $\sigma(S_1)$ denotes surface of the partial triangle, which is positioned in front of the first node, and $\sigma(S)$ is surface of the whole finite element. Obviously, addition of all three simplex coordinates in an arbitrary point is unitary

Fig. 3.2B.6 *Matter of simplex coordinates*

$$\xi_1 + \xi_2 + \xi_3 = 1.$$ (3.2B.5)

As shown in **[21]**, this conclusion can be generalized for an arbitrary dimension and for an arbitrary order of an approximation polynomial.

Now, we turn our attention to Lagrange interpolation polynomials.

Lagrange polynomial of nth can be expressed (using simplex coordinate *x*) as

$$R_m(n, \xi) = \frac{1}{m!} \prod_{k=0}^{m-1} (n\xi - k) \quad m \geq 1 \quad R_0(n, \xi) = 1.$$ (3.2B.6)

Here, *n* is order of the approximation polynomial. Eqn. (3.2B.6) describes the whole family of polynomials: family members differ in the index *m*, which can vary from zero to the order of polynomial *n*.
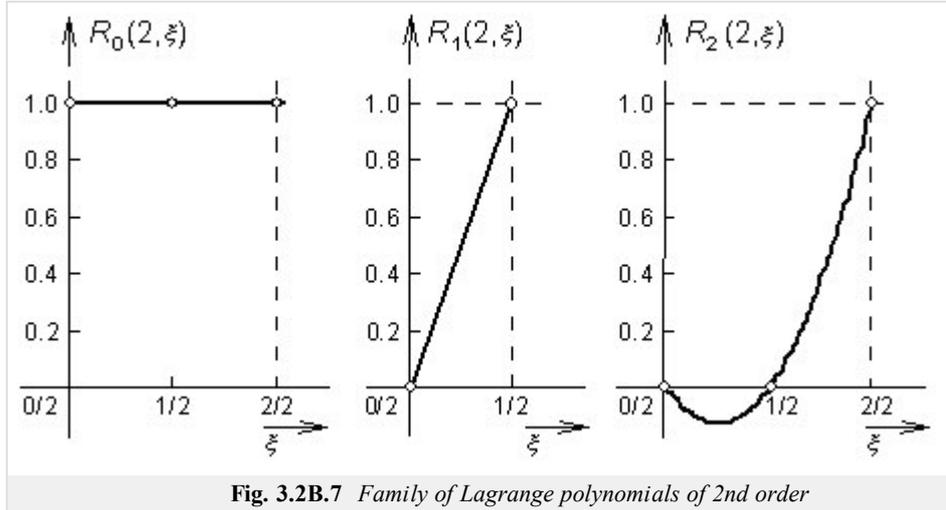
Fig. 3.2B.7 *Family of Lagrange polynomials of 2nd order*

Nulls of polynomials $R_m(n)$ are equidistantly placed on coordinates $\xi = 0$, $1/n$ till $(m-1)/n$, the polynomial is of unitary value in $\xi = m/n$. Hence, $R_m(n)$ is of m equidistantly placed nulls at the left from the coordinate $\xi = m/n$ and of zero nulls at the right.

The above-given statement is illustrated by fig. **3.2B.7**, where all the members of the family of quadratic polynomials $R(2)$ are depicted. Figure demonstrates the above-described equidistant distribution of nulls. The family member of index 0, i.e. $R_0(2)$, does not have any null at the left from the coordinate 0 and is of unitary value at the coordinate 0. The family member of index 1, i.e. $R_1(2)$, is of single null at the coordinate 0 and is unitary at $1/2$. Finally, the family member of index 2, i.e. $R_2(2)$, is of nulls at coordinates 0 and $1/2$ and is unitary at the coordinate 1.
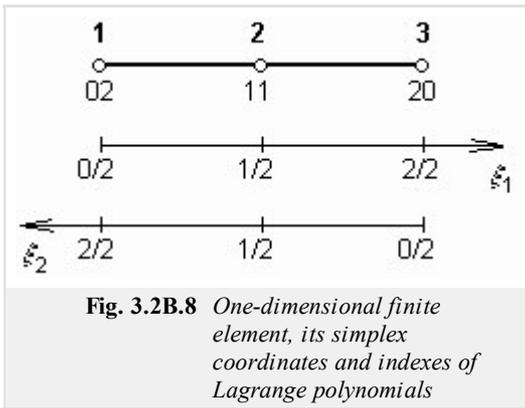
Fig. 3.2B.8 *One-dimensional finite element, its simplex coordinates and indexes of Lagrange polynomials*

Using Lagrange polynomials, we compose quadratic shape functions for one-dimensional finite element. Simplex coordinate ξ_1 is oriented from left to right on this element, the coordinate ξ_2 goes from the right to the left (fig. **3.2B.8**). The shape function related to the node 1 (unitary value in the node 1, zero value in the nodes 2 and 3) is then composed by multiplying Lagrange polynomial of variable ξ_1 and index 0 (constant function of value 1) by Lagrange polynomial of variable ξ_2 and index 2 (since the coordinate ξ_2 is oriented fro the right to the left, the course of the function $R_2(2)$ from fig. **3.2B.7** has to be reverted).

Similarly, shape functions for nodes 2 and 3 can be composed. For the node 2, Lagrange polynomials of variables ξ_1 and ξ_2 and of index 1 are mutually multiplied. For the node 3, Lagrange polynomial of variable ξ_1 and index 2 is multiplied by Lagrange polynomial of variable ξ_2 and index 0. Indexes of Lagrange polynomials, which form shape functions of respective nodes, are written at these nodes (fig. **3.2B.8**) in the form of a fraction; numerator is an index of Lagrange polynomial of the coordinate ξ_1 and denominator is an index of Lagrange polynomial of the coordinate ξ_2. Adding numerator and denominator, order of approximation polynomial n has to be obtained.

In general, the shape function of the node (i, j) of a one-dimensional finite element can be expressed as

$$\alpha_{ij} = R_i\left(n, \xi_1\right)R_j\left(n, \xi_2\right) \quad i + j = n, \tag{3.2B.7a}$$

here n is order of an approximation polynomial, R denotes Lagrange polynomials defined by eqn. (3.2B.5) and ξ are simplex coordinates.

In the next step, we turn our attention to a two-dimensional finite element. The only change, which has to be done, is adding a new simplex coordinate ξ_3 to two existing coordinates ξ_1 and ξ_2. Two multiplicands, which appear in relations for shape functions of a one-dimensional finite element, are completed by the third multiplicand, corresponding to Lagrange polynomial of a new simplex coordinate ξ_3

$$\alpha_{ijk} = R_i\left(n, \xi_1\right)R_j\left(n, \xi_2\right)R_k\left(n, \xi_3\right) \quad i + j + k = n. \tag{3.2B.7b}$$

Here, ξ_1, ξ_2 and ξ_3 denote simplex coordinates of a two-dimensional finite element, n is order of an approximation polynomial and R are Lagrange polynomials.

Substituting to (3.2B.6) a (3.2B.7), we get for linear approximation the following shape functions

$$\alpha_{100} = N_1^{(n)} = \xi_1, \quad \alpha_{010} = N_2^{(n)} = \xi_2, \quad \alpha_{001} = N_3^{(n)} = \xi_3. \tag{3.2B.8}$$

Now, we are familiar with basis functions for the approximation of a scalar function E_z. Therefore, we can turn our attention to the approximation of vector function $\mathbf{E}_t$. The approximation of a vector function formally corresponds to the approximation of a scalar function;

only basis functions are of vector nature

$$\mathbf{E}_t^{(n)} = \mathbf{N}_{t,01}^{(n)} e_{t,01}^{(n)} + \mathbf{N}_{t,12}^{(n)} e_{t,12}^{(n)} + \mathbf{N}_{t,20}^{(n)} e_{t,20}^{(n)} = \sum_{i,j} \mathbf{N}_{t,ij}^{(n)} e_{t,ij}^{(n)}. \tag{3.2B.9}$$
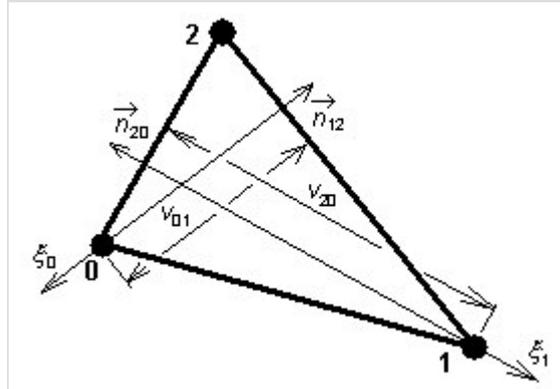
Fig. 3.2B.9 *To the explanation of behavior of vector shape function; for simplicity, superscripts (n) are missed*

In eqn. (3.2B.9) $e_{t,01}^{(n)}$ denotes the edge approximation coefficient for the approximation of the distribution of transversal components of electric-field intensity vector, which is related to the edge 0-1 of nth finite element. Next, $\mathbf{N}_{t,01}^{(n)}$ denotes vector shape function, which is multiplied by the edge coefficient 0-1 in order to evaluate submission of this coefficient to the approximation of the distribution of transversal electric-field intensity vector over n^{th} finite element. Similar situation can be observed at the other two vector shape functions.

The vector shape function can be expressed as

$$\mathbf{N}_{t,ij}^{(n)} = -\xi_i \frac{l_{k,i}^{(n)}}{2A^{(n)}} \mathbf{n}_{k,i}^{(n)} + \xi_j \frac{l_{j,k}^{(n)}}{2A^{(n)}} \mathbf{n}_{j,k}^{(n)}. \tag{3.2B.10}$$

Here, $A^{(n)}$ denotes surface of nth finite element, $l_{k,i}^{(n)}$ is the length of the edge k-i of nth finite element, $\mathbf{n}_{k,i}^{(n)}$ is the normal to the edge k-i of nth finite element and ξ_i is a classic simplex coordinate, which is unitary in node i and which is zero in the opposite edge. The meaning of the rest of symbols is similar.

Let us observe behavior of the shape function (2.3B.10) in node 0. Here, the simplex coordinate ξ_0 is of unitary value and ξ_1 is zero. In node 0, the shape function (2.3B.10) is perpendicular to the edge 2-0 and is oriented inside the finite element (due to the negative sign). Its magnitude equals to the reverse value of the height v_{20}. In node 1, shape function is of the direction of the normal to the edge 1-2 and its magnitude equals to the reverse value of the height v_{01}. Moving from node 0 to node 1 along the edge 0-1, the direction of the shape function (2.3B.10) continuously changes from $-\mathbf{n}_{01}^{(n)}$ to $+\mathbf{n}_{12}^{(n)}$ and its magnitude changes from the value $(1/v_{20})$ to $(1/v_{01})$. Since the shape function (2.3B.10) does not depend on the coordinate ξ_2, the described behavior is the same along all the parallels of the edge 0-1.

Now, all the components of the electric-field intensity vector are approximated in a formal way. Therefore, the next step of the finite-element method can be done: the formal approximation is substituted to the solved equation, and an approximation error (the difference between the approximate solution and the exact one) is minimized. Performing these steps, we obtain the final matrix equation:

$$\begin{bmatrix} \mathbf{S}_t^{(n)} - k_0^2 \mu_r^{(n)} \tilde{\varepsilon}_r^{(n)} \mathbf{T}_t^{(n)} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{E}_t^{(n)} \\ \mathbf{E}_z^{(n)} \end{bmatrix} = \gamma^2 \begin{bmatrix} \mathbf{T}_t^{(n)} & \mathbf{G}^{(n)} \\ \mathbf{G}^{(n)\,\mathrm{T}} & \mathbf{S}_z^{(n)} - k_0^2 \mu_r^{(n)} \tilde{\varepsilon}_r^{(n)} \mathbf{T}_z^{(n)} \end{bmatrix} \begin{bmatrix} E_t^{(n)} \\ E_z^{(n)} \end{bmatrix},$$

where

$$\mathbf{T}_t^{(n)} \mathbf{E}_t^{(n)} = \sum_{i,j} \left\{ e_{t,ij}^{(n)} \iint_{S^{(n)}} \left[\mathbf{N}_{t,rs}^{(n)} \cdot \mathbf{N}_{t,ij}^{(n)} \right] dS \right\}, \tag{3.2B.11a}$$

$$\mathbf{G}^{(n)} \mathbf{E}_z^{(n)} = \sum_m \left\{ e_{z,m}^{(n)} \iint_{S^{(n)}} \left[\mathbf{N}_{t,rs}^{(n)} \cdot \left(\nabla_t N_{z,m}^{(n)} \right) \right] dS \right\}, \tag{3.2B.11b}$$

$$\mathbf{S}_z^{(n)} \mathbf{E}_z^{(n)} = \sum_m \left\{ e_{z,m}^{(n)} \iint_{S^{(n)}} \left[\left(\nabla_t N_{z,q}^{(n)} \right) \cdot \left(\nabla_t N_{z,m}^{(n)} \right) \right] dS \right\}, \tag{3.2B.11c}$$

$$\mathbf{T}_z^{(n)} \mathbf{E}_z^{(n)} = \sum_m \left\{ e_{z,m}^{(n)} \iint_{S^{(n)}} \left[N_{z,q}^{(n)} N_{z,m}^{(n)} \right] dS \right\},$$

(3.2B.11d)

$$\mathbf{S}_t^{(n)} \mathbf{E}_t^{(n)} = \sum_{i,j} \left\{ e_{t,ij}^{(n)} \iint_{S^{(n)}} \left[\left(\nabla_t \times \mathbf{N}_{t,rs}^{(n)} \right) \cdot \left(\nabla_t \times \mathbf{N}_{t,ij}^{(n)} \right) \right] dS \right\}.$$

(3.2B.11e)

In the above-given relations, $\mathbf{E}_t^{(n)}$ is column vector of three unknown edge approximation coefficient (related to the approximation of transversal components of electric-field intensity vector) over n^{th} finite element and $\mathbf{E}_z^{(n)}$ denotes column vector of three unknown nodal approximation coefficients (related to the approximation of longitudinal component of electric-field intensity vector) over n^{th} finite element. Next, γ denotes complex propagation constant propagation constant, k_0 is wave number in vacuum, $\mu_r^{(n)}$ is relative permeability of nth finite element and $\tilde{\varepsilon}_r^{(n)}$ is complex relative permittivity of the same element. Symbol dS denotes elementary facet for the integration over n^{th} finite element and $S^{(n)}$ is the total surface of nth finite element. Summation over the index m represents addition over all nodes of the finite element (i.e., $m = 0, 1, 2$) and summation over indexes i, j represents addition over all edges of the element (i.e., $i, j = 0\text{-}1, 1\text{-}2, 2\text{-}0$). Symbols $e_{t,ij}^{(n)}$ represent edge approximation coefficients, symbols $e_{z,m}^{(n)}$ represent nodal approximation coefficients.

Matrices $\mathbf{T}_t^{(n)}$, $\mathbf{G}^{(n)}$, $\mathbf{S}_z^{(n)}$, $\mathbf{T}_z^{(n)}$ and $\mathbf{T}_t^{(n)}$ are matrices of coefficients of nth finite element of the size 3 x 3. Elements of the above-described matrices were computed by the integration of the product of basic functions and weighting ones (or their derivatives) over n^{th} finite element (in simplex coordinates, of course). Those matrices can be evaluated exploiting the following relations:

$$\mathbf{S}_t^{(n)} = \frac{1}{A^{(n)}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

(3.2B.12a)

$$\mathbf{T}_t^{(n)} = \frac{1}{12} \sum_{i=0}^{2} \mathbf{Q}_i \, cotg\left[\theta_i^{(n)} \right],$$

(3.2B.12b)

$$\mathbf{G}^{(n)} = \frac{1}{6} \sum_{i=0}^{2} \mathbf{C}_i \, cotg\left[\theta_i^{(n)} \right],$$

(3.2B.12c)

$$\mathbf{S}_z^{(n)} = \frac{1}{2} \sum_{i=0}^{2} \mathbf{D}_i \, cotg\left[\theta_i^{(n)} \right],$$

(3.2B.12d)

$$\mathbf{T}_z^{(n)} = \frac{A^{(n)}}{12} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

(3.2B.12e)

where

$$\mathbf{Q}_0 = \begin{bmatrix} 3 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix}, \quad \mathbf{Q}_1 = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 1 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 3 \end{bmatrix},$$

$$\mathbf{C}_0 = \begin{bmatrix} 0 & -2 & 2 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{C}_1 = \begin{bmatrix} -1 & 0 & 1 \\ 2 & 0 & -2 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 0 \\ -2 & 2 & 0 \end{bmatrix},$$

$$\mathbf{D}_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \quad \mathbf{D}_1 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{D}_2 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$A^{(n)}$ is surface of nth finite element and $\theta_i^{(n)}$ is angle at i^{th} vertex of nth finite element. The above-given relations are valid for the following organization of nodes and edges:

$$\mathbf{E}^{(n)} = \begin{bmatrix} e_{z,0}^{(n)} & e_{z,1}^{(n)} & e_{z,2}^{(n)} & e_{t,12}^{(n)} & e_{t,20}^{(n)} & e_{t,01}^{(n)} \end{bmatrix}^T.$$

(3.2B.12f)

Solving the matrix equation for the vector of unknown approximation coefficients, the solution of the problem is obtained. Substituting approximation coefficients to the formal approximation, a real approximation of a sought function in every point of nth finite element is obtained. Associating approximations over all finite elements, a global approximation of the solution of the partial differential equation in all the points of the analyzed structure is found.

In layer C, a matlab program analyzing a shielded microstrip transmission line by the described finite-element method is introduced. A practical programmer's description is given in the layer D.

3.2 Shielded microstrip tranmission lines

Matlab program

The presented program serves to the computation of electric-field intensity of modes, which propagate along the shielded microstrip transmission line. The shielding waveguide is of the dimensions 12.700 mm x 12.705 mm and is fabricated from the perfect-electric conductor. At the bottom of the waveguide, a dielectric substrate is placed (lossless dielectrics of the relative permittivity $\varepsilon_r = 4.2$ and height h = 1.27 mm). At the center of the upper side, a microstrip of the width w = 1.27 mm and of the negligible thickness $t \approx 0$ mm is etched; a perfect electric conductivity of the microstrip is assumed again. Above the substrate, vacuum is expected. Thanks to the symmetry of the structure, a half of the transmission line is analyzed only. Plane of symmetry is replaced by perfect electric conductor is odd modes are computed, and by perfect magnetic conductor if even modes are investigated. The term even (odd) mode is related to such modes, which electric-field intensity distribution is even (odd) function with respect to the plane of the symmetry.
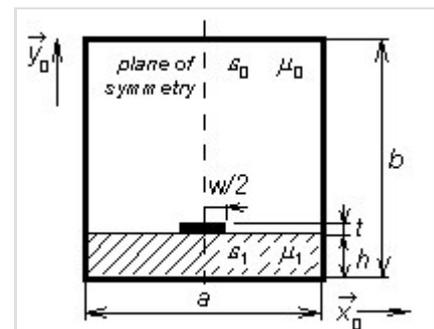
Fig. 3.2C.1 *Shielded microstrip transmission line*

The program `microstrip` can be found (including all the necessary m-files) here. Hence, we have to run Matlab and have to set the path to the folder **Microstrip**. Typing `microstrip` to the command window of Matlab, the program is started. The next steps follow:
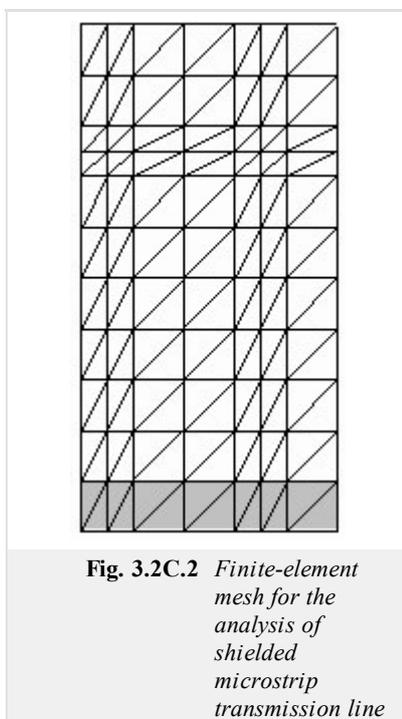

Fig. 3.2C.2 *Finite-element mesh for the analysis of shielded microstrip transmission line*

1. We type ε_{r1} (see fig. **1**) to the input line **relative permittivity of substrate**, we type tan δ to the input line **dissipation factor of substrate** and we type the frequency, on which the analysis is performed, to the input line **frequency [GHz]**. Then, we select modes (buttons Even modes , Odd modes) we want to compute. Then, the program performs analysis of the transmission line using the finite-element mesh, which is depicted in fig **2**.

2. When the selected type of the analysis is finished, a dialog containing the line **Mode number** is opened. The lowest mode is of number 1, higher modes are numbered sequentially according to the magnitude of the phase constant. Requested mode can be chosen by the scrollbar. Only modes, which can propagate along the transmission line on a given frequency, are available. Pressing the button Draw , two results are displayed: the first one shows the distribution of the transversal components of electric-field intensity, the second one shows the distribution of the longitudinal component.

3. Pressing the button End , the program is terminated.

The program `microstrip` is described from the programmer's point of view in the layer D.

3.2 Shielded microstrip tranmission lines

Developing Matlab

Software implementation of the finite-element method applied to the analysis of a shielded microstrip transmission line is similar to the analysis of waveguides (part 3.1, layer D). Therefore, we briefly discuss common features and differences of the approaches.

1. **Numbering of nodes and edges.** Local numbering of edges stays the same as given in the description of the analysis of rectangular waveguide. Considering triangles lying on the edge, numbering starts at the left down vertex and continues in the anti-clockwise direction. Considering triangles standing on a vertex, numbering starts at the right upper vertex and continues in the clockwise direction. Number of a local edge is then given by number of the opposite vertex. The described numbering is depicted in fig. **1**.
The global numbering starts by numbering diagonal edges from the left down element and continues through rows of finite bi-elements to the upper right element. Then, the numbering continues the same way for horizontal edges and for vertical ones. At the end, nodes are numbered the same way as in the case of rectangular waveguide.
Nodes and edges can be numbered by an arbitrary way (the described numbering is nor *compulsory*). Nevertheless, the numbering convention has to be kept in the whole program, i.e., when the matrix **C** is built, when local nodes and edges are associated to global ones, when boundary conditions are introduced, when results are visualized, etc.

Fig. 3.2D.1 *Numbering local edges*

Fig. 3.2D.2 *Shielded microstrip transmission line*
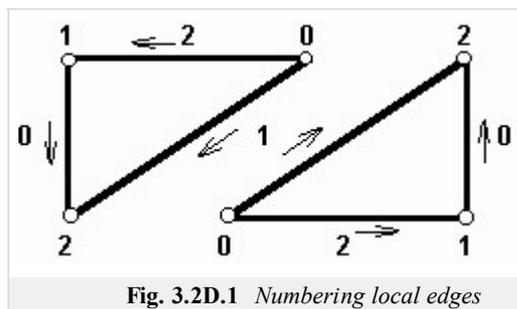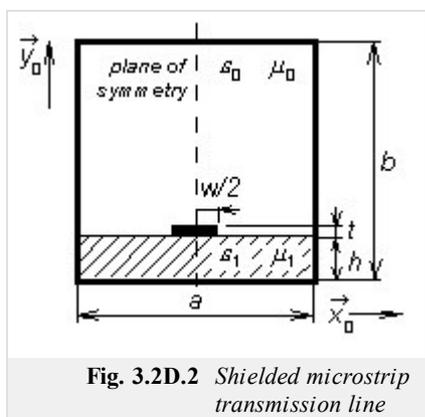
2. **Composing matrix C.** The matrix **C** serves to associating local edges and nodes to global ones. The part of **C** related to associating nodes stays the same as in the case of a rectangular waveguide (in the first column, ones are put to all rows, which correspond to local nodes associated to the first global node, etc.). The part of **C**, which associates local edges, contains maximally two ones in a column (maximally two local edges can form the global one). Moreover, one of ones has to be negative because two neighboring edges are of the opposite orientation (see fig. **1**).

3. **Boundary conditions.** Analyzing the structure, we have to ensure that electric-field intensity components, which are in parallel to perfectly electric-conducting surfaces, are zero. Therefore, we have to eliminate those columns and rows of global matrices, which correspond to nodes and edges on such surfaces. Moreover, we have to eliminate columns and rows, which correspond to nodes and edges inside metallic parts.

Except for the above-described exceptions, the program is identical comparing to the program for the analysis of a homogeneous waveguide.

3.2 Shielded microstrip tranmission lines

Java applet

3.2 Shielded microstrip tranmission lines

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

In the shielded microstrip transmission line, hybrid modes can propagate. Those modes …

Possible answers for question 1:

- … are created by the mutual coupling of transversal electric and magnetic modes.
- … are a special form of the transversally electromagnetic wave (the structure looks like a coaxial transmission line).
- … are created by the mutual coupling of longitudinal section electric and magnetic modes.
- I won't answer this question

Question 2

When analyzing microstrip transmission lines …

Possible answers for question 2:

- … non-simplified Maxwell equations have to be considered.
- Maxwell equations are not needed – computing current on the microstrip is satisfactory.
- Maxwell equations can be simplified – longitudinal components have to be considered only.
- I won't answer this question

Question 3

Analyzing microstrip transmission lines by finite elements …

Possible answers for question 3:

- … there is no difference compared to the analysis of the hollow waveguide.
- … transversal field components have to be approximated by edge elements to preserve field continuity at boundaries of dielectrics.
- … longitudinal field components have to be approximated by edge elements to preserve field continuity at boundaries of dielectrics.
- I won't answer this question

Question 4

Speaking about finite elements spatial distribution of field components can be approximated by the linear combination of…

Possible answers for question 4:

- … known basis functions and unknown coefficients.
- … known basis functions and unknown time expansions.
- … eigenvalues and eigenvectors.
- I won't answer this question

Question 5

The critical frequency of the dominant mode of a shielded planar transmission line …

see result

Chapter 4: Antennas

Introduction

Antennas can be considered as one of the most important parts of the radio-communication chain because antennas transform an electromagnetic wave propagating along the transmission line to the wave propagating in free-space and vice versa. Consequently, antenna parameters (directivity pattern, impedance, gain) significantly influence final properties of the radio link.

In this chapter, an overview of antennas and technical details of their design and exploitation are not presented. We consider these topics as well understandable and well described in the existing literature (in [1], e.g.). We turn our attention to the principles of radiation and to the ways of its mathematical modeling.

The initial paper describes the modeling of a wire dipole by the method of moments. Using this method, a current distribution on the antenna is computed as a primarily quantity. If the current distribution is known, all the other parameters can be computed. The complete description is given in the layer A. In the layer B, a Dutch translation of the layer A is given.

The second paper deals with the analysis of wire antennas consisting of more elements. We become familiar with the method of induced electromotoric forces, which enables to compute properties of not only self-standing dipoles but too their mutual coupling.

Even the third paper is devoted to the analysis of wire antennas consisting of more elements. Our attention is turned to Yagi antenna and to its analysis by the method of moments. Comparing papers 2 and 3, the reader can compare both the approaches.

In the fourth paper and in the fifth paper, wire antennas are left and the attention is turned to microstrip antennas. First, a microstrip dipole is analyzed using one-dimensional method of moments. Second, a patch antenna analyzed by two-dimensional method of moments.

Analysis of microstrip antennas is much more complicated compared to the analysis of wire antennas. This is given by the fact that microstrip antennas are manufactured exploiting dielectric substrates, and therefore, analysis methods have to respect fulfilling continuity conditions on the boundary of two media.

4.1 Wire dipole

Basic theory

In this chapter, we introduce the reader to the moment-method computing of parameters of a wire dipole. We present all the information in English in this layer so that the reader can become familiar with English terms, which are used in the area of antennas and computational electromagnetics. The Dutch translation of this chapter can be found in the layer B.
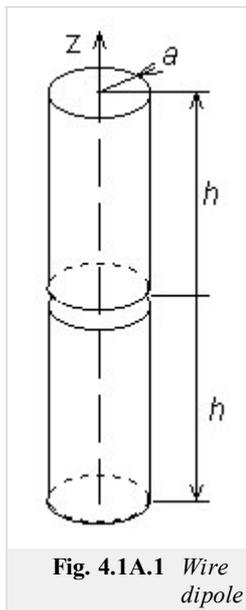

Fig. 4.1A.1 *Wire dipole*

All the technical parameters of antennas (gain, input impedance, directivity pattern) can be computed if current distribution on the antenna surface is known. Unfortunately, computation of current distribution makes troubles because integral equations have to be solved.

There are two basic approaches to the solution of integral equations – iterative and moment ones. Iterative methods are based on rough approximation of the current distribution (e.g. sinus one) that is iteratively made more accurate. On the other hand, moment methods transform integral equations to a set of simultaneous linear equations, which are solved by matrix operations.

In this chapter, our attention is turned to the moment analysis of wire antennas. In all the cases, antennas are supposed to be circular cylinders of the radius a and of the length $2h$. The antenna axis is situated to the axis z (fig. **4.1A.1**) of the cylindrical coordinate system (r, ρ, z). The antenna is placed to the vacuum $(\mu = \mu_0, \varepsilon = \varepsilon_0, \sigma = 0)$ and no losses are considered.

In the center of the cylinder $(z=0)$, there is a short gap. In the gap, we assume a hypothetical harmonic generator, which produces an azimuthally symmetric exciting electrical field (fig. **4.1A.2**). The voltage across the gap

$$V = -\int_{gap} E_z dz \qquad (4.1A.1)$$

is supposed to be 1V. In (4.1A.1), E_z is the z-component of the exciting electrical field intensity on the interpolated antenna surface (fig. **4.1A.2**). Outside the gap, E_z is zero due to the perfect conductivity> of the cylinder.
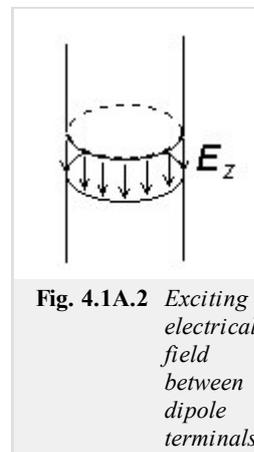

Fig. 4.1A.2 *Exciting electrical field between dipole terminals*

I. Methods of Moments

Let us to consider a general integral equation

$$\int_a^b f(z, \xi)d\xi = g(z), \qquad (4.1A.2)$$

where f is an unknown function, $<a,b>$ is the integration interval and g is a known function describing sources. The moment solution of (4.1A.2) can be reached within 3 steps:

1. The unknown function f is approximated by a linear combination of known basis functions f_n and unknown coefficients c_n

$$f \approx \tilde{f} = \sum_{n=1}^N c_n f_n. \qquad (4.1A.3)$$

2. The approximation of the unknown function $\tilde{f}$ is substituted back into the solved equation (4.1A.2). After that, the summation and the integration are swapped. This yields

$$\sum_{n=1}^N c_n \int_a^b f_n(z, \xi)d\xi = g(z) + R(z). \qquad (4.1A.4)$$

   Here, $R(z)$ is the residuum which expresses the fact that the approximation $\tilde{f}$ does not meet (4.1A.2) exactly. Equation (4.1A.4) is one equation for N unknown coefficients c_n.

3. The approximation $\tilde{f}$ is as accurate as possible if the residuum R is minimal. Hence, the residuum is minimized by the method of weighted residua : product of a weighting function w and the residuum R integrated over the region of interest $<a, b>$ has to be zero **[5]**. If N weighting functions are used then the set of N simultaneous linear equations for N unknown coefficients c_n is obtained

$$\int_a^b w_m(z) R(z) dz = 0 \quad m = 0, 1, \ldots N, \tag{4.1A.5a}$$

$$\sum_{n=1}^N c_n \int_a^b w_m(z) \int_a^b f_n(z, \xi) d\xi dz = \int_a^b w_m(z) g(z) dz. \tag{4.1A.5b}$$

Both basis functions and weighting functions have to be linearly independent on the interval $<a,b>$.

II. Basis Functions

Basis functions can be global or local ones. Global basis functions are defined on all the region of interest $<a,b>$. E.g., system of functions

$$f_n(z) = \cos\left(\frac{\pi n z}{h}\right) \tag{4.1A.6}$$

is on $<a,b>$ linearly independent and coefficients c_n in the approximation

$$f(z) \approx \tilde{f}(z) = \sum_{n=1}^N c_n f_n = \sum_{n=1}^N c_n \cos\left(\frac{\pi n z}{h}\right) \tag{4.1A.7}$$

have got then meaning of Fourier coefficients of the current distribution.

Approximation based on the global basis functions is called the single-basis approximation.

Local basis functions are defined on all the region too but each of them is non-zero only on a sub-region of the interval of interest $<a,b>$ as can be seen in fig. **4.1A.3**. If basis functions are normalized then coefficients c_n have got the meaning of nodal values (samples) of the computed function f (fig. **4.1A.3**). Approximation based on the local basis functions is called the multi-basis approximation.
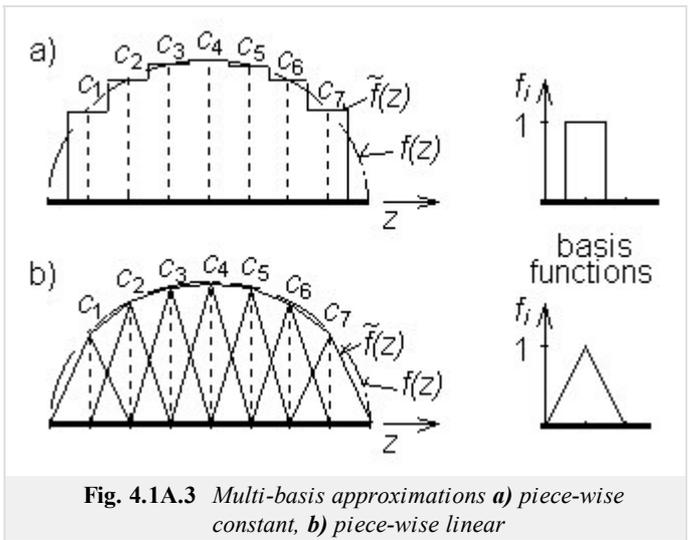

Fig. 4.1A.3 *Multi-basis approximations **a)** piece-wise constant, **b)** piece-wise linear*

III. Weighting functions

Point matching and Galerkin's methods are the most common ways of the residuum minimization.

Point matching (or collocation) uses Dirac pulses, which are place to points where the values of unknown current distribution are computed, as weighting functions

$$w_m(z) = \delta(z - z_m). \tag{4.1A.8}$$

Point matching method exhibits very low computational requirements because one integration is eliminated in 4.1A.5b) thanks to the filtering property of Dirac pulses

$$\sum_{n=1}^N c_n \int_a^b f_n(z_m, \xi) d\xi = g(z_m). \tag{4.1A.9}$$

On the other hand, the residuum minimization is related to the matching points z_m only.

In Galerkin's method, weighting functions are identical with basis ones

$$w_m(z) = f_m(z). \tag{4.1A.10}$$

Galerkin's method exhibits higher computational requirements in comparison with point matching because one of integrations is not eliminated in this case. On the other hand, the residuum minimization is performed with all the points $z \in <a,b>$.

IV. Wire antennas

Assume the cylindrical antenna of fig. **4.1A.1**. Then, the radiated electromagnetic field can be expressed in terms of vector potentials **A** and scalar potentials φ, respectively. Potentials have to meet inhomogeneous wave equations [2]

$$\frac{\partial^2 A_z(z)}{\partial z^2} + k^2 A_z(z) = -\mu_0 J_z(z),$$ (4.1A.11a)

$$\frac{\partial^2 \varphi(z)}{\partial z^2} + k^2 \varphi(z) = -\frac{\rho(z)}{\varepsilon_0}.$$ (4.1A.11b)

Here, J_z is the z-component of the current density [A.m^{-2}] impressed to the antenna by the source, ρ is the volume charge density [C.m^{-3}] on the antenna, A_z is the z-component of the vector potential and φ is the scalar potential, $k = 2\pi/\lambda$ is the wave-number and λ is the wavelength.

The current flowing on the antenna causes charge accumulation at the antenna cylinder. This fact can be described by the continuity equation [2]

$$\frac{\partial J_z(z)}{\partial z} + j\omega\rho(z) = 0.$$ (4.1A.12a)

If radius of the antenna cylinder is much smaller than the wavelength $a \ll \lambda$ then the current and charge can be assumed to be concentrated in the axis of the cylinder [5], and solving (4.1A.11) yields [2]

$$A_z(z) = \frac{\mu}{4\pi} \int_{2h} I_z(\xi) \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)} d\xi,$$ (4.1A.12b)

$$\varphi(z) = \frac{1}{4\pi\varepsilon} \int_{2h} \sigma(\xi) \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)} d\xi.$$ (4.1A.12c)

Here, $I_z(\xi)$ is the current [A] flowing in the axis of wire, $\sigma(\xi)$ denotes the length charge density [C.m^{-1}] on the axis of wire, $R(z,\xi)$ is the distance between the location ξ of electromagnetic field sources $I_z(\xi)$ and $\sigma(\xi)$ and the location z potentials $A(z)$ and $\varphi(z)$.

On the basis of $A(z)$ and $\varphi(z)$, electrical intensity of the field radiated by the antenna can be computed [2]

$$E_z^S(z) = -j\omega A_z(z) - \frac{\partial \varphi(z)}{\partial z}.$$ (4.1A.12d)

Electrical intensity has to meet the boundary condition on the antenna surface S

$$E_z^i + E_z^S = 0 \text{ on } S$$ (4.1A.12e)

E_z^i denotes electrical intensity of the incident wave. In the case of an usual transmitting antenna, E_z^i is the intensity by the feeding source (on the antenna surface), i.e. the intensity in the exciting gap (fig. **4.1A.2**).

If the current distribution on the antenna is to be computed then the set of equations (4.1A.12) has to be solved.

In order to meet the boundary condition 4.1A.12e), electrical intensity (and consequently potentials) has to be computed on the surface of the wire. That is why the distance R is described by the equation

$$R(z,\xi) = \sqrt{a^2 + (z-\xi)^2}.$$ (4.1A.13)

In the following paragraphs, piece-wise constant basis functions and Dirac weighting functions are used to solve (4.1A.12).

In the first step, the analyzed structure has to be discretized. Segmentation of the antenna is depicted in fig. **4.1A.4**. Lower bounds of segments are signed by "-", upper ones by "+". Lower bound of the first segment and upper bound of the last one are shifted from the ends of the antenna in order to meet the condition $I(-h)=I(h)=0$. Segments' lengths are $\Delta = 2\alpha$.

Substitung piece-wise constant approximation to the integral equations 4.1A.12b,c) yields

$$A_z(z) \approx \frac{\mu}{4\pi} \sum_{n=1}^{N} I_n \int_{-h+(n-0,5)\Delta}^{-h+(n+0,5)\Delta} \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)} d\xi,$$ (4.1A.14b)

Fig. 4.1A.4 *Piece-wise constant approximation*

$$\varphi(z) \approx \frac{1}{4\pi\varepsilon} \sum_{n=1}^{N} \sigma_n \int_{-h+(n-0,5)\Delta}^{-h+(n+0,5)\Delta} \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)} d\xi.$$ (4.1A.14c)
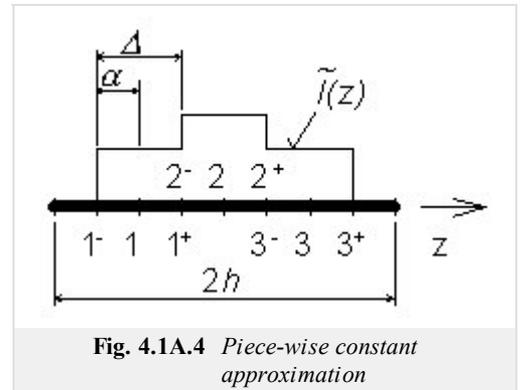
Here, I_n and σ_n are nodal values of current and charge density distributions.

Since the first derivative of the piece-wise constant approximation is zero on the constant sections and does not exist on their borders, 4.1A.12a) and 4.1A.12d) are rewritten in terms of finite differences. If the fact that $I_n = I_z(-h+n\Delta)$ is considered then the continuity equation can be expressed as

$$\frac{I_z\big(-h+(n+1)\Delta\big)-I_z\big(-h+n\Delta\big)}{\Delta} + j\omega\sigma\big(-h+(n+0,5)\Delta\big) \approx 0 \qquad (4.1A.15a)$$

and the relation for computing electrical intensity is then of the form

$$E_z^s\big(-h+n\Delta\big) \approx -j\omega A_z\big(-h+n\Delta\big) - \frac{\varphi\big[-h+(n+0,5)\Delta\big]-\varphi\big[-h+(n-0,5)\Delta\big]}{\Delta}. \qquad (4.1A.15d)$$

Relations 4.1A.15a) and 4.1A.15d) show that Dirac pulses for point matching have to be placed to the center of segments for the vector potential

$$A_z\big(-h+m\Delta\big) \approx \frac{\mu}{4\pi} \sum_{n=1}^{N} I_n \int_{-h+(n-0,5)\Delta}^{-h+(n+0,5)\Delta} \frac{\exp\big[-jkR(-h+m\Delta,\,\xi)\big]}{R(-h+m\Delta,\,\xi)} d\xi \qquad (4.1A.15b)$$

and to borders of segments for the scalar potential

$$\varphi\big[-h+(m+0,5)\Delta\big] \approx \frac{1}{4\pi\varepsilon} \sum_{n=1}^{N} \sigma_{n+} \int_{-h+n\Delta}^{-h+(n+1)\Delta} \frac{\exp\big\{-jkR\big[-h+(m+0,5)\Delta,\,\xi\big]\big\}}{R\big[-h+(m+0,5)\Delta,\,\xi\big]} d\xi. \qquad (4.1A.15c)$$

In 4.1A.15c), $\sigma_{n+} = \sigma[-h+(n+0.5)\Delta]$.

Now, (4.1A.15) can be rewritten into a more compact form

$$\sigma_{n+} \approx \frac{-1}{j\omega}\left[\frac{I_{n+1}-I_n}{\Delta}\right], \qquad (4.1A.16a)$$

$$A_z(m) \approx \frac{\mu}{4\pi} \sum_{n=1}^{N} I_n \int_{\Delta_n} \frac{\exp\big[-jkR(m,\,\xi)\big]}{R(m,\,\xi)} d\xi, \qquad (4.1A.16b)$$

$$\varphi(m^+) \approx \frac{1}{4\pi\varepsilon} \sum_{n=1}^{N} \sigma_{n+} \int_{\Delta_{n+}} \frac{\exp\big[-jkR(m^+,\,\xi)\big]}{R(m^+,\,\xi)} d\xi, \qquad (4.1A.16c)$$

$$-E_z^i(m) \approx -j\omega A_z(m) - \frac{\varphi(m^+)-\varphi(m^-)}{\Delta}. \qquad (4.1A.16d)$$

In 4.1A.16d), the boundary condition 4.1A.12e) is included.

Now, let's have a look at the continuity theorem 4.1A.16a); it expresses the fact that segments of the antenna can be replaced by elementary electrical dipoles (fig. **4.1A.5**). Taking this idea in mind, submission of nth segment to the scalar potential can be computed on the basis of 4.1A.16c) as

$$\varphi(m^+) = \frac{1}{j\omega\varepsilon}\left[I_n \int_{\Delta_{n+}} \frac{\exp(-jkR)}{4\pi R} d\xi - I_n \int_{\Delta_{n-}} \frac{\exp(-jkR)}{4\pi R} d\xi \right]\frac{1}{\Delta}. \qquad (4.1A.17)$$

Substituting (4.1A.17) and (4.1A.16b) to 4.1A.16d) and multiplying both sides by Δ yields

$$\mathbf{E}_z^i\Delta = \mathbf{Z}\,\mathbf{I}, \qquad (4.1A.18)$$

where

$$Z_{mn} = j\omega\mu\Delta \int_{\Delta_n} \frac{\exp\big[-jkR(m,\,\xi)\big]}{4\pi R(m,\,\xi)} d\xi + \qquad (4.1A.19)$$

$$+\frac{1}{j\omega\varepsilon}\left[\int\limits_{\Delta_n+}\frac{\exp\left[-jkR(m^+,\xi)\right]}{4\pi R(m^+,\xi)}d\xi-\int\limits_{\Delta_n-}\frac{\exp\left[-jkR(m^+,\xi)\right]}{4\pi R(m^+,\xi)}d\xi\right]\frac{1}{\Delta}-$$

$$-\frac{1}{j\omega\varepsilon}\left[\int\limits_{\Delta_n+}\frac{\exp\left[-jkR(m^-,\xi)\right]}{4\pi R(m^-,\xi)}d\xi-\int\limits_{\Delta_n-}\frac{\exp\left[-jkR(m^-,\xi)\right]}{4\pi R(m^-,\xi)}d\xi\right]\frac{1}{\Delta},$$

denotes submission of current and charge on n^{th} segment to the voltage induced on the mth segment.

Since electrical intensity is zero on all the segments except of the source gap, elements of voltage vector are zero except of the gap-segment corresponding element that equals 1. Then, (4.1A.18) provides the current distribution **I**. Ratio of input voltage and input current gives then the input impedance of the analyzed antenna.

An example of the analysis results is depicted on fig. **4.1A.6**; module and phase of the current distribution of the dipole $h = \lambda$ and $a = 0.001588\,\lambda$ is plotted there.

The presented results (fig. **4.1A.6**) can be obtained using a computer program, whose user's guide is presented in the layer C. In the layer D, the program is described from the programmer's point of view.
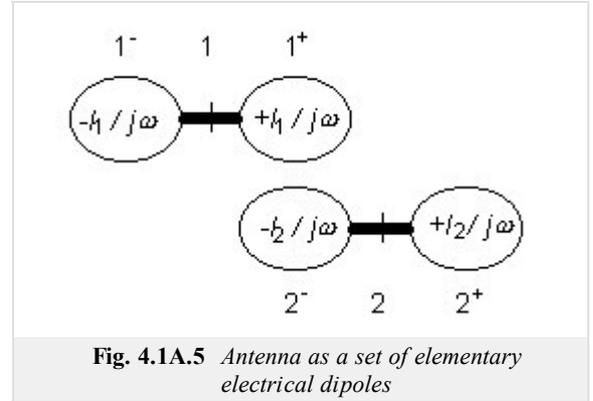
Fig. 4.1A.5 *Antenna as a set of elementary electrical dipoles*
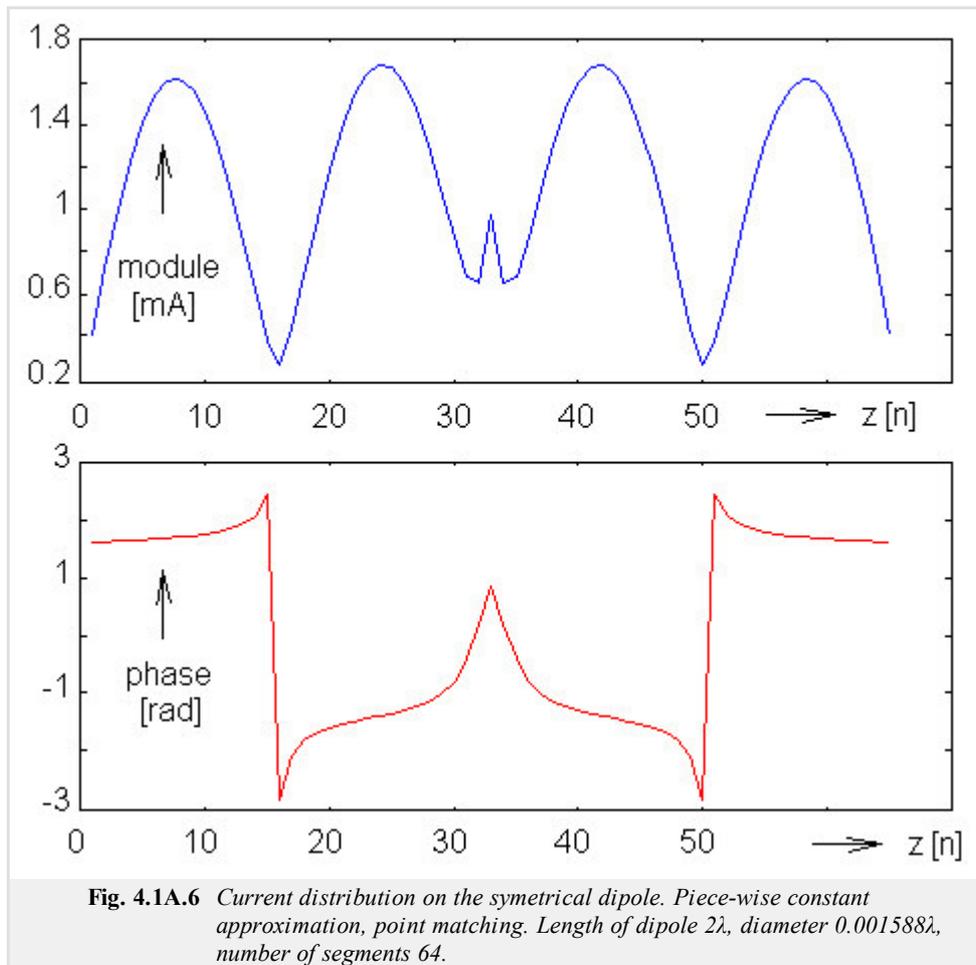
Fig. 4.1A.6 *Current distribution on the symetrical dipole. Piece-wise constant approximation, point matching. Length of dipole 2λ, diameter 0.001588λ, number of segments 64.*

4.1 Wire dipole

Advanced theory

In dit hoofdstuk introduceren we de lezer in de moment-methode berekening van parameters van een wiredipole. We presenteren deze informatie in het Nederlands in laag B zodat de lezer vertrouwd kan geraken met de Nederlandse termen die gebruikt worden bij antennes. De Engelse vertaling van dit hoofdstuk vindt men in laag A.
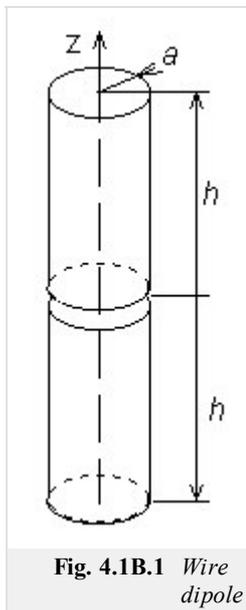

Fig. 4.1B.1 *Wire dipole*

De technische parameters van antennes (versterking, ingangsimpedantie, derectivity pattern) kunnen berekend worden als de current distribution van het antennedraagvlak gekend is. Spijtig genoeg geeft de berekening van current distribution problemen omdat er integraal vergelijkingen moeten worden opgelost.

Er zijn 2 basisbenaderingen voor het oplossen van integraalvergelijkingen. Deze zijn interatieve en momentopnames. Interative methodes zijn gebaseerd op de ruwe benadering van de current distribution (v.b. sinus) waardoor deze zo nauwkeuriger wordt gemaakt. Anderzijds, moment methodes vormen de integraal vergelijkingen om tot een set van parallelle lineaire vergelijkingen, welke opgelost kunnen worden door matrix bewerkingen.

In dit hoofdstuk gaat onze aandacht vooral naar het moment analyse van draadantennes. In alle gevallen worden de antennes verondersteld om cirkelvormige cilinders te zijn met een straal a en een lengte $= 2h$. De as van de antenne wordt voorgesteld door de z-as (fig. **4.1B.1**) van het cilindervormig coördinaten systeem (r, ρ, z). De antenne wordt in een vacuüm omgeving geplaatst ($\mu = \mu_0$, $\varepsilon = \varepsilon_0$, $\sigma = 0$) en er worden geen verliezen verondersteld.

In het centrum van de cilinder ($z=0$), is er een korte opening. In deze opening veronderstellen we een harmonische generator welke een symmetrisch elektrisch veld genereert (fig. **4.1B.2**). De spanning wordt verondersteld

$$V = -\int_{gap} E_z dz \qquad (4.1B.1)$$

hier 1V te bedragen. In (4.1.B.1) is E_z de z-component van de elektrische veldintensiteit van de antenne oppervlakte (fig. **4.1B.2**). Buiten het gat, E_z is nul omwille van de perfecte geleiding van de cilinder.
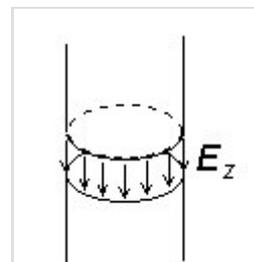
I. Moment methode

Laten we een algemene integraalvergelijking nemen.

Fig. 4.1B.2 *Exciting electrical field between dipole terminals*

$$\int_a^b f(z, \xi)d\xi = g(z). \qquad (4.1B.2)$$

In deze formule is f een niet gekende functie, $\langle a,b \rangle$ is het interval en g is een gekende functie die wordt voorgesteld door de bron. De moment oplossing van (4.1.B.2) kan men bekomen in 3 stappen:

1. De ongekende functie f kan men herschrijven door een lineaire combinatie van een gekende basisfunctie f_n en een ongekende coëfficiënt c_n.

$$f \approx \tilde{f} = \sum_{n=1}^N c_n f_n. \qquad (4.1B.3)$$

2. De benadering van de ongekende functie $\tilde{f}$ is terug vervangen door de opgeloste vergelijking (4.1B.2). Hier zijn de sommatie en de integratie verwisseld. Dit bekomt

$$\sum_{n=1}^N c_n \int_a^b f_n(z, \xi)d\xi = g(z) + R(z). \qquad (4.1B.4)$$

Hier is, $R(z)$ het residuum welk het feit uitdrukt dat de benadering $\tilde{f}$ niet helemaal overeen komt met (4.1B.2). De vergelijking (4.1B.4) is een vergelijking voor N aantal ongekende coëfficienten c_n.

3. De benadering $\tilde{f}$ is zo nauwkeurig mogelijk als het residuum R minimaal is. Vandaar dat het residuum geminimaliseerd wordt door de weighted residua methode : Het produkt van een weegfunctie w en het residuum R geintegreerd over het interval $\langle a, b \rangle$ moet nul zijn [5]. Als N aantal weegfuncties worden gebruikt, dan is de set van N aantal gelijkaardige lineaire vergelijkingen voor N aantal ongekende

coëfficiënten c_n gegeven in

$$\int_a^b w_m(z)R(z)dz = 0 \quad m = 0,1,\dots N,$$

(4.1B.5a)

$$\sum_{n=1}^N c_n \int_a^b w_m(z) \int_a^b f_n(z,\xi)d\xi dz = \int_a^b w_m(z)g(z)dz.$$

(4.1B.5b)

Beide basis functies en weegfuncties moeten lineair onafhankelijk zijn van het interval $<a,b>$.

II. Basis Functies

Basis functies kunnen globaal of locaal zijn. Globale basis functies worden afgebakend op het hele gebied $<a,b>$. V.b., systeemfuncties

$$f_n(z) = \cos\left(\frac{\pi n z}{h}\right)$$

(4.1B.6)

zijn op $<a,b>$ lineair onafhankelijk en de coëfficiënten c_n in volgende benadering

$$f(z) \approx \tilde{f}(z) = \sum_{n=1}^N c_n f_n = \sum_{n=1}^N c_n \cos\left(\frac{\pi n z}{h}\right)$$

(4.1B.7)

hebben dan betekenis op de Fourier coëfficiënten van de stroomdichtheids verdeling.

Deze benadering gebaseerd op de globale basis functies noemt men de single-basis benadering.

Locale basis functies worden ook afgebakend op het hele gebied, maar elk van hen is niet gelijk aan nul voor de deelgebieden van het interval $<a,b>$ zoals men kan zien in fig. **4.1B.3**. Als basis functies worden genormaliseerd, dan hebben de coëfficiënten c_n de betekenis van knoopwaardes van de berekende functie f (fig. **4.1B.3**). Benadering gebaseerd op de locale basis functies noemt men multi-basis benaderingen.
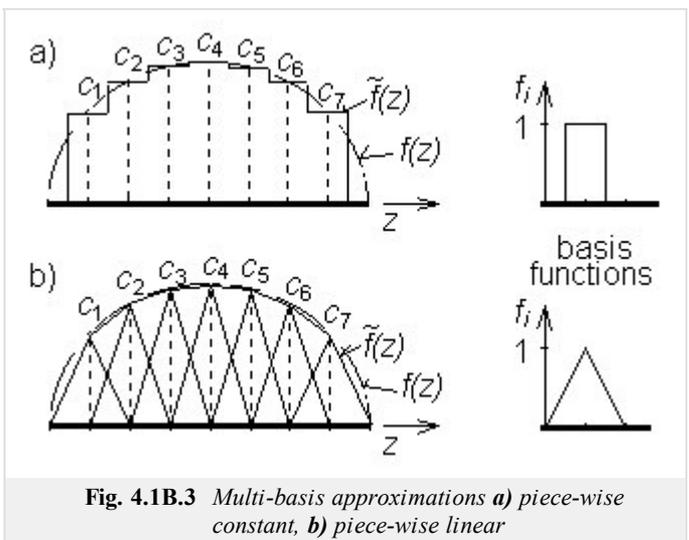

Fig. 4.1B.3 *Multi-basis approximations **a)** piece-wise constant, **b)** piece-wise linear*

III. Weeg functies

Punt matching en Galerkins methodes zijn de meest algemene vormen van residuum minimalisatie.

Punt matching (of rangregeling) maken gebruik van Dirac pulsen, welke geplaatst worden op plaatsen waar men de ongekende waarde van de stroom distributie wil berekenen, zoals weeg functies

$$w_m(z) = \delta(z - z_m).$$

(4.1B.8)

Punt matching methodes vereisen zeer weinig berekeningen omdat onze integratie geëlimineerd wordt in (4.1B.5b) dankzij de filtering van de Dirac pulsen

$$\sum_{n=1}^N c_n \int_a^b f_n(z_m,\xi)d\xi = g(z_m).$$

(4.1B.9)

Aan de andere kant is de minimalisatie van het residuum enkel verwant tot de matching points z_m.

In de Galerkins methode zijn weeg functies identiek aan basis functies

$$w_m(z) = f_m(z).$$

(4.1B.10)

Galerkins methode vertoont meer beredeneringen in vergelijking met de matching point methode omdat een van de integraties niet kan worden weggewerkt. Anderzijds is de residuum minimalisatie hier voorgesteld voor elk punt $z \in <a,b>$.

IV. Draad antennes

Veronderstel de cylindervormige antenne van fig. **4.1B.1**. Dan kan het uitgestraalde elektromagnetische veld uitgedrukt worden in een vectorpotentiaal **A** en scalair potentiaal φ. Deze moeten overeenkomen met niet-homogene golfvergelijkingen **[2]**

$$\frac{\partial^2 A_z(z)}{\partial z^2} + k^2 A_z(z) = -\mu_0 J_z(z), \qquad (\,4.1\text{B}.11\text{a}\,)$$

$$\frac{\partial^2 \varphi(z)}{\partial z^2} + k^2 \varphi(z) = -\frac{\rho(z)}{\varepsilon_0}. \qquad (\,4.1\text{B}.11\text{b}\,)$$

Hier is J_z de z-component van de stroomdichtheid [A.m^{-2}] opgewekt door de bron in de antenne. ρ is het volume van de ladingsdichtheid [C.m^{-3}] in de antenne, A_z is de z-component van het vector potentiaal en φ is het scalair potentiaal, $k=2\pi/\lambda$ is het golfaantal en λ is de golflengte.

De stroom die vloeit in de antenne veroorzaakt een spanningsophoping in de cilinder. Dit verschijnsel noemt men de continuïteitsvergelijking van de regressievergelijking **[2]**

$$\frac{\partial J_z(z)}{\partial z} + j\omega\rho(z) = 0. \qquad (\,4.1\text{B}.12\text{a}\,)$$

Als de straal van de antenne-cilinder veel kleiner is dan de golflengte $a << \lambda$ dan mag men veronderstellen dat de stroom zich concentreerd in de as van de cilinder **[5]**, en door het oplossen van (4.1B.11) bekomt men **[2]**

$$A_z(z) = \frac{\mu}{4\pi} \int_{2h} I_z(\xi) \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)} d\xi, \qquad (\,4.1\text{B}.12\text{b}\,)$$

$$\varphi(z) = \frac{1}{4\pi\varepsilon} \int_{2h} \sigma(\xi) \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)} d\xi. \qquad (\,4.1\text{B}.12\text{c}\,)$$

Hier is $I_z(\xi)$ de stroom [A] die vloeit in de as van de cilinder, $\sigma(\xi)$ is de ladingsdichtheid [C.m^{-1}] in de as van de cilinder, $R(z,\xi)$ is de afstand tussen de plaats ξ van de elektromagnetische veldbronnen $I_z(\xi)$ en $\sigma(\xi)$ en de locatie z potentiëlen $A(z)$ en $\varphi(z)$.

Op basis van $A(z)$ en $\varphi(z)$, kan de elektrische intensiteit van de veldstralen van de antenne berekend worden **[2]**

$$E_z^s(z) = -j\omega A_z(z) - \frac{\partial \varphi(z)}{\partial z}. \qquad (\,4.1\text{B}.12\text{d}\,)$$

De elektrische intensiteit moet hier gelijk zijn aan de grenswaarde van het antennedraagvlak S

$$E_z^i + E_z^s = 0 \text{ on } S \qquad (\,4.1\text{B}.12\text{e}\,)$$

E^i_z veronderstelt hier de elektrische intensiteit van de golf. In het geval van een gebruikelijke antenne is E^i_z de intensiteit van de voedingsbron (op het antennedraagvlak). V.b. de intensiteit in het gat (fig. **4.1B.2**).

Als men de stroomdistributie van de antenne wil berekenen, dan zal men eerste de set van vergelijkingen (4.1B.12) moeten uitwerken.

Als men de grensvoorwaarden wil bereiken (4.1B.12e) zal men de elektrische intensiteit moeten berekenen op het draagvlak van de draad. Dit is waarom de afstand R beschreven staat als volgt

$$R(z,\xi) = \sqrt{a^2 + (z-\xi)^2}. \qquad (\,4.1\text{B}.13\,)$$

In de volgende paragrafen worden constante basis functies, en Dirac weeg functies stapsgewijs gebruikt voor het uitwerken van (4.1B.12).

In de eerste stap wordt de geanalyseerde structuur gediscritiseerd. Opsplitsing van antenne is vertoont in fig. **4.1B.4**. Lagere delen worden voorgesteld door een "-", hogere delen door een "+". Lagere delen van het eerste segment en hogere delen van het laatste segment zijn weggelaten van het einde van de antenne om zo de conditie I(-h)=I(h)=0 te bereiken. De segmentlengtes zijn $\varDelta = 2\alpha$.

Het stapsgewijs vervangen van constante benaderingen naar de integraal vergelijking (4.1B.12b,c) bekomt

Fig. 4.1B.4 *Piece-wise constant approximation*
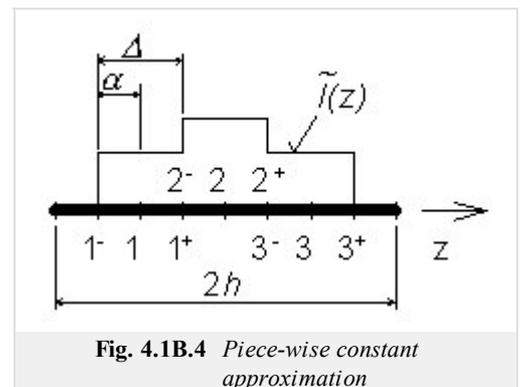
$$A_z(z) \approx \frac{\mu}{4\pi} \sum_{n=1}^{N} I_n \int_{-h+(n-0,5)\Delta}^{-h+(n+0,5)\Delta} \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)}\, d\xi, \qquad (4.1\text{B}.14\text{b})$$

$$\varphi(z) \approx \frac{1}{4\pi\varepsilon} \sum_{n=1}^{N} \sigma_n \int_{-h+(n-0,5)\Delta}^{-h+(n+0,5)\Delta} \frac{\exp[-jkR(z,\xi)]}{R(z,\xi)}\, d\xi. \qquad (4.1\text{B}.14\text{c})$$

I_n en σ_n zijn knoopwaardes van de stroomladings dichtheid distributie.

Omdat de eerste afleiding van de stapsgewijze constante benadering nul is voor de constante secties en niet bestaat in de grensgevallen, zijn (4.1B.12a) en (4.1B.12d) herschreven in termen van eindige verschillen. In het geval dat $I_n = I_z(-h+n\Delta)$ wordt dan verondersteld dat de continuiteitsvergelijking als volgt kan worden uitgedrukt

$$\frac{I_z(-h+(n+1)\Delta) - I_z(-h+n\Delta)}{\Delta} + j\omega\sigma(-h+(n+0,5)\Delta) \approx 0 \qquad (4.1\text{B}.15\text{a})$$

en de relatie voor het berekenen van elektrische intensiteit is dan van de vorm

$$E_z^s(-h+n\Delta) \approx -j\omega A_z(-h+n\Delta) - \frac{\varphi[-h+(n+0,5)\Delta] - \varphi[-h+(n-0,5)\Delta]}{\Delta}. \qquad (4.1\text{B}.15\text{d})$$

De relaties (4.1B.15a) en (4.1B.15d) tonen on dat de Dirac pulsen voor punt matching in het centrum van de segmenten worden geplaatst voor het vector potentiaal

$$A_z(-h+m\Delta) \approx \frac{\mu}{4\pi} \sum_{n=1}^{N} I_n \int_{-h+(n-0,5)\Delta}^{-h+(n+0,5)\Delta} \frac{\exp[-jkR(-h+m\Delta,\xi)]}{R(-h+m\Delta,\xi)}\, d\xi \qquad (4.1\text{B}.15\text{b})$$

en de grensen van de segmenten voor het scalair potentiaal

$$\varphi[-h+(m+0,5)\Delta] \approx \frac{1}{4\pi\varepsilon} \sum_{n=1}^{N} \sigma_{n+} \int_{-h+n\Delta}^{-h+(n+1)\Delta} \frac{\exp\{-jkR[-h+(m+0,5)\Delta,\xi]\}}{R[-h+(m+0,5)\Delta,\xi]}\, d\xi. \qquad (4.1\text{B}.15\text{c})$$

In (4.1B.15c), $\sigma_{n+} = \sigma[-h+(n+0.5)\Delta]$.

(4.1B.15) kan men herschrijven in een compactere vorm

$$\sigma_{n+} \approx \frac{-1}{j\omega}\left[\frac{I_{n+1} - I_n}{\Delta}\right], \qquad (4.1\text{B}.16\text{a})$$

$$A_z(m) \approx \frac{\mu}{4\pi} \sum_{n=1}^{N} I_n \int_{\Delta_n} \frac{\exp[-jkR(m,\xi)]}{R(m,\xi)}\, d\xi \qquad (4.1\text{B}.16\text{b})$$

$$\varphi(m^+) \approx \frac{1}{4\pi\varepsilon} \sum_{n=1}^{N} \sigma_{n+} \int_{\Delta_{n+}} \frac{\exp[-jkR(m^+,\xi)]}{R(m^+,\xi)}\, d\xi, \qquad (4.1\text{B}.16\text{c})$$

$$-E_z^i(m) \approx -j\omega A_z(m) - \frac{\varphi(m^+) - \varphi(m^-)}{\Delta}. \qquad (4.1\text{B}.16\text{d})$$

In (4.1B.16d), is de randvoorwaarde (4.1B.12e) opgenomen.

Laten we nu een kijkje nemen op de continuiteitsvergelijking (4.1B.16a); het drukt het feit uit dat genmenten van de antenne vervangen kunnen worden door elementaire elektrische dipolen (fig. **4.1B.5**). Dit in gedachten genomen, productie van het n^{th} segment van het scalair potentiaal kan berekend worden op basis van (4.1B.16c)

$$\varphi(m^+) = \frac{1}{j\omega\varepsilon}\left[I_n \int\limits_{\Delta_n+} \frac{\exp(-jkR)}{4\pi R}d\xi - I_n \int\limits_{\Delta_n-} \frac{\exp(-jkR)}{4\pi R}d\xi\right]\frac{1}{\Delta}. \qquad (4.1B.17)$$

Vervangen we (4.1B.17) en (4.1B.16b) in (4.1B.16d) en multiplexen we beide kanten door Δ dan bekomen we

$$\mathbf{E}_{\underset{z}{}}^i\Delta = \mathbf{Z}\,\mathbf{I}, \qquad (4.1B.18)$$

waar

$$Z_{mn} = j\omega\mu\Delta \int\limits_{\Delta_n} \frac{\exp[-jkR(m,\xi)]}{4\pi R(m,\xi)}d\xi +$$

$$+ \frac{1}{j\omega\varepsilon}\left[\int\limits_{\Delta_n+} \frac{\exp[-jkR(m^+,\xi)]}{4\pi R(m^+,\xi)}d\xi - \int\limits_{\Delta_n-} \frac{\exp[-jkR(m^+,\xi)]}{4\pi R(m^+,\xi)}d\xi\right]\frac{1}{\Delta} - \qquad (4.1B.19)$$

$$- \frac{1}{j\omega\varepsilon}\left[\int\limits_{\Delta_n+} \frac{\exp[-jkR(m^-,\xi)]}{4\pi R(m^-,\xi)}d\xi - \int\limits_{\Delta_n-} \frac{\exp[-jkR(m^-,\xi)]}{4\pi R(m^-,\xi)}d\xi\right]\frac{1}{\Delta},$$

de stroomlading op het n^{th} segment voorgesteld tot de spanning opgewekt in het m^{th} segment.

Omdat de elektrische intensiteit nul is voor alle segmenten behalve het voedingsgat, de elementen van de spanningsvector bedragen nul behalve voor het gatsegment dat gelijk is aan 1. (4.1B.18) voorziet de stroomdistributie $\mathbf{I}$. De verhouding van de ingangsspanning en de ingangsstroom geeft dan de ingangsimpedantie van de geanalyseerde antenne.

Een voorbeeld van de resultaten van de analyse is voorgesteld in fig. **4.1B.6**; module en fase van de stroomdichtheidsverdeling van de dipool $h = \lambda$ en $a = 0.001588\,\lambda$ is daar bekomen.

Fig. 4.1B.5 *Antenna as a set of elementary electrical dipoles*

Men kan meer informatie vinden over de bekomen resultaten (fig. **4.1B.6**) door middel van een computer programma. Dit programma is voorgesteld in laag C. In laag D is het programma uitgelegd door de programmeur.
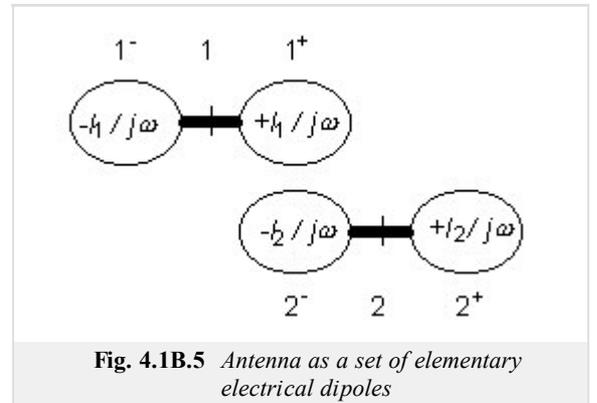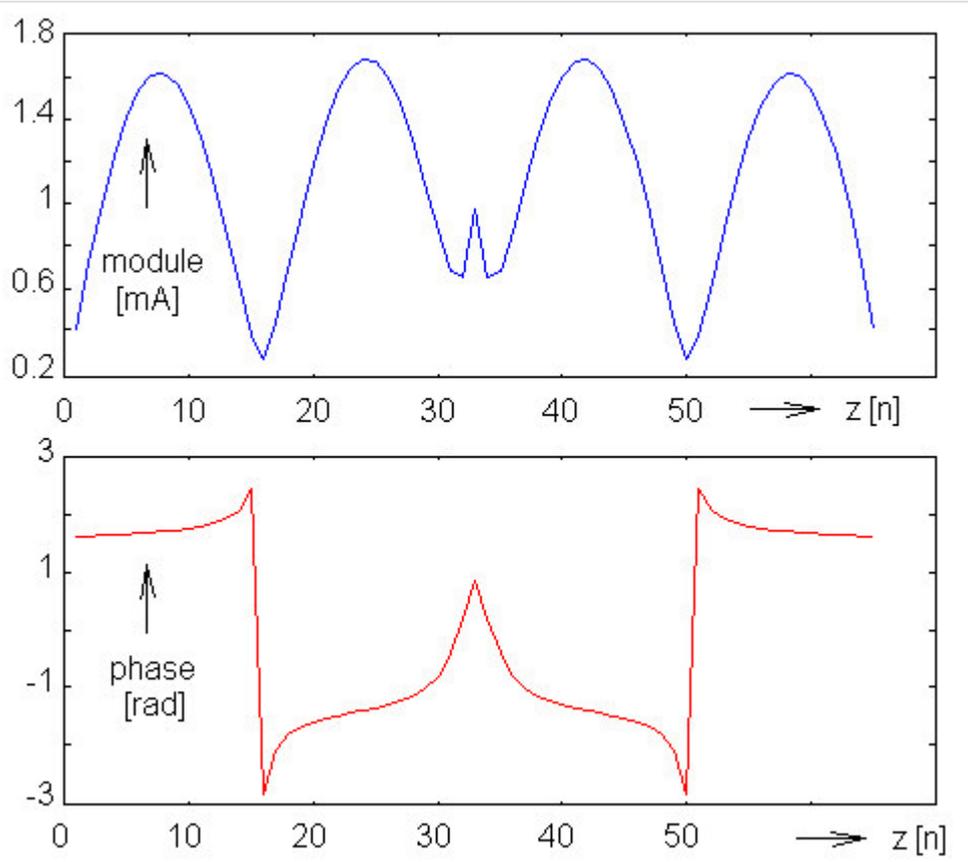
Fig. 4.1B.6 *Current distribution on the symetrical dipole. Piece-wise constant approximation, point matching. Length of dipole 2λ, diameter 0.001588λ, number of segments 64.*

4.1 Wire dipole

Matlab program

The computer program, accompanying the paper about the moment analysis of a wire dipole is available here. The program is of very simple user's interface. We run Matlab and set the path to the folder **Dipol**. In the command window, the basic m-file `dipol.m` has to be run and the appearing form has to be filled in. In the form, we type the length of the dipole (related to wavelength) and the radius of the dipole (related to wavelength). The program prints input impedance of the dipole. In two independent windows, current distribution on the antenna and directivity pattern are plotted.

The program considers the fixed number of discretization segments (`N` = 33). The excitation gap is assumed in the middle of the wire.

4.1 Wire dipole

Developing Matlab

The Matlab program for the analysis of wire antennas by the method of moments can be written relatively simply.

In the first step, the integral

$$\psi = \int_{\Delta} \left[\frac{\exp(-jkR)}{4\pi R} \right] d\xi \qquad (\ 4.1D.1\ )$$

is numerically evaluated for all the possible mutual distances between segments of the discretized antenna. In Matlab syntax:

```
alpha = 0.5*delta; % one half of the segment's length
psi = zeros( 1, N+1); % numerical integration
for m=1:(N+1)
    x = (m-1)*delta;
    psi(m) = quadl( 'green', -alpha, alpha, 1e-5, [], x, a, k);
end
psi = psi/delta;
```

In the above-given source code, `quadl` denotes the standard Matlab function, which performs numeric integration. Next, `green` is a string corresponding to the name of the m-file, which contains the integrated function. Symbols `-alpha` and `alpha` denotes integration limits, `1e-5` is the allowed integration error, and `z`, `a`, `k` are parameters (`z` is the distance between the source segment and the observation one, `a` denotes radius of the antenna wire and `k` is wave-number).

The integrated function is programmed by the following way:

```
function out = green(ksi, z, a, k)

R = sqrt( a^2 + (z-ksi)^2); % ksi ∈ (-Δ/2; +Δ/2)
out = exp(-j*k*R)/(4*pi*R); % Green's function
```

When integrals for all the possible distances between segments are evaluated, the impedance matrix can be built:

```
for m = 1:A
    for n = m:N
        dist = abs(m-n); % distance between source and destination
        hlp = 2*psi(1+dist) - psi(1+abs(dist-1)) - psi(1+abs(dist+1));
        Z(m,n) = j*omega*mi*del*psi(1+dist) + hlp/(j*omega*epsilon*del);
        Z(n,m) = Z(m,n); % matrix is symmetrical
    end
end
```

In the above-given list, `mi` and `epsilon` are permeability and permittivity of vacuum, `omega` denotes angular frequency and `j` is imaginary unit.

In the final step, the impedance matrix is inverted. The column of the inverted matrix, which corresponds to the excitation gap, directly provides nodal values of the current distribution on the antenna.

4.1 Wire dipole

Java applet

Output / Vstup dat	Module	Phase	Dir. Pattern

Input / Vstup dat

Enter the radius of the wire dipole antenna:

Radius of wire / Polomr vodie diplu

0.002

Enter the length of the wire dipole antenna:

Length of dipole / Dlka diplu

0.5

How many segments does the antenna have?

# of Segments / Poet segment (max. 250)

33

Enter the wavelength of the signal:

Wavelength / Vlnov dlka

1

Calculate / Vypoti

Help

4.1 Wire dipole

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**10** of **10**:

Question 1

The wire dipole is an antenna, which can be characterized by the following statement:

Possible answers for question 1:

- ○ The antenna can be build from linear segments of an arbitrary radius.
- ○ The radius of the antenna wire is much smaller than the antenna length and the wavelength.
- ○ The antenna is built from an arbitrary copper wire.
- ◉ I won't answer this question

Question 2

The moment method…

Possible answers for question 2:

- ○ … exploits time instants for evaluating Maxwell equations in the integral form.
- ○ … is a general numeric method for solving Maxwell equations.
- ○ … is a numeric technique for solving Maxwell equations in the integral form.
- ◉ I won't answer this question

Question 3

Directivity pattern of the antenna …

Possible answers for question 3:

- ○ … gives magnitude of field intensity of the radiated wave for all the angles in the specified plane.
- ○ … describes geometry of a wire antenna (directions to which antenna wires are oriented).
- ○ … gives an information about polarization of the radiated wave.
- ◉ I won't answer this question

Question 4

Input impedance of the antenna …

Possible answers for question 4:

- ○ … is an impedance at the input of an antenna amplifier (if no amplifier is connected to the antenna, then the input impedance is zero).
- ○ … is an impedance, which can be measured on input terminals of the antenna.
- ○ … does not depend on the power radiated by the antenna.
- ◉ I won't answer this question

Question 5

Basis functions …

Question 6

Method of weighted residua …

Possible answers for question 6:

- ⊙ … is used for minimizing the difference between the numeric solution of Maxwell equations and the exact field distribution.
- ⊙ … is exploited for matching a wire antenna (minimizing residual reflections on a feeder).
- ⊙ … is a special case of Galerkin method.
- ⦿ I won't answer this question

Question 7

Impedance matrix in the moment analysis of the antenna …

Possible answers for question 7:

- ⊙ … consists of input impedances of the antenna for different radii (rows) and different lengths (columns).
- ⊙ … contains self-impedances of antenna segments on the diagonal and mutual impedances of segments out of the diagonal.
- ⊙ … consists of input impedances for different time instants and different directions of arrival of the incident wave.
- ⦿ I won't answer this question

Question 8

Sidelobes appear in the directivity pattern of the symmetric dipole if the antenna length …

Possible answers for question 8:

- ⊙ … $l > 1.2\ \lambda$, λ denotes wavelength.
- ⊙ … $0.6\ \lambda \leq l \leq 0.7\ \lambda$, λ denotes wavelength.
- ⊙ … $l \leq 0.5\ \lambda$, λ denotes wavelength.
- ⦿ I won't answer this question

Question 9

Input impedance of the symmetric dipole is purely real (the radius a = $10^{-3}\ \lambda$ and λ denotes wavelength, the number of discretization segments N = 33) …

Possible answers for question 9:

- ⊙ … for $l = 0.485\ \lambda$.
- ⊙ … for $l = 0.462\ \lambda$.
- ⊙ … for $l = 0.473\ \lambda$.
- ⦿ I won't answer this question

Question 10

If the number of discretization segments in the moment analysis of a wire antenna is increased then …

Possible answers for question 10:

- ○ … accuracy of results decreases.
- ○ … accuracy of results raises.
- ○ … accuracy of results is not influenced.
- ● I won't answer this question

see result

4.2 Mutual impedance

Basic theory

Radiating elements (dipoles, loops, etc.) are usually grouped into an antenna array and fed from a common source in order to primarily reach the desired directivity pattern of the radiation.

Designing the feeding system of an antenna array, even the mutual influence among the antenna elements has to be considered and a true value of the impedance at the input of each antenna element has to be determined. Knowledge of these entries enables to determine input current of each antenna element and even currents on the feeding ports of the array. Moreover, the structure of the feeding system can be modified to reach a suitable value of the input impedance of the antenna array in its whole.

In this chapter, we concentrate on the method of calculating input impedance of a linear antenna (dipole) by the method of induced electromotoric forces. The approach is then going to be generalized in order to compute impedances of single elements in the antenna array.

Radiation impedance Z_Σ of the antenna can be considered as a proportionality constant between the complex power P_Σ, which is radiated by the antenna, and the squared value of the related current I

$$P_\Sigma = Z_\Sigma |I|^2 = Z_{\Sigma vst}|I_{vst}|^2 = Z_{\Sigma m}|I_m|^2. \qquad (\,4.2A.1\,)$$

As a related current, we consider the input current of the element I_{vst} or the value of the current I_m in the maximum of the standing wave. Since currents I_{vst} and I_m of the same antenna (under identical conditions) usually differ, but the radiated power stays the same, even the values of the radiation impedance differ - the value of $Z_{\Sigma vst}$ related to the input current I_{vst} and the value of $Z_{\Sigma m}$ related to the current in its maximum I_m are not the same.

Neglecting losses of the antenna, the radiation impedance $Z_{\Sigma vst}$ equals to the input impedance of the antenna Z_{vst} computed as a ratio of input voltage U_{vst} and input current I_{vst}

$$U_{vst} = Z_{vst}I_{vst}. \qquad (\,4.2A.2\,)$$

The above-defined values of input impedance Z_{vst} are usually provided for basic types of antennas. These values are valid for antennas placed in *free space*, i.e. in satisfactorily long distance from other antennas or objects. A detailed algorithm of computing Z_{vst}, which is based on the radiated power, is given in the layer B.

In an antenna array, single antenna elements mutually influence themselves and impedance of each element depends on the type of elements in the surrounding, on the way of their positioning and feeding. In order to evaluate and compute the mutual influence, we change our view to the equation (4.2A.2).

The current I is assumed to be excited in the antenna any way. The current I is characterized by the current I_{vst} and by a function of current distribution. Antenna radiates and creates a given intensity of electric field E in its surrounding, which is proportional to the current on the antenna. A certain field intensity E_t appear even on the surface of the radiating antenna. At the same moment, antenna acts as the receiving antenna, and the receive results in a given voltage at the input port of the antenna. Voltage U_{vst} in eqn. (4.2A.2) can be therefore considered as a voltage, produced on the antenna by the receive of the own radiation. This own radiation is proportional to the current magnitude I_{vst} at its input and the quantity Z_{vst} (self-impedance) plays the role of the proportionality constant.

The above-described consideration can be simply applied to the whole antenna array. Even in the array, there is a certain field intensity E_t on the surface of each antenna element. This field is not created not only by the self-radiation of the antenna element but too by the radiation of the other elements.

In fig. **4.2A.1**, the antenna array consisting of n elements (dipoles), which are fed by currents $I_{vst\,i}$ on their input terminals, is depicted. In analogy to (4.2A.2), input voltages of antenna elements are given by the following set of equations

$$U_{vst1} = Z_{11}I_{vst1} + Z_{12}I_{vst2} + \ldots\ldots + Z_{1n}I_{vstn}$$
$$U_{vst2} = Z_{21}I_{vst1} + Z_{22}I_{vst2} + \ldots\ldots + Z_{2n}I_{vstn}$$
$$\vdots \qquad (\,4.2A.3\,)$$
$$U_{vstn} = Z_{n1}I_{vst1} + Z_{n2}I_{vst2} + \ldots\ldots + Z_{nn}I_{vstn}$$

Fig. 4.2A.1 *Array of n dipoles*

Impedance coefficients Z_{jk} (4.2A.3) express the mutual coupling between j-th and k-th element and are called mutual impedance. Mutual impedance is a complex quantity fulfilling $Z_{jk} = Z_{kj}$. Its magnitude depends on the shape, on the dimensions and on the mutual position of antenna elements, and even on their current distribution.

Coefficient Z_{jj} is called self-impedance and determines relation between the current and the voltage at the input of antenna elements out of the
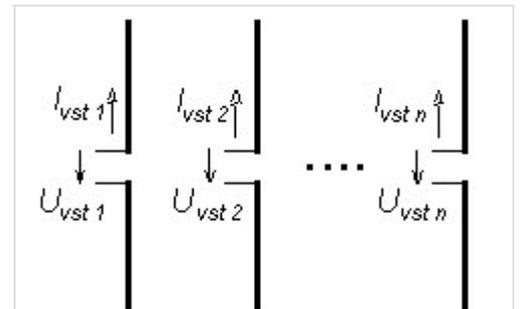
array and equals to the radiation impedance of a separated j-th antenna element in free space.

Typical dependency of the components of the mutual impedance $Z_{jk} = R_{jk} + jX_{jk}$ on the distance d (multiplied by wave-number $k = 2\pi/\lambda$) between two parallel dipoles of the same length is depicted in fig. **4.2A.2**.
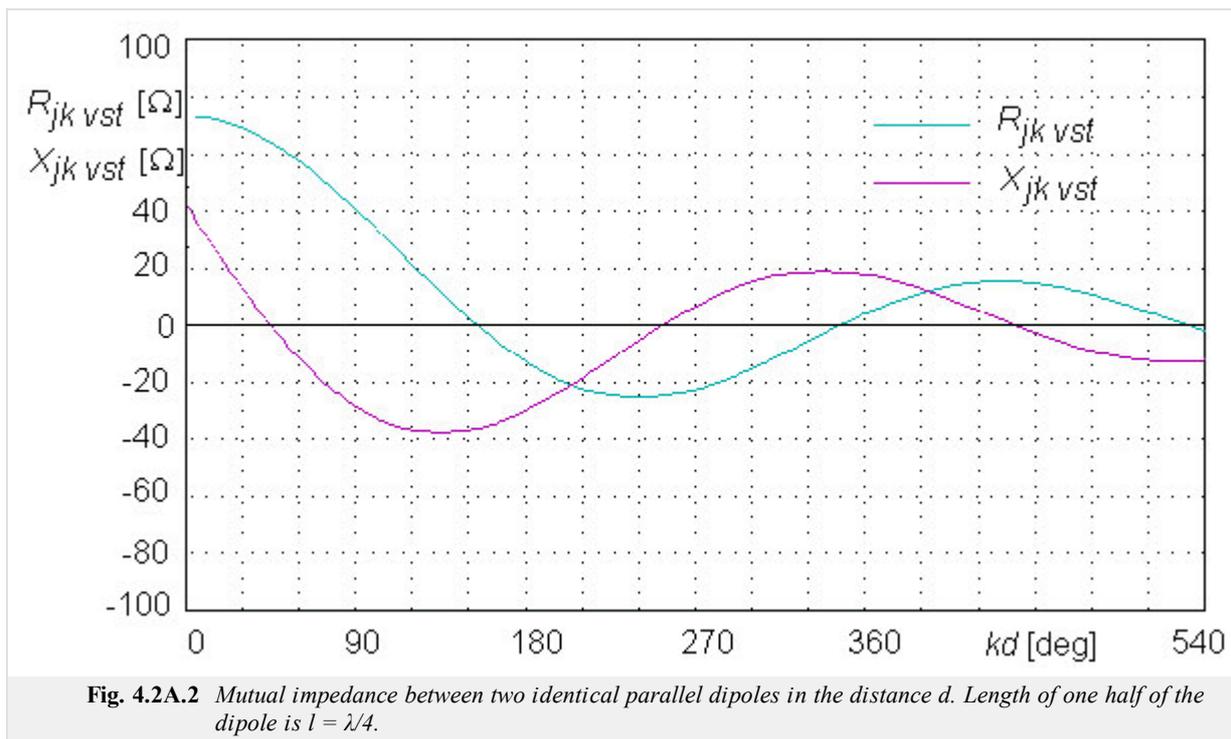

Fig. 4.2A.2 *Mutual impedance between two identical parallel dipoles in the distance d. Length of one half of the dipole is l = λ/4.*

Magnitude of mutual impedance depends on dimensions and on the distance of antenna elements. Real and imaginary component of mutual impedance can be both positive and negative and its maximum values decrease when the distance between elements d rises (the influence of very distant elements negligible).

Values of mutual impedance, which are provided for technical needs in charts, are usually related to the current in maximum. Re-computation to input terminals can be done considering the sine current distribution on the antenna

$$Z_{jkvst} = \frac{Z_{jkm}}{\sin^2(kl)}, \qquad (\ 4.2A.4\ )$$

where $kl = (2\pi/\lambda)\ l$, if antenna length l is not close to the integral multiple of $\lambda/2$.

More accurate values of mutual impedance can be computed using the program from the layer C. The program computes values of mutual impedance Z_{jk}, related to the maximum current or to the input one for a couple of two parallel dipoles of the same length and of the same standing wave on the wire of the dipoles. The length of the dipole l, the distance between dipoles d and the magnitude of the axial shift of the feeding ports of the dipoles h (fig. **4.2A.3**) have to be entered when multiplied by wave-number $k = 2\pi/\lambda$. Graphic representation of the magnitude of both components of mutual impedance R_{jk} and X_{jk} on the variation of a elected quantity (l, d or h) is suitable when the influence of the spatial arrangements of the antenna array to the impedance relations is investigated.
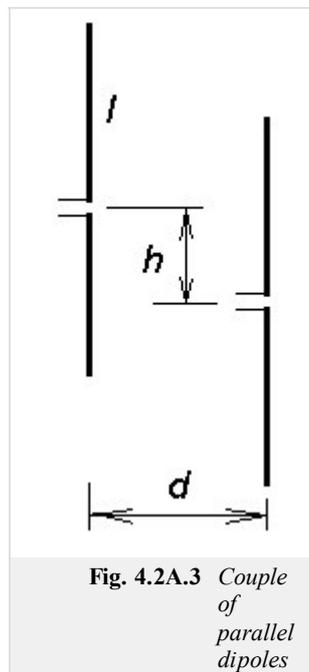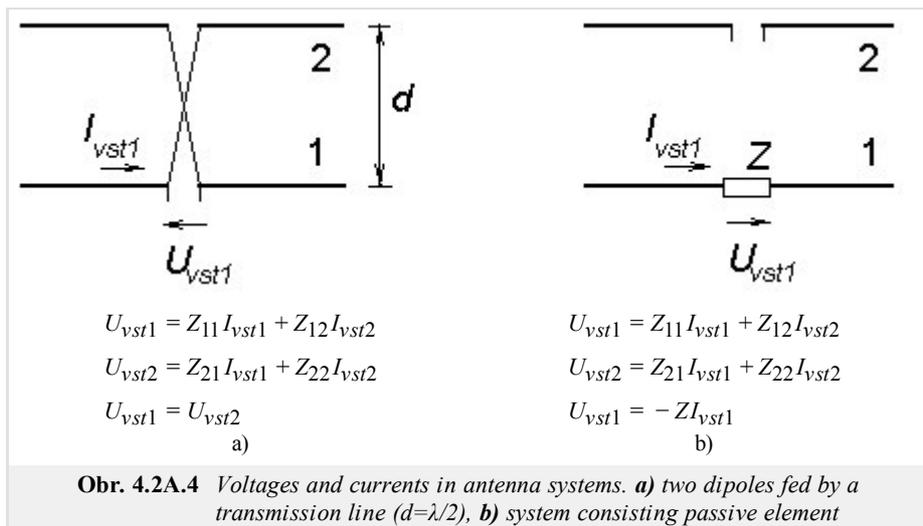
The same way, components of the radiation impedance Z_{ii} at the input of an isolated antenna element can be computed if *the distance between elements d* is put to be equal to the radius of the antenna wire a.

Fig. 4.2A.3 *Couple of parallel dipoles*

Computing voltages and currents in the system of n antenna elements, $2n$ independent quantities (n voltages and n currents) have to be determined. Considering the way of feeding, other (n-1) equations can be built. For elected magnitude of a single voltage (current), the ($2n$-1) unknown quantities can be computed. In antenna arrays, where single antenna elements are fed by the system of transmission lines, computations are complicated by the fact that impedance of each element depends on currents in other elements. That way, relations on transmission lines and radiators are mutually coupled. In case of single fed element, computations are rather simple.

Deriving equations (4.2A.3) for a given situation, we have to respect the fact that orientation of voltage $U_{vst\ i}$ and of a current $I_{vst\ i}$ correspond to the feeding of an element by a generator. In case of a passive element (voltage on the input port is caused by the current I_{vst} flowing through the load Z of the input port) is the orientation of the voltage opposite, and therefore, the sign has to be changed in (4.2A.3). For two typical

situations, corresponding sets of equations are given in fig. **4.2A.4**.

$$U_{vst1} = Z_{11}I_{vst1} + Z_{12}I_{vst2}$$
$$U_{vst2} = Z_{21}I_{vst1} + Z_{22}I_{vst2}$$
$$U_{vst1} = U_{vst2}$$
a)

$$U_{vst1} = Z_{11}I_{vst1} + Z_{12}I_{vst2}$$
$$U_{vst2} = Z_{21}I_{vst1} + Z_{22}I_{vst2}$$
$$U_{vst1} = -ZI_{vst1}$$
b)

Obr. 4.2A.4 *Voltages and currents in antenna systems. **a)** two dipoles fed by a transmission line (d=λ/2), **b)** system consisting passive element*

Set (4.2A.3) enables to compute the ratio between any couple of unknown quantities. Input impedance of *i*-th element $Z_{vst\,i}$ is obtained when dividing *i*-th eqn. (4.2A.3) by input current of *i*-th element $I_{vst\,i}$

$$Z_{vsti} = \frac{U_{vsti}}{I_{vsti}} = Z_{i1}\frac{I_{vst1}}{I_{vsti}} + Z_{i2}\frac{I_{vst2}}{I_{vsti}} + \dots\dots + Z_{ii} + \dots\dots + Z_{ni}\frac{I_{vstn}}{I_{vsti}}.$$
(4.2A.5)

Impedance of every element of an antenna array is therefore given by the sum of its own impedance Z_{ii} and of submissions of the other elements, depending on the product of mutual impedances Z_{jk} and respective currents (amplitudes and phases) flowing through the input ports of those elements. The change of feeding any antenna element therefore causes the change of impedances of all the antenna elements in the array. As an example, let us consider results obtained for the arrays depicted in fig. **4.2A.4**. In the case of the couple of dipoles fed by the crossed transmission line of the length $d = \lambda/2$ (fig. **4.2A.4a**), currents of the same magnitude and the same phase flow in both the dipoles, and both of the dipoles are of the input impedance

$$Z_{vst1} = Z_{vst2} = Z_{11} + Z_{12}\frac{I_{vst2}}{I_{vst1}} = Z_{11} + Z_{12}.$$
(4.2A.6)

Impedance on the port of the dipole "1" equals to the half of the value Z_{vst1}. For the couple of the half-wavelength-long dipoles ($kl = 90°$), we ge $Z_{11} = (73,1 + j42,5)$ Ω and $Z_{12} = (-12,5 + j30)$ Ω. Assuming in-phase feeding (fig. **4.2A.4a**), the resultant value of the impedance of both the dipoles equals to $Z_1 = (60,6 + j12,5)$ Ω. Considering anti-phase feeding, the resultant input impedance equals to the difference of Z_{11} and Z_{12}, which yields the result $Z_1 = (85,6 + j72,5)$ Ω. The same result is obtained for the dipole in the distance $\lambda/4$ from the planar reflector. Impedance on the port "1" in fig. **4.2A.4a** equals to the half of the computed value Z_1, i.e. $Z_{soust} = (30,3 + j6,2)$ Ω.

Input impedance of the dipole "2" in fig. **4.2A.4b** is given by the relation

$$Z_{vst2} = Z_{22} - \frac{z_{21}^2}{Z + Z_{11}}.$$
(4.2A.7)

Substituted values of both the impedances Z_{11} and Z_{12} have to be related to the input current of the dipole I_{vst} in both the cases.

4.2 Mutual impedance

Advanced theory

Radiation impedance $Z_{\Sigma vst}$ on the input port of the antenna can be computed from the value of complex power P_{Σ} which is radiated by the antenna, when dividing power by input current $|I_{vst}|^2$ of the antenna (eqn. 4.2A.1 in the layer A). Hence, turn our attention to computing the radiated power.

During the process of the radiation of electromagnetic waves, energy delivered to the lossless antenna in free space is transformed to the energy of electromagnetic wave, which propagates in free space, and to the energy, which is periodically exchanged between the antenna and the near field of the antenna. The radiated power P_{Σ} (complex, in general) can be computed integrating Poynting vector along the closed surface around the antenna. The obtained result is influenced by the selection of the region where the integration is performed.

In the far-field zone (radiation zone), electric field intensity E and magnetic field one H of the radiated electromagnetic wave are in phase. Poynting vector is purely real, and performing its integration along the closed surface (usually a sphere with the center in the antenna), a purely real power P_{Σ} is obtained. Dividing the power by squared input current I_{vst}, the real part of the radiation impedance $R_{\Sigma vst}$ is obtained.

Computing complex radiated power P_{Σ}, integration of Poynting vector has to be performed in immediate vicinity of the antenna (in case of linear antennas, we integrate along the surface of the antenna wire). Let us start with radiation of a symmetrical dipole of the length $2l$ and of the radius of antenna wire $2a$, which is placed to the coordinate system as depicted in fig. **4.2B.1**. On the surface of the cylindrical antenna wire, the current $I(z)$ excites a longitudinal component of electric field intensity $E_z(z)$ and a transversal tangential component of magnetic field intensity $H_\varphi(z)$.

Magnetic field intensity $H_\varphi(z)$ on the surface of a relatively thin wire is primarily determined by the current $I(z)$ in the immediate vicinity of the element dz. Applying Amper low, we get

$$H_\varphi(z) = \frac{I(z)}{2\pi a}. \qquad (4.2B.1)$$

Electric field intensity $E_z(z)$ on the surface of the antenna wire can be expressed in term of vector potential $A_z(z)$ considering a known current distribution $I(z)$

$$A_z(z) = \frac{\mu}{4\pi} \int_{-l}^{l} I(\zeta) \frac{e^{-jkr}}{r} d\zeta, \qquad (4.2B.2)$$

$$r = \sqrt{(z-\zeta)^2 + a^2}.$$

Fig. 4.2B.1 *Field intensities on the surface of antenna wire*

Electric field intensity $E_z(z)$ equals then to

$$E_z(z) = -j\omega\left(A_z + \frac{1}{k^2} grad_z div(A_z)\right) = -j\omega\left(A_z + \frac{1}{k^2}\frac{\partial^2 A_z}{\partial z^2}\right). \qquad (4.2B.3)$$

If the radius of the antenna wire a is significantly smaller than its length $2l$, and if radiation of the wire ends is neglected, Poynting vector is of the radial component only

$$\Pi = -E_z(z)H_\varphi^*(z). \qquad (4.2B.4)$$

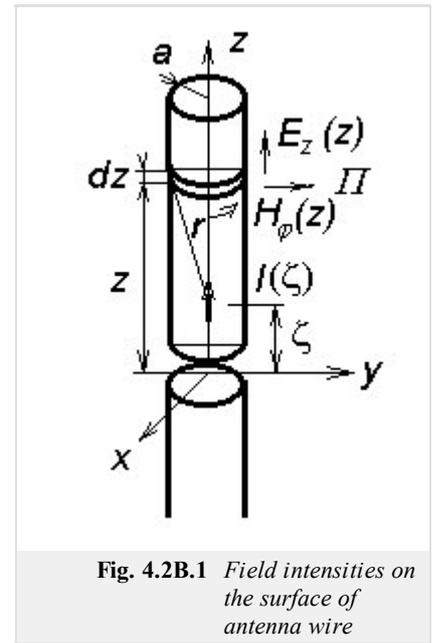An element of the wire surface $dS = 2\pi$ and dz radiates then the power

$$dP_{\Sigma} = \Pi 2\pi a\, dz = -E_z(z)I_z^*(z)\, dz, \qquad (4.2B.5)$$

where asterisk denotes complex conjunction.

Complex power radiated by the whole antenna P_{Σ} is obtained integrating along the whole dipole

$$P_{\Sigma} = \int_{-l}^{l} dP_{\Sigma} = -\int_{-l}^{l} E_z(z)I_z^*(z)\, dz. \qquad (4.2B.6)$$

Components of radiation impedance $Z_{\Sigma vst}$ on the input port of the antenna can be obtained when dividing the power P_{Σ} by squared input current I_{vst}

$$Z_{\Sigma vst} = -\frac{1}{|I_{vst}|^2} \int_{-l}^{l} E_z(z) I_1^*(z)\, dz. \qquad (\,4.2B.7\,)$$

The described approach is formally simple, but solving a given problem, evaluation of integrals is rather problematic. The main advantage of this approach is given by the possibility to compute impedances of elements in an antenna array as shown in the layer A.

Now, turn our attention to a small detail of the approach. We should disagree with the orientation of the vector Π in fig. **4.2B.1**, which does not correspond to the clockwise system **E** x **H**. Moreover, the cause of the sign "minus" in eqn. (4.2B.4) is not clear. Maybe, we may argue that the opposite orientation and the opposite sign can mutually compensate. Next, the reader can ask why antennas are fabricated from well-conductive metals: using perfect conductors, tangential component of electric field intensity vanishes, value of Π is zero and such a *perfect* antenna does not radiate. Hence, more detailed attention to the above-described approach is demanded.

Let us start with the zero tangential electric field component and assume perfect conductivity of the antenna wire. If our computations consider an *exact* current distribution then zero tangential electric field component is obtained everywhere on the surface of the antenna wire except of the infinitesimally short excitation gap at $z = 0$. Understanding Poynting vector as a surface power density, antenna seems to radiate in the excitation gap only.

Finite conductivity of real metals does not change anything on this fact. Tangential component on the surface is not indeed zero, but the real power penetrating *inside* the conductor is relatively small (this power is converted to heat in the conductor). The only region, where real component of Poynting vector is directed out of the antenna, is the feeding gap.

This conclusion seems to be rather surprising and might to invoke doubts. At least, this conclusion does not correspond to the notion of an antenna as a set of radiating elementary dipoles. Fortunately, these problems can be relatively simply solved.

First, nothing proves the fact that the antenna radiates in the feeding gap only. Poynting theorem says only that *integrating* the product $E_t(z)$ $H(z)$ along the *closed* surface equals to the total incoming power. Meaning of the surface power density is only *attributed* to this product because it can be *usually* done. Therefore, no contradiction with respect to the theory is revealed. Moreover, there is no need for asking the question where the power is radiated, and the so-far notions can be kept.

Even from the mathematical point of view, no serious problems appear. Due to the practical reasons, computations are not performed with an *exact* current distribution $I(z)$ but with the *approximate* one $I'(z)$ (of sine nature, usually). Even a small difference between $I(z)$ and $I'(z)$ causes the tangential component of electric field intensity to differ from zero on the surface of the antenna wire, and the integral (4.2B.6) can be evaluated. Computations can be therefore performed using an approximate current distribution $I'(z)$. Satisfactory correctness of the approach can be proven following A. A. Pistolkors:
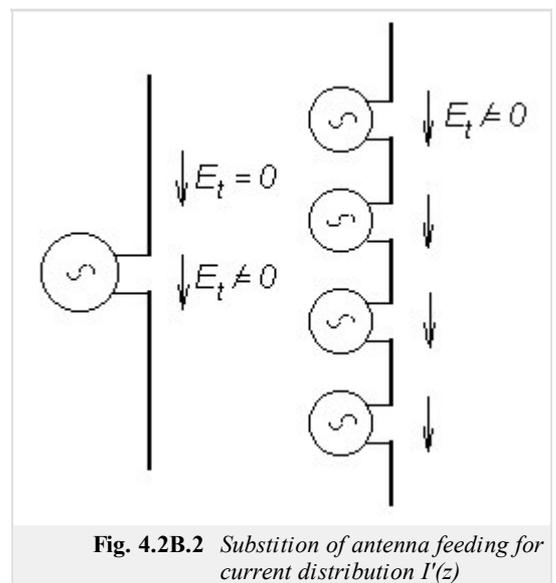
The real antenna (symmetric dipole) is depicted in fig. **4.2B.2a**. The antenna is fabricated from a perfect conductor, the current distribution is given by $I(z)$ and on the whole surface (except of the excitation gap) the tangential component of electric field intensity equals to zero. Since an approximate current distribution $I'(z)$ is considered in computations, E_t differing from zero is obtained, and consequently, a certain radiated power is computed.

In order to prove the computed power being approximately equal to the really radiated power, we can imagine to analyze an imaginary antenna of the same shape and the same dimensions as the real antenna but of $I'(z)$ as an *exact* current distribution. Then the imaginary antenna has to be fed by a different way. We have to admit the excitation consisting of a vertical row of elementary sources (fig. **4.2B.2b**), which voltages are set to enforce the current distribution following the function $I'(z)$. The fact that the tangential component of electric field intensity is non-zero on the surface of the imaginary antenna is correct because the whole surface of the imaginary antenna is covered by excitation gaps. The power we have computed is valid for our imaginary antenna. If the difference between $I(z)$ and $I'(z)$ is small, then the computed power approaches the power radiated by the real antenna.

Fig. 4.2B.2 *Substition of antenna feeding for current distribution I'(z)*

Finishing our discussion, we can go back to the basic idea.

In the set of parallel dipoles (fig. **4.2A.1** in the layer A), the power radiated by the first element of the array is expressed first. The total electric field intensity $E_{z1}(z)$ on the surface of the first element equals to the sum of contributions of all the elements of the array

$$E_{z1}(z) = E_{11}(z) + E_{12}(z) + \ldots\ldots + E_{1n}(z). \qquad (\,4.2B.8\,)$$

Power $P_{\Sigma 1}$ radiated by the first element is obtained when substituting to eqn. (4.2B.6). Formal rearrangement yields the relation

$$P_{\Sigma 1} = - \int_{-l}^{l} \left[E_{11}(z) I_1^*(z) + E_{12}(z) I_1^*(z) + \ldots\ldots + E_{1n}(z) I_1^*(z) \right] dz =$$

$$= I_{vst1}^* \left[-I_{vst1} \int_{-l}^{l} \frac{E_{11}(z)}{I_{vst1}} \frac{I_1^*(z)}{I_{1vst}^*} dz - I_{vst2} \int_{-l}^{l} \frac{E_{12}(z)}{I_{vst2}} \frac{I_1^*(z)}{I_{1vst}^*} dz - \ldots\ldots - I_{vstn} \int_{-l}^{l} \frac{E_{1n}(z)}{I_{vstn}} \frac{I_1^*(z)}{I_{1vst}^*} dz \right] \qquad (\ 4.2B.9\ )$$

.

Integrals

$$Z_{jk} = \int_{-l}^{l} \frac{E_{jk}(z)}{I_{vstk}} \frac{I_j^*(z)}{I_{vstj}^*} \qquad (\ 4.2B.10\ )$$

are called mutual impedances and are given in Ohms. Thanks to the formal rearrangement, integrals (4.2B.10) do not depend on the magnitudes of currents in antenna elements because the field intensity E_{jk} excited on the surface of j-th element by the radiation of k-th element is proportional to the current $I_{vst\,k}$ in k-th element of the antenna array.

For parallel dipoles, results of the integration in (4.2B.10) are expressed in the form of tabulated functions (integral sine, integral cosine). In order to get numerical values of mutual impedance components Z_{jk}, computer program described in the layer C can be used. Results describe coupling between the couple of parallel dipoles of the same length $2l$, which is related to the maximal current $I_{\max}$. Values $Z_{jk\ vst}$, which are related to the input current of the dipole, are obtained performing re-computing according to eqn. (4.2A.4) in the layer A.

Dependency of the magnitude of the components of mutual impedance on the distance between dipoles was already demonstrated in the layer A (fig. **4.2A.2**) for the couple of parallel symmetric dipoles with a standing current way. Negative values of the mutual impedances R_{jk} and X_{jk} express the fact that for a given spatial arrangement, radiation of k-th element reduces the magnitude of real or imaginary part of the power $P_{\Sigma j}$ which is radiated by j-th element of the antenna array. Decreasing magnitude of mutual impedance for long distances corresponds to the decrease of the electric field intensity E_{jk} for increasing distance.

Exploiting eqn. (4.2B.10), self-impedance of a stand-alone antenna element Z_{jj} can be computed if the distance between elements d is set to the value of the radius of the antenna wire a. This approach corresponds to the situation described when deriving eqn. (4.2B.7): we considered electric field intensity on he surface of j-th antenna wire E_{jj} excited by a current flowing through the axis of the same wire.

The method of induced electromotoric forces enables to compute the radiated power, and consequently the values of mutual impedances even in the case of non-parallel antenna wires. Then, we have to respect a different spatial orientation of vectors of electric field intensity E_{jk} with respect to the surface of j-th antenna wire. Decomposing vector E_{jk} to respective components and exploiting more general expression for the power contribution dP_{Σ}, we obtain formally more complicated relations. That way, coupling between crossed wires of a turnstile antenna and further situations can be solved.

4.2 Mutual impedance

Matlab program

The program computes real and imaginary part of the mutual impedance of antennas depending on the length of the dipole *l*, on the shift between the centers of the dipoles *h* and on the distance of the dipole axes *d*. Parameters are entered to the program multiplied by the wave-number, i.e. as *kl*, *kh* and *kd*.

First the path of Matlab has to be redirected to the folder **Impedance**. The program is run calling the m-file `vzimp.m`. Leaving the introductory window, which brings the basic information about the program and which serves for the selection of a chart to be displayed (which variable is going to vary), values of *kl*, *kh* and *kd* have to be entered. The type of entered parameters depends on the setting in the introductory window (which chart is going to be displayed and which variable is going to vary). For the varying variable, its minimum value and its maximum one have to be given. Pushing the button Compute (if one of values is swept), a new window is opened, where the course of the real part and the imaginary one of the mutual impedance is depicted with respect to the selected varying parameter.

The button Exit serves for finishing the program and the button Back can be used for a new run of the program.

4.2 Mutual impedance

Developing Matlab

Following relations can be derived for computing mutual impedance Z_{12} of two dipoles. Real part of the mutual impedance R_{12} is given by:

$$R_{12} = 15\left\{(K_1 \sin q + L_1 \cos q) + [K_2 \sin(q+2p) + L_2 \cos(q+2p)] + [K_3 \sin(q-2p) + L_3 \cos(q-2p)]\right\}. \qquad (4.2\text{D}.1)$$

Imaginary part of the mutual impedance X_{12} can be expressed as:

$$X_{12} = 15\left\{(M_1 \sin q + N_1 \cos q) + [M_2 \sin(q+2p) + N_2 \cos(q+2p)] + [M_3 \sin(q-2p) + N_3 \cos(q-2p)]\right\}, \qquad (4.2\text{D}.2)$$

where

$$
\left.
\begin{aligned}
K_1 &= 2[2f_2(\delta, q) - f_2(\delta, q+p) - f_2(\delta, q-p)] \\
L_1 &= 2[2f_3(\delta, q) - f_3(\delta, q+p) - f_3(\delta, q-p)] \\
K_2 &= f_2(\delta, q) - 2f_2(\delta, q+p) + f_2(\delta, q+2p) \\
L_2 &= f_3(\delta, q) - 2f_3(\delta, q+p) + f_3(\delta, q+2p) \\
K_3 &= f_2(\delta, q) - 2f_2(\delta, q-p) + f_2(\delta, q-2p) \\
L_3 &= f_3(\delta, q) - 2f_3(\delta, q-p) + f_3(\delta, q-2p) \\
M_1 &= 2[2f_4(\delta, q) - f_4(\delta, q+p) - f_4(\delta, q-p)] \\
N_1 &= -2[2f_1(\delta, q) - f_1(\delta, q+p) - f_1(\delta, q-p)] \\
M_2 &= f_4(\delta, q) - 2f_4(\delta, q+p) + f_4(\delta, q+2p) \\
N_2 &= -f_1(\delta, q) + 2f_1(\delta, q+p) - f_4(\delta, q+2p) \\
M_3 &= f_4(\delta, q) - 2f_4(\delta, q-p) + f_4(\delta, q-2p) \\
N_3 &= -f_1(\delta, q) + 2f_1(\delta, q-p) - f_3(\delta, q-2p)
\end{aligned}
\right\}. \qquad (4.2\text{D}.3)
$$

In these relations, variables given by the user p, q, δ are present:

$$p = kl \qquad q = kh \qquad \delta = kd. \qquad (4.2\text{D}.4)$$

Finally, we have to define the following functions:

$$
\left.
\begin{aligned}
f_1(\delta, u) &= \sin\left(\sqrt{u^2 + \delta^2} + u\right) + \sin\left(\sqrt{u^2 + \delta^2} - u\right) \\
f_2(\delta, u) &= \sin\left(\sqrt{u^2 + \delta^2} + u\right) - \sin\left(\sqrt{u^2 + \delta^2} - u\right) \\
f_3(\delta, u) &= \cos\left(\sqrt{u^2 + \delta^2} + u\right) + \cos\left(\sqrt{u^2 + \delta^2} - u\right) \\
f_4(\delta, u) &= \cos\left(\sqrt{u^2 + \delta^2} + u\right) - \cos\left(\sqrt{u^2 + \delta^2} - u\right)
\end{aligned}
\right\}. \qquad (4.2\text{D}.5)
$$

Here, u is a parameter, which changes according to the variable, which is used in our relations.

The above-given relations are programmed. We do not present here the transcription into the Matlab syntax due to the worse readability. In the program one of the parameters kl, kh or kd vary.

4.2 Mutual impedance

Java applet

4.2 Mutual impedance

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

Mutual impedance between two dipoles …

Possible answers for question 1:

○ … is the difference between input impedances of those two dipoles.

○ … quantifies the contribution of the current on the first dipole to the voltage induced on the second dipole.

○ … can be neglected in practical computations.

◉ I won't answer this question

Question 2

Mutual impedance between two dipoles depends on …

Possible answers for question 2:

○ … the distance of those two dipoles and the wavelength.

○ … the distance, but is independent on the wavelength.

○ … the distance and the input impedance of the dipoles.

◉ I won't answer this question

Question 3

Input impedance of the dipole in the array is influenced by …

Possible answers for question 3:

○ … input impedances of neighboring dipoles and their currents.

○ … mutual impedances of neighboring dipoles and their currents.

○ … the linear combination if input impedances and mutual ones.

◉ I won't answer this question

Question 4

Mutual impedance between the first dipole and the second one …

Possible answers for question 4:

○ … can equal (under special conditions) to the mutual impedance between the second dipole and the first one.

○ … can never equal to the mutual impedance between the second dipole and the first one.

○ … equals any time to the mutual impedance between the second dipole and the first one.

◉ I won't answer this question

Question 5

The method of electromagnetic forces can be applied …

Possible answers for question 5:

- ◯ … when dipoles are parallel.
- ◯ … to an arbitrary array consisting of arbitrary antennas.
- ◯ … to the array of arbitrarily oriented dipoles.
- ⦿ I won't answer this question

see result

4.3 Yagi antenna

Basic theory

In this paragraph, the knowledge from the paragraph 4.1 is going to be applied and extended to modeling Yagi antenna. First, a brief description of the antenna is given. Second, the principle of its operation is explained, and third; results of the paragraph 4.1 are exploited in order to build a numerical model of the antenna.

Yagi antenna [1] is a row of dipoles exhibiting longitudinal radiation. A single element (active one) is fed in galvanic way and the rest of elements are excited by the radiation (passive elements). In passive elements, the wave, which is radiated by the active element, induces currents those regressively influence the radiation pattern of the whole antenna.

Thanks to the construction simplicity and good parameters, Yagi became one of the most popular antennas in the meter-wave band and decimeter-wave one.

As already noted, Yagi antenna consists of a single active element and several passive ones. As an active element, a symmetric dipole (or a folded one), which works in one-quarter-wavelength resonance, is used. Passive elements are constructed as dipoles without the excitation gap. A single passive element, which is longer than the active one, is positioned at the back and plays the role of a reflector. The rest of passive elements, which are shorter than the active one, are placed in the front and are called directors (fig. **4.3A.1**).

Depending on the number of directors, the gain of the antenna is between 10 and 15 dB. Lengths of passive elements with respect to the length of the active one are not chosen randomly but their proper value conditions a proper functionality of the antenna.

Fig. 4.3A.1 *Yagi antenna - description and example of realization*
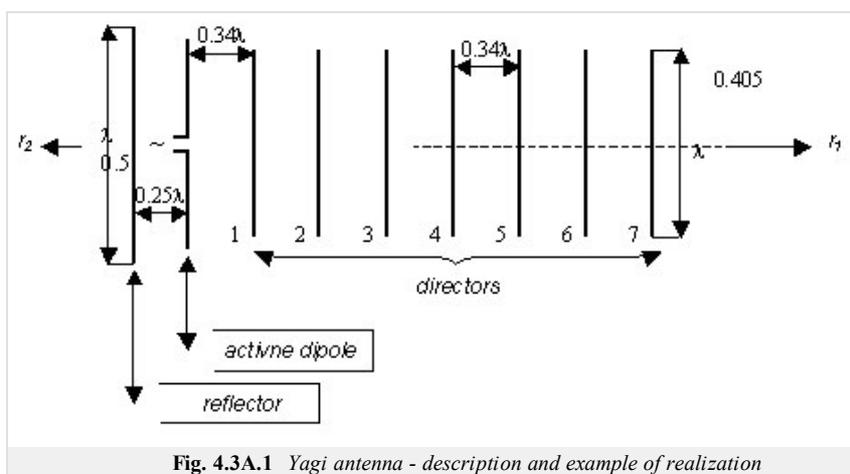
In fig. **4.3A.1**, the schematic of Yagi antenna is given. The antenna consists of a symmetric dipole as an active element, of a reflector and of seven directors. If the antenna is properly tuned [6], the reflector attenuates the energy flow in the direction r_2, and the first director amplifies the energy flow in the direction r_1. At the same moment, amplification of the energy flow in the direction r_1 creates conditions for the excitation of the second director. The second director subsequently amplifies the energy flow in the direction r_1 and creates conditions for the excitation of the third director, etc.

We can therefore say that directors form a *wave-guiding channel* in a fact. Energy distributed in this wave-guiding channel travels from the active element to the director in the largest distance, and there, the wave is partially reflected. As a result, the total field in the wave-guiding channel is given by the superposition of the forward wave and the backward one. From that point of view, Yagi antenna belongs to the traveling wave antennas.

If directors are properly tuned, amplitude of the reflected wave at the last director is very small, and Yagi antenna behaves as a traveling wave antenna, which is matched at its end. Nevertheless, Yagi antenna differs from traveling wave antennas in principle because the current distribution need not be the superposition of a single forward wave and a single backward one. In general, the current distribution of Yagi antenna can be described by the sum of several forward waves and backward ones, which are of different propagation constants.

If the radiated wave should be attenuated in the direction r_2, the reflector of Yagi antenna has to be designed that way [6] so that the phase of induced current anticipates the current phase in the active dipole. Moreover, field intensities of the active dipole and of the reflector have to be of such phases in r_1, so that the phase of the current induced in the first director can be retarded with respect to the active dipole (so that the energy flow in r_1 can be amplified). Similarly, phase of the current induced in the second director has to be retarded with respect to the first director, etc. The antenna has to be tuned so that all the above-described conditions are fulfilled.

An elementary analysis, which is based on the method of induced electromotoric forces, shows [6], that the total radiation impedance of the reflector (related to the maximal current) has to be of positive, *inductive* reactance and that the total radiation impedance of each director (related to the maximal current) has to be of negative, *capacitive* reactance.

The polarity of the reactance of elements of Yagi antenna can be influenced changing their length. Positive reactance is reached by increasing the length with respect to resonant one, and vice versa [1]. Speaking about the resonant length, one-quarter-wavelength resonance is assumed.

Since radiation impedance of the dipole depends on its radius, even the magnitude of shortening or extending antenna elements depends on this radius. Moreover, resultant values of reactance of the reflector and the directors depend on the distances among antenna elements and on the number of directors [6].

In fig. **4.3A.1**, concrete parameters of Yagi antenna are given [6] when symetric dipole is used as an active element. The length of the reflector is 0.5λ, the length of the directors is 0.405λ, the length of the active element corresponds to the resonant length; is a bit shorter than 0.5λ. Smaller the radiation impedance of the active dipole is, shorter its resonant length (with respect to 0.5λ) is. Diameter of all the dipoles equals to 0.002λ. The distance between the active element and the

reflector equals to 0.25λ, the distance between the active element and the first director is 0.34λ, and the distance among directors equals to 0.34λ.

Yagi antenna is very popular these days. There are many technical realizations of this antenna, which differ in the number of directors, distances among antenna elements and lengths of those elements. In most cases, the number of elements, distances and lengths are determined experimentally. Nevertheless, today's computers and advanced numerical techniques enable to develop efficient computer programs for modeling and optimizing Yagi antennas on a PC. This topic is going to be investigated in the next paragraphs. In order to fully understand the explanation, the theory presented in the paragraph 4.1 has to be studied.

Before starting the numerical analysis of Yagi antenna, several simplifications are introduced [7]:

1. Antenna is situated in a lossless medium.

2. Antenna elements are fabricated from the perfect electric conductor.

3. Currents and charges are concentrated on the axes of antenna wires.

4. Each antenna element (reflector, dipole, directors) is divided into segments of the same length.

5. Current distribution on antenna elements is approximated using piecewise constant functions.

As a consequence of the assumption 2, the tangential component of electric field intensity has to be zero on each antenna segment except of the excitation gap. Obviously, field on the surface of each antenna segment is influenced not only by currents and charges of the respective antenna element but too by currents and charges of other antenna elements. This it the main difference compared to a simple dipole analysis in the paragraph 4.1.

Nevertheless, the main steps of building the impedance matrix of Yagi antenna are done in analogy with the wire antenna approach. Therefore, modifications of the computational algorithm are described here only.

Computing impedance matrix of the stand-alone active antenna element, relations described in paragraph 4.1 can be directly applied. In case of two different antenna elements, the mutual impedance of segments is computed a similar way; moreover the mutual distance of antenna elements has to be included to the computation of the distance between two respective segments.

Using matrices of mutual impedances and second Kirchhoff low, the following set of linear matrix equations is obtained for Yagi antenna:

Reflector:	$[0] = [Z_{rr}][I_r] + [Z_{rd}][I_d] + [Z_{r1}][I_1] + [Z_{r2}][I_2] + \ldots + [Z_{rn}][I_n],$	
Active dipole:	$[U_0] = [Z_{dr}][I_r] + [Z_{dd}][I_d] + [Z_{d1}][I_1] + [Z_{d2}][I_2] + \ldots + [Z_{dn}][I_n],$	(4.3A.1)
First director:	$[0] = [Z_{1r}][I_r] + [Z_{1d}][I_d] + [Z_{11}][I_1] + [Z_{12}][I_2] + \ldots + [Z_{1n}][I_n],$	
The n-th director:	$[0] = [Z_{nr}][I_r] + [Z_{nd}][I_d] + [Z_{n1}][I_1] + [Z_{n2}][I_2] + \ldots + [Z_{nn}][I_n],$	

here, $[I_r]$ and $[I_d]$ are vectors of currents in the reflector and in the active dipole, $[I_1]$, $[I_2]$, ..., $[I_n]$ are vectors of currents in the first director, in the second one, up to the n-th one. Next, $[Z_{rr}]$, $[Z_{dd}]$, $[Z_{11}]$, $[Z_{22}]$, ... , $[Z_{nn}]$ are impedance matrices containing mutual impedances of segments belonging to a given element (reflector, active element, first director, second director up to the n-th director), $[Z_{rd}]$ denotes the matrix of mutual impedances between segments of the reflector and the active dipole, $[Z_{d1}]$, ..., $[Z_{dn}]$ are matrices of mutual impedances between segments of the active dipole and directors, $[Z_{r1}]$, ..., $[Z_{rn}]$ denote matrices of mutual impedances between segments of the reflector and directors, $[Z_{ik}]$ is matrix of mutual impedances between segments of i-th and k-th director, $[0]$ is zero vector, the vector $[U_0]$ contains a single non-zero element corresponding to the voltage in the excitation gap of the active dipole.

Considering the principle of the reciprocity, following relations can be obtained:

$$[Z_{rd}] = [Z_{dr}]^{\mathrm{T}},\ [Z_{dn}] = [Z_{nd}]^{\mathrm{T}},\ [Z_{rn}] = [Z_{nr}]^{\mathrm{T}},\ [Z_{ik}] = [Z_{ki}]^{\mathrm{T}},\ [Z_{rr}] = [Z_{rr}]^{\mathrm{T}},\ [Z_{dd}] = [Z_{dd}]^{\mathrm{T}},\ [Z_{11}] = [Z_{11}]^{\mathrm{T}},\ \ldots,\ [Z_{nn}] = [Z_{nn}]^{\mathrm{T}}$$

$$[Z_{dd}] = [Z_{dd}]^{\mathrm{T}},\ [Z_{11}] = [Z_{11}]^{\mathrm{T}},\ \ldots,\ [Z_{nn}] = [Z_{nn}]^{\mathrm{T}},$$

Here, $^{\mathrm{T}}$ denotes the matrix transposition.

The size of vectors of current distribution on respective antenna elements $[I_r]$, $[I_d]$, $[I_1]$, ..., $[I_n]$ equals to the number of segments, the element is divided for. The matrix of self-impedances of antenna elements $[Z_{rr}]$, $[Z_{dd}]$, $[Z_{11}]$, ..., $[Z_{nn}]$ are square matrices, which size corresponds to the number of segments again. Matrices $[Z_{rd}]$, $[Z_{dn}]$, $[Z_{rn}]$ and $[Z_{ik}]$ can be in general rectangular and are of the size R x D, D x N and R x N, where R is the number of segments of the reflector, D is the number of segments of the active dipole and N is the number of segments of the n-th director.

The set of equations (4.3A.1) can be rewritten to a more compact form:

$$[Z][I] = [U]$$

(4.3A.2)

where

$$[Z] = \begin{bmatrix} [Z_{rr}] & [Z_{rd}] & [Z_{r1}] & \ldots & [Z_{rn}] \\ [Z_{dr}] & & & & \\ [Z_{1r}] & & & & \\ & & & & \\ [Z_{nr}] & [Z_{nd}] & [Z_{n1}] & & [Z_{nn}] \end{bmatrix}, \quad [I] = \begin{bmatrix} [I_r] \\ [I_d] \\ [I_1] \\ \\ [I_n] \end{bmatrix}, \quad [U] = \begin{bmatrix} [0] \\ [U_0] \\ [0] \\ \\ [0] \end{bmatrix}$$

(4.3A.3)

The matrix $[Z]$ is symmetric, i.e. $[Z] = [Z]^T$.

Solving eqn. (4.3A.2), current distribution on every antenna element can be computed

$$[I] = [Z]^{-1}[U]$$ (4.3A.4)

where $[Z]^{-1}$ denotes the inverse of the matrix $[Z]$.

Input impedance of the antenna is computed as

$$Z_{vst} = \frac{1}{[Z]^{-1}_{feed, feed}} = \frac{1}{[Y]_{feed, feed}}$$ (4.3A.5)

where *feed* is the index of the element of the inverted (admitance) matrix, which corresponds to the excitation segment of the active element.

If the current distribution on antenna elements is known, directivity pattern can be computed. In our computations, only the plane *E* is considered (the plane identical with fig. **4.3A.1**; the plane *H* is perpendicular to the Figure and to antenna elements as well). Since each discretization segment of the antenna can be considered as an elementary dipole, the following relation can describe electric field intensity in the surrounding of the antenna:

$$E_{\vartheta} = 60 j \frac{k}{2} \sin\Theta \frac{\exp(-jkr)}{r} \sum_{n=1}^{M} I_n \exp(-jkr\Delta r_n)$$ (4.3A.6)

Fig. 4.3A.2 *Principal scheme for computing phase shifts among segments*

Here, θ is an angle measured from the dipole axis, M is the total number of segments of the whole antenna, j denotes imaginary unit, k is wave-number, and Δr_n denotes spatial differences of waves radiated by different elementary dipoles (segments). The rest of symbols is explained in fig. **4.3A.2**.

The directivity pattern of the antenna can be expressed considering (4.3A.6) as

$$F(\Theta) = j \frac{k}{2} \sin\Theta \sum_{n=1}^{M} I_n \exp(-jk\Delta r_n)$$ (4.3A.7)

Finally, spatial differences of waves Δr_n have to be expressed for all the segments of the antenna. As a reference, the first segment of the reflector is elected (see fig. **4.3A.2**). Then, spatial difference are described by the following equations:

$$
\begin{aligned}
1^{st} \text{ segment:} \quad & \Delta r_1 = 0 \\
2^{nd} \text{ segment:} \quad & \Delta r_2 = \Delta l \cos\Theta \\
3^{rd} \text{ segment:} \quad & \Delta r_3 = 2\Delta l \cos\Theta \\
4^{st} \text{ segment:} \quad & \Delta r_4 = 3\Delta l \cos\Theta \\
5^{st} \text{ segment:} \quad & \Delta r_5 = (x_1 - x_2)\sin\Theta + \Delta l \cos\Theta \\
6^{st} \text{ segment:} \quad & \Delta r_6 = (x_1 - x_2)\sin\Theta + 2\Delta l \cos\Theta \\
& \text{etc.}
\end{aligned}
$$ (4.3A.8)

On the basis of the above-given description, a Matlab program for the analysis of Yagi antenna was developed (for more information, see the layer C and the layer D). Functionality of the program is verified by analyzing a three-element Yagi antenna (reflector, active dipole and director). Following parameters are considered:

- Wavelength: 0.680 m
- Length of reflector: 0.408 m
- Radius of antenna wires: 0.001 m
- Distance reflector - dipole: 0.100 m
- Distance dipole - director: 0.150 m
- Length of dipole: 0.357 m
- Length of director: 0.306 m
- Reflector: 15 identical segments
- Dipole 13 identical segments
- Director 11 identical segments

Fig. 4.3A.3 *Current distribution on dipole*

Fig. **4.3A.3** to **4.3A.5** show the current distribution on the reflector, on the dipole and on the director. Fig. **4.3A.6** to **4.3A.7** show the directivity pattern of the antenna both in the Cartesian coordinates and in the polar ones. In patterns, the angle β is related to the angle θ as depicted in fig. **4.3A.2**: $\theta = 360° - \beta$.
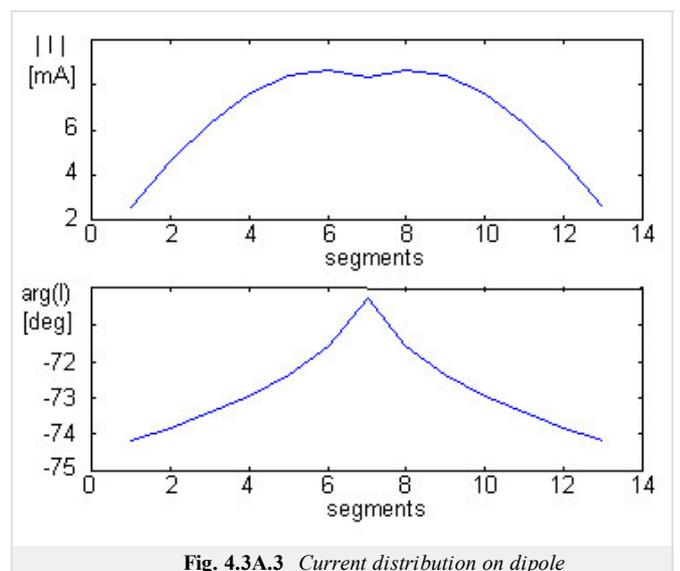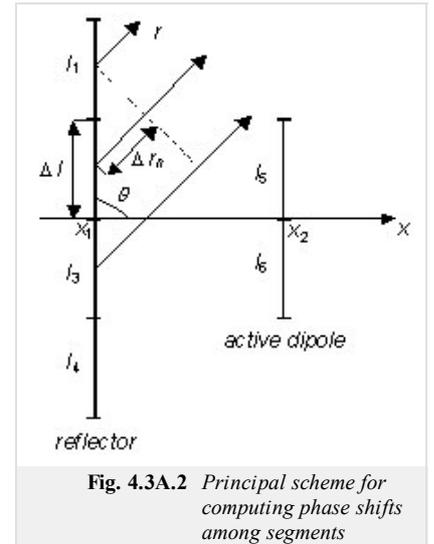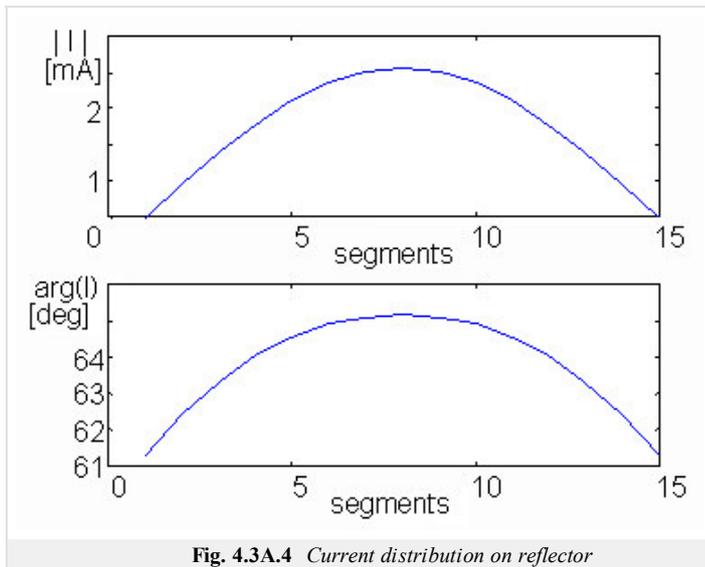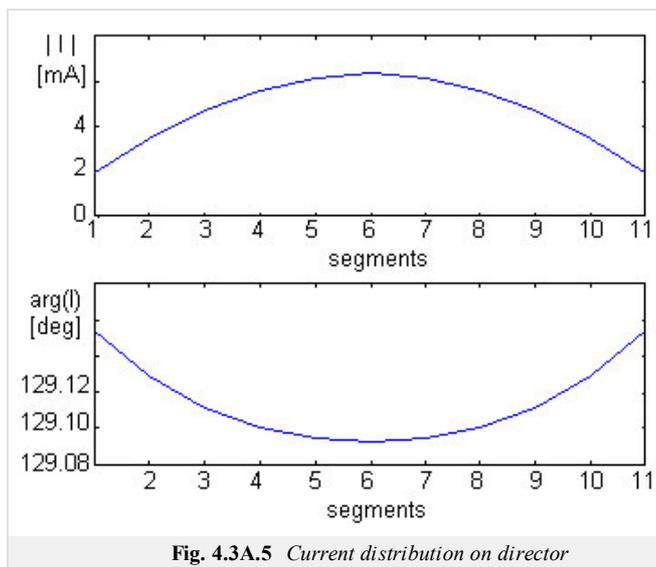
Fig. 4.3A.4 *Current distribution on reflector*

Fig. 4.3A.5 *Current distribution on director*
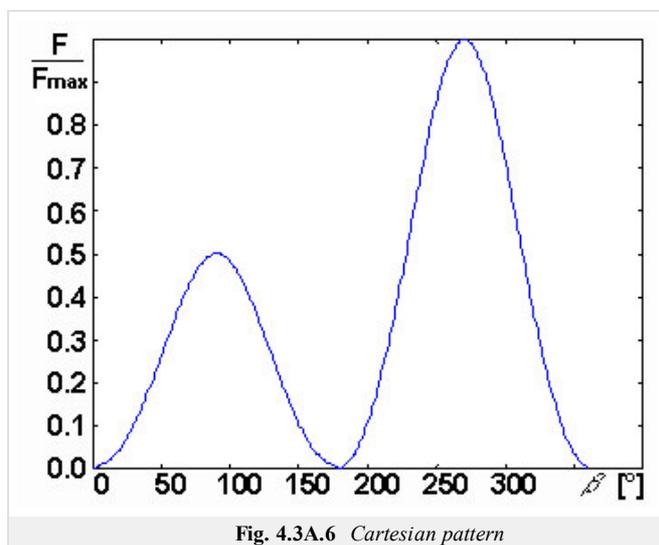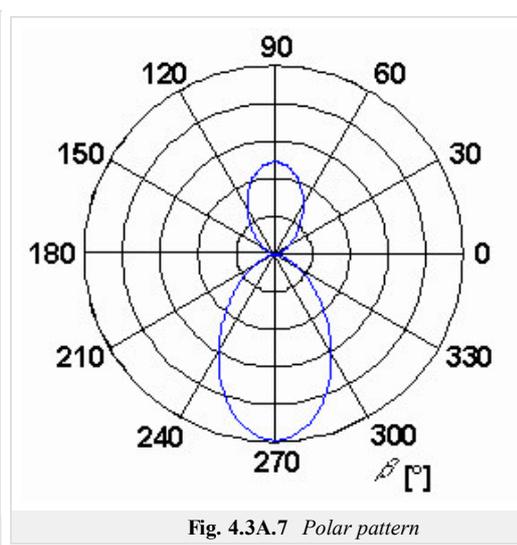

Fig. 4.3A.6 *Cartesian pattern*

Fig. 4.3A.7 *Polar pattern*

Directivity patterns (fig. **4.3A.6** and **4.3A.7**) show that the antenna radiation is much stronger in the direction from the dipole to the directors ($\beta = 270°$) than in the opposite direction. Extending the antenna for additional two directors can create even more dominant main beam. The rest of the parameters stay the same except of:

- The distance of the first director from the dipole is 0.107 m.
- The second director is in the distance 0.150 m from the dipole.
- The third director is in the distance 0.272 m from the dipole.

Fig. **4.3A.8** shows the final directivity pattern of the antenna in polar coordinates.

We can conclude that there are more ways of analyzing Yagi antenna (e.g., sine current distribution is a priori assumed on each antenna element, and consequently, input impedance and directivity pattern are computed **[6]**). Obviously, the moment method is able to reach better accuracy. These days, when powerful computers are at our disposal, the moment method is highly preferred.
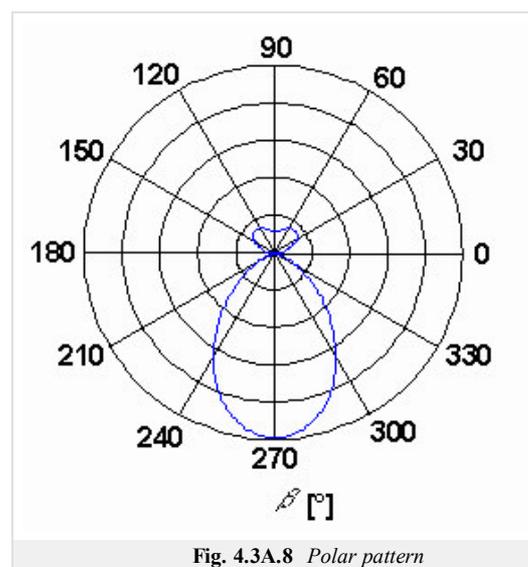

Fig. 4.3A.8 *Polar pattern*

4.3 Yagi antenna

Matlab program

In this layer, we present a user's guide to the program for the analysis of Yagi antenna (the programmer's guide can be found in the layer D). The program is started by entering `Yagi_Gui_Master` to the command window of Matlab (path is set to folder **Yagi**). Menu items of the main form are of the following meaning:

Wavelength	wavelength of the supplying voltage;
Radius	radius of the wire of antenna elements;
Reflector length	length of the whole reflector;
Reflector-dipole	distance between the reflector and the active dipole;
Number of elements	the number of segments to which the reflector is going to be divided (odd number required);
Difference 1	the number of segments for which the active element is shorter than the reflector;
Difference 2	the number of segments for which the directors are shorter than the reflector;
Input element	index of the feeding segment of the active dipole in the complete current vector (an information item, which cannot be edited);
Number of elements	the total number of antenna elements;
Dipole-director	the distance between the active dipole and the nearest director;
Director-director	the distance between two neighboring directors;
Dipole length	the length of the active dipole (an information item, which cannot be edited);
Director lengths	the length of all the directors (an information item, which cannot be edited).

Pressing the button $\boxed{\text{OK}}$ on the main form, the computation is started. Respective charts (current distribution, directivity pattern) are sequentially displayed in a single form after pressing the key $\boxed{\text{Enter}}$.

Pressing the button $\boxed{\text{Apply}}$ on the main form, information items display values of real lengths (recomputed from the number of segments) of the active dipole and directors, and the value of the index of the excitation segment. If needed, the number of segments of antenna elements can be changed before starting the computations.

Pressing the button $\boxed{\text{Cancel}}$ on the main form, the program is terminated.

When the computations are finished, the actual value of input impedance is printed in the header of the main form. In charts of directivity patterns, the angle 270° corresponds to the direction from the dipole to the directors, the angle 90° corresponds to the direction from the dipole to the reflector.

Since the graphic interface of the program does not support the setting of individual distances among directors, a more general function `Yagi_Antenna` has to be used as a *stand-alone program*. for this purpose. When this m-function is started, parameters `N` and `feed` have to be given (their meaning is explained in the layer D). The m-function returns the input impedance of the antenna. Further parameters can be changed in the source code of the function. Individual distances among antenna elements are set indirectly changing their coordinates with respect to the active dipole (the positive orientation goes from the dipole to the reflector). Respective charts (current distribution, directivity pattern) are sequentially displayed in a single form after pressing the key $\boxed{\text{Enter}}$.

4.3 Yagi antenna

Developing Matlab

In this paragraph, the moment analysis of Yagi antenna is described from the programmer's point of view. The description assumes the knowledge of the description presented in the paragraph 4.1 and in the layer A of this chapter.

The program consists of files `Yagi_Gui_Master.mat` and `Yagi_Gui_Master.m` , which create the form of the user's interface, of `Yagi_Antena_OK_Master.m`, which is an reaction to pressing the button OK in the main form, of `Yagi_Antena_Apply_Master.m`, which is an reaction to pressing Apply and of `Yagi_Antenna_Slave.m`, which is responsible for all the computations.

First, we explain the meaning of symbols appearing in the source code:
    `delta` is the segment length,
    `alfa` is one half of the segment length,
    `k` denotes wave-number,
    `M` is the number of antenna elements,
    `N` is the number of segments to which the reflector is divided,
    `epsilon` denotes permittivity of vacuum,
    `omega` denotes angular frequency,
    `j` is imaginary unit,
    `Z` denotes impedance matrix,
    `Y` is admittance matrix,
    `I` is the vector of currents,
    `feed` is the index of the excitation segment,
    `diff(1)` is the number of segments, for which the active dipole is shorter than the reflector,
    `diff(2)` is the number of segments, for which the directors are shorter than the reflector,
    `beta` is an angle given by $\beta = 360° - \theta$ (see the layer A – fig. **4.3A.2**) and
    `F` denotes directivity pattern of the antenna.

Now, we turn our attention to the basic blocks of the program.

First, $[\exp(-jkr)/(4\pi r)]$ is numerically integrated to all the possible distances among segments of both a single antenna element and different antenna elements of Yagi antenna (see paragraph 4.1, the layer D).

Second, results of numerical (vector `psi`) are composed into the impedance matrix. Individual sub-matrices $[Z_{rr}]$, $[Z_{dd}]$, $[Z_{11}]$, $[Z_{22}]$, ..., $[Z_{nn}]$ are composed the similar way as described in the paragraph 4.1, the layer D.

On the basis of a known impedance matrix, the admittance matrix can be computed

```
Y = inv(Z);
```

current distribution on antenna elements can be obtained

```
for m=1:(N-2*diff(1)) % current distribution of the dipole
    I(m)=Y(m+N,feed);
    n(m)=m;
end

for m=1:N % current distribution of the reflector
    I(m)=Y(m,feed);
    n(m)=m;
end

for o=3:M
    for m=1:(N-2*diff(2)) % current distribution of directors
        Id(m)=Y(2*N-2*diff(1)+m+(o-3)*(N-2*diff(2)),feed);
        u(m)=m;
    end
end
```

input impedance can be evaluated

```
1/Y(feed,feed)
```

and directivity pattern can be computed.

All the above-given parameters of the antenna can be computed for an arbitrary wavelength of the excitation voltage, for an arbitrary length of the reflector, for arbitrary distances reflector - active dipole, active dipole - director and director - director (the way of individual setting of the distance among directors is described in the layer C), for an arbitrary radius of antenna wire, for an arbitrary number of antenna elements and

for an arbitrary number of segments to which antenna elements are divided. The length of the active dipole is for `diff(1)` segments shorter than the length of the reflector, the length of each director is for `diff(2)` segments shorter than the length of the reflector. Magnitude of the feeding voltage is normalized. The radius of antenna elements has to be properly chosen in order to meet conditions ensuring functionality of the method (see the layer A). More details are given in the source code of the program in the form of comments.

4.3 Yagi antenna

Java applet

Graph / Graf 1	Graph 2	Graph 3	Graph 4

Input / Vstup dat	Output / Vstup dat

Wavelength of the supplying voltage
0.68

Radius of the antenna wire
0.001

Length of the whole reflector
0.408

Distance between the reflector and the active dipole
0.1

Number of segments of reflector (odd number & max
15

N of segments for which active element is shorter th
1

Number of segments for which directors are shorter
2

Index of the excitation segment (feed)
22

The total number of antenna elements (minimum 3)
4

Distance between the active dipole and the nearest c
0.107

Distance between two neighboring directors
0.043

Calculate / Vypotl

Help

4.3 Yagi antenna

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**10** of **10**:

Question 1

Yagi antenna consists of…:

Possible answers for question 1:

○ … an active dipole, of a reflector and of several directors.

○ … an active dipole and of several directors (one of them is longer than the dipole, several of them are shorter).

○ … several active dipoles, of a reflector and of several directors.

◉ I won't answer this question

Question 2

The directors of Yagi antenna are excited by electromagnetic wave, and induced currents. The directors…

Possible answers for question 2:

○ … influence gain of the antenna.

○ … concentrate the antenna radiation into one half-space only.

○ … influence all the parameters of the antenna (gain, directivity pattern, input impedance).

◉ I won't answer this question

Question 3

Depending on the number of directors, the gain of Yagi antenna is usually in the interval …

Possible answers for question 3:

○ … 5 to 9 db.

○ … 10 to 15 db.

○ … 16 to 20 db.

◉ I won't answer this question

Question 4

The lobes in the directivity pattern of Yagi antenna appear …

Possible answers for question 4:

○ … both in front of reflector and even behind the reflector.

○ … in front of reflector only.

○ … behind the reflector only.

◉ I won't answer this question

Question 5

If Yagi antenna is analyzed by the method of moments, mutual- and self-impedances of antenna segments …

○ … of the active dipole appear in the impedance matrix; the influence of the reflector and directors are neglected.

○ … of the active dipole appear in the impedance matrix; the influence of the reflector and directors is incorporated to the analysis by the method of electromagnetic forces.

○ … of all the wires (active dipole, directors, reflector) appear in the impedance matrix.

◉ I won't answer this question

Question 6

A Yagi antenna is said to have a power gain over a dipole antenna for the same frequency band because …

Possible answers for question 6:

○ … it radiates more power than a dipole.

○ … it concentrates the radiation in one direction.

○ … it can be used for more than one band.

◉ I won't answer this question

Question 7

The maximum radiation from a three element Yagi antenna is …

Possible answers for question 7:

○ … in the direction of the reflector end of the boom.

○ … in the direction of the director end of the boom.

○ … at right angles to the boom.
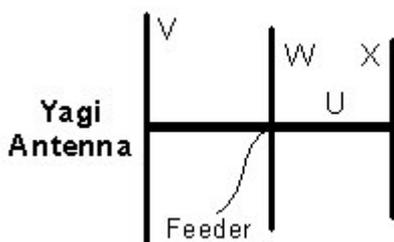
◉ I won't answer this question

Question 8

If the distance between the reflector and the active dipole decreases, the module of the current distribution of the reflector …

Possible answers for question 8:

○ … increases.

○ … decreases.

○ … is not affected.

◉ I won't answer this question

Question 9

In this diagram the item U, V & X correspond to respectively the …

Possible answers for question 9:

○ … boom, driven element & reflector.

○ … driven element, reflector & director.

○ … boom, reflector & director.

◉ I won't answer this question

Question 10

Changing λ (the wavelength of the supplying voltage) between 0 & 1 affects the directivity pattern …

Possible answers for question 10:

- ○ … more than changing λ between 1 & 2.
- ○ … less than changing λ between 1 & 2.
- ○ … almost the same amount as changing λ between 1 & 2.
- ◉ I won't answer this question

see result

Changing λ (the wavelength of the supplying voltage) between 0 & 1 affects the directivity pattern …

Possible answers for question 10:

- ○ … more than changing λ between 1 & 2.
- ○ … less than changing λ between 1 & 2.
- ○ … almost the same amount as changing λ between 1 & 2.
- ○ I won't answer this question

4.4 Microstrip dipole

Basic theory

Microstrip antennas are frequently used in today's wireless communication systems. Thanks to their low profile, they can be mounted to the walls of buildings, to the fuselages of airplanes or to the reverse sides of mobile phones. Moreover, microstrip antennas are fabricated using the same technology as producing printed circuit boards. Therefore, the fabrication is relatively simple and well reproducible. Finally, microstrip antennas can be simply integrated directly to microwave circuits, which are based on microstrip technology, and therefore, no special transmission lines, symmetrization circuits or connectors are needed on the contrary to classical antennas [8].
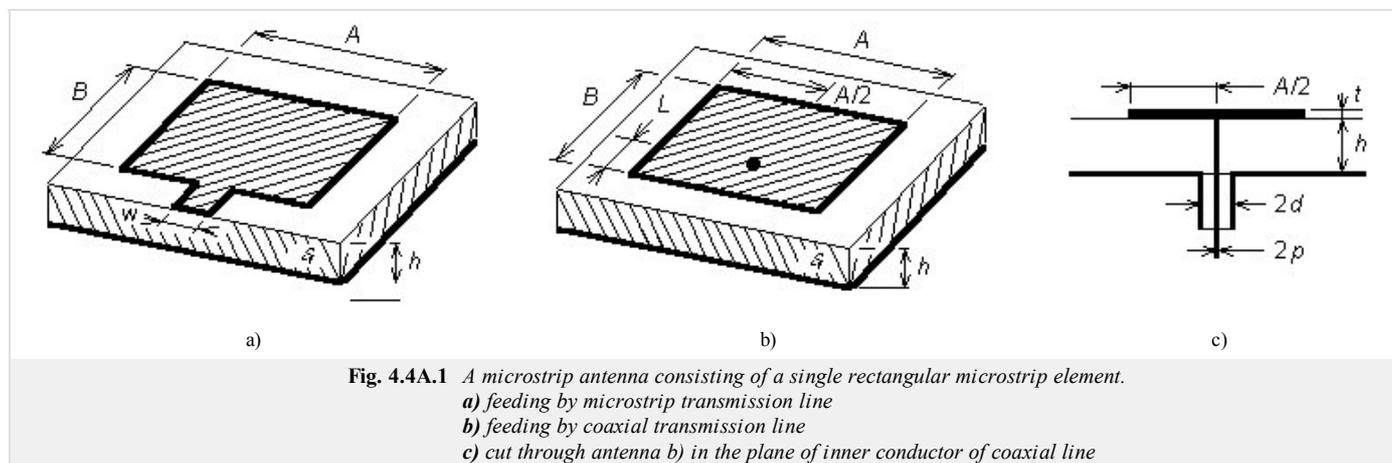
Nevertheless, there are several disadvantages of microstrip antennas. Narrow operation band is the main disadvantage. Due to this property, the design of microstrip antenna arrays exhibiting sufficiently low level of side-lobes is a really hard nut to be broken. Even the *parasitic radiation* of the feeding microstrip network, which can deform the directivity pattern [8].

One of the most frequently used types of microstrip antennas, the patch antenna, is depicted in fig. **4.4A.1a**. The antenna consists of a conductive rectangle of the dimensions $A \times B$, which is etched on a dielectric substrate. The antenna is fed by the microstrip transmission line (fig. **4.4A.1**, the microstrip goes from the front edge of the substrate crossways from the left). The second side of the substrate (on fig. **4.4A.1a** depicted as the bottom one) is continuously electroplated. The electroplated side plays the role of a *reflector* in the sense of zero potential (from the point of view of feeding) and in the sense of limiting radiation in the direction behind the reflector. Further, we call the electroplated side the ground plane.

The microstrip antenna that is fed by the microstrip transmission line (fig. **4.4A.1a**) can be considered as an open (non-shielded) open-ended transmission line, which is significantly widened at its end. If electromagnetic wave propagates along such transmission line, electromagnetic energy is primarily radiated into surrounding at the non-homogeneities (spontaneous widening of the microstrip at the border of the feeding line and the antenna element and the open end of this element) of the transmission line. The structure therefore behaves as a transmitting antenna. Moreover, if the length of the microstrip antenna element equals to the half of the wavelength on this widened transmission line, then input impedance of such an antenna is purely real [8]. Then, the antenna is said to be in resonance.

Radiation of a microstrip antenna can be explained in different ways. We can come out of the current distribution on the antenna element, which can be consequently understood as a wire antenna consisting of a very wide and a very thin antenna conductor. Next, we can come out from the line of electric intensity at the front side and at the back side of the antenna element (from the point of view of the feeding transmission line) and we can explain the radiation as an effect of a strong horizontal component (i.e. of the component, which is oriented in parallel with the ground plane) of electric field intensity vector at those edges.

As an alternative to the microstrip feeding of the microstrip antenna, a feeding by coaxial probe can be used (fig. **4.4A.1b**). Whereas the outer conductor of the coaxial cable is connected to the grounding plane, the inner conductor goes through a gap in the grounding plane and continues through the substrate to the microstrip antenna element, where it is fixed (fig. **4.4A.1c**). Significant reduction of the parasitic radiation of the feeding transmission line is the main advantage [8].

Fig. 4.4A.1 *A microstrip antenna consisting of a single rectangular microstrip element.*
a) feeding by microstrip transmission line
b) feeding by coaxial transmission line
c) cut through antenna b) in the plane of inner conductor of coaxial line

On the other hand, the coaxial feeding losses the main advantages of the microstrip feeding - a simple realization of a feeding system when microstrip antenna elements are grouped into arrays.

Now, turn our attention to the mathematical modeling of microstrip antennas.

Thanks to the popularity of microstrip antennas, several types of models of those antennas have been developed. Numerical models are of different validity domains (some are valid for *lower* microwave frequencies, some are valid for a limited interval of dimensions, etc.). Numerical models significantly differ even in their CPU-time demands (some models are based on closed-form expressions, which leads to low computational requirements, and some models are based on numerical methods, which leads to high CPU-time demands). The methods significantly differ even in the reached accuracy.

In the next, we concentrate on modeling microstrip antennas by the method of moments, which was described in the paragraph 4.1. As already said, the method comes out of Maxwell equations in the integral form, and therefore, we compute voltages and currents on the microstrip antenna element instead of searching for the distribution of electromagnetic field in the antenna surrounding.

In this paragraph, we are going to simplify the analysis. Instead of analyzing a patch antenna, we deal with a microstrip dipole (fig. **4.4A.2**) only. At the dipole, a small width of the antenna microstrip B is assumed. Then, only longitudinal currents (direction x) need to be considered and one-dimensional problem is going to be solved, which is an analogy to the solution of a wire dipole (paragraph 4.1)

Analyzing the antenna, we assume an infinitely wide substrate, infinitely small thickness of metallic layers $t \approx 0$, perfect electric conductivity of all metallic parts, lossless dielectrics and vacuum in the antenna surrounding.

Assume a plane wave of an angular frequency ω, which impinges the microstrip dipole. The electric field intensity vector $\mathbf{E}^I$ (the upper index I denotes the incident wave) describes the incident wave. This wave induces conductive currents in the antenna wire, which are described by the current density vector $\mathbf{J}$. Since the current has to be zero at the ends of the microstrip dipole, a charge described by the charge density ρ is accumulated here. In the following half-wave, the orientation of conductive currents is changed and the accumulated charge leaves the dipole **[9]**.

Fig. 4.4A.2 *Microstrip dipole plus reflector.*
a) global view
b) discretization net for x component of current density

The effects of currents are described by the vector potential

$$\mathbf{A}(\mathbf{r}) = \iint_S \{ G_A(\mathbf{r}|\mathbf{r}_0) \cdot \mathbf{J}(\mathbf{r}_0) \} dS_0$$ (4.4A.1a)

and the effects of charges are described by the scalar potential

$$V(\mathbf{r}) = \iint_S \{ G_V(\mathbf{r}|\mathbf{r}_0) \rho(\mathbf{r}_0) \} dS_0.$$ (4.4A.1b)

Value of the vector potential $\mathbf{A}(\mathbf{r})$ in the *observation pint* $\mathbf{r}$ on the microstrip dipole can be computed by the consecutive addition of contributions of components of the current density vector $\mathbf{J}$ from all the points of the microstrip antenna element $\mathbf{r}_0$ whereas the current density vector $\mathbf{J}$ is multiplied by the respective column of the dyadic Green function $\mathbf{G}_A$. The observation point, where the value of the vector potential is computed, is determined by the vector $\mathbf{r}$. Summation of contributions of all the currents flowing on the surface of the antenna element to the vector potential $\mathbf{A}(\mathbf{r})$ in the point $\mathbf{r}$ is performed by the integral over all the surface of the microstrip dipole S. The *source points*, where currents contributing to the vector potential to the point $\mathbf{r}$ are consecutively determined by the vector $\mathbf{r}_0$. The argument of the dyadic Green function $\mathbf{r}|\mathbf{r}_0$ shows the contribution of the current in $\mathbf{r}_0$ the potential in $\mathbf{r}$.

Dealing with the computation of the scalar potential, the dyadic Green function G_A is replaced by the scalar Green function G_V, and instead of the current density vector, the scalar charge density r appears. Except of this, eqn. (4.4A.1b) is identical with eqn. (4.4A.1a).

As already said, current density and charge density on the microstrip dipole are mutually associated. Therefore, eqns. (4.4A.1a) and (4.4A.1b) have to be completed by the continuity theorem given by

$$-j\omega\rho = \nabla \cdot \mathbf{J}.$$ (4.4A.1c)

Eqn. (4.4A.1c) expresses the fact that current flowing from a closed region (see the current divergence as the right-hand side) has to equal to the charge reduction in this region within one second (see time derivative at the left-hand side).

If both the vector potential and the scalar one on the microstrip dipole are expressed, then the electric field intensity, which is radiated by the antenna, can be computed

$$\mathbf{E}^S = -j\omega\mathbf{A} - \nabla V.$$ (4.4A.1d)

Time derivative of the vector potential $j\omega\mathbf{A}$ describes a *dynamic contribution* of electric charge on the dipole to the transmitted wave (electric conductive currents, i.e. charges in motion, are sources of vector potential). Gradient scalar potential gives a *static contribution*of electric charge on the dipole to the transmitted wave (static charges, which are concentrated this moment at the ends of the dipole, are sources of scalar potential).

The final problem, which has to be solved, are boundary conditions. Since all the metallic parts of the antenna are perfectly electrically conductive, components of electric field intensity, which are tangential to the antenna surface, have to be zero on this surface

$$\mathbf{n}_0 \times \mathbf{E}^S = -\mathbf{n}_0 \times \mathbf{E}^I.$$ (4.4A.1e)

In the above-given relation, $\mathbf{E}^S$ denotes electric intensity of a wave, which is transmitted by the antenna, and $\mathbf{E}^I$ is electric intensity of a wave, arriving to the antenna from outside. The vector $\mathbf{n}_0$ is unitary vector, which is perpendicular to perfect electrically conductive surfaces.

As already explained in the paragraph 4.1, analytical solution of (4.4A.1) is unknown, and therefore, we utilize the moment method for its solution.

In the first step of the analysis of the microstrip dipole, the antenna is placed to Cartesian coordinate system (fig. **4.4.2**). Next, vector equations (4.4A.1) are rewritten in the elected coordinate system to the scalar form. Moreover, we assume a special case when moving on the surface of the microstrip dipole (thanks to the boundary conditions, value of the tangential component of electric field intensity is known, which can be used further).

We start at eqn. (4.4A.1a), which describes dependency of vector potential on current density vector. The mutual relation between those two quantities is
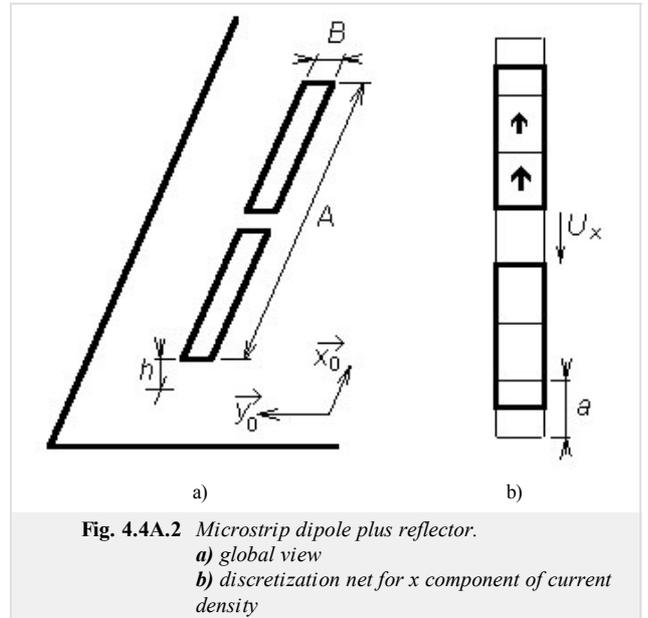
described by dyadic Green function. In a fact, dyadic Green function is a matrix 3×3 which elements are functions describing dependency of components x, y, z of vector potential on the components of current density vector. A single scalar element G_A^{st} of this dyadic Green function can be understood as s-th component of vector potential, which is *supplied* by an elementary electric dipole (described by a constant current density vector) in the direction t. Obviously, only the component xx of dyadic Green function is non-zero (assuming a very small width of the microstrip dipole, y-component of current density vector is zero and the only non-zero J_{xx} can be source of x-component of vector potential A_x; other components of vector $\mathbf{A}$ are of zero value).

If x component of vector potential (i.e., the component oriented along the dipole) is going to be computed, then (4.4A.1a) can be rewritten to

$$A_x(x_m, y_n) = \iint_S \left\{ G_A^{xx}(x_m, y_n | x', y') J_x(x', y') \right\} dx' dy'. \tag{4.4A.2}$$

Here, (x_m, y_n) are coordinates of a point on the surface of the dipole, where vector potential A_x is computed. Coordinates (x', y') specify the position of x-component of current density, which supplies x-component of vector potential. During the integration, coordinates (x', y') walk through all the points of the antenna surface S. The symbol J_x denotes x-component of current density vector, the symbol G_a^{xx} represents x diagonal term of dyadic Green function.

That way, (4.4A.1a) is adopted for the case of the analyzed antenna, and therefore, the attention is turned to the rearrangement of (4.4A.1b). Since (4.4A.1b) is a scalar equation containing scalar quantities only, the adoption consists in considering the introduced coordinated system

$$V(x_m, y_n) = \iint_S \left\{ G_V(x_m, y_n | x', y') \rho(x', y') \right\} dx' dy'. \tag{4.4A.3}$$

Again, scalar potential V is computed on the surface of the dipole in the point (x_m, y_n). Computing this potential, the product of scalar Green function G_V and charge density r is integrated over the whole surface of the dipole. The movement on the surface during integration is done by changing coordinates (x', y').

Next, we turn our attention to continuity equation (4.4A.1c). Since only x-component of current density vector is non-zero (conductive currents can flow in the direction of dipole axis only), the relation can be rewritten to the form

$$-j\omega \rho(x, y) = \frac{\partial J_x(x, y)}{\partial x}. \tag{4.4A.4}$$

Considering (4.4A.4), charge density r can be expressed as a function of current density J_x and can be substituted to (4.4A.3). That way, charge density is elliminated from (4.4A.3) and both potentials are expressed as functions of components of current density vector

$$V(x_m, y_n) - \frac{1}{j\omega} \iint_S \left\{ G_V(x_m, y_n | x', y') \left[\frac{\partial J_x(x', y')}{\partial x} \right] \right\} dx' dy'. \tag{4.4A.5}$$

Further, vector potential (4.4A.2) and scalar potential (4.4A.5) are substituted to (4.4A.1d), which enables us to compute electric field intensity of the radiated wave. Since both vector potential (4.4A.2) and scalar one (4.4A.5) are functions of an unknown current distribution on the surface of the dipole, even electric field intensity is a function of this current distribution

$$E_x^S(x_m, y_n) = -j\omega A_x(x_m, y_n) - \frac{\partial V(x_m, y_n)}{\partial x},$$

$$E_x^S(x_m, y_n) = -j\omega \iint_S \left\{ G_A^{xx}(x_m, y_n | x', y') J_x(x', y') \right\} dx' dy' + \frac{1}{j\omega} \iint_S \frac{\partial}{\partial x} \left\{ G_V(x_m, y_n | x', y') \left[\frac{\partial J_x(x', y')}{\partial x} \right] \right\} dx' dy'. \tag{4.4A.6}$$

Magnitude of electric field intensity on the surface of microstrip dipole can be determined, assuming perfect electric conductivity, from boundary condition (4.4A.1e). Current distribution J_x is the only unknown in (4.4A.6). And the equation (4.4A.6), which contains the unknown function $J_x(x, y)$, is going to be solved by moment method

The way of obtaining a piecewise-constant approximation of current distribution on the basis of (4.4A.6) was described in the paragraph 4.1. Therefore, the approach is here reminded only.

1. The region, where the solution of the integral equation is going to be found, has to be discretized (the surface of the dipole is divided to sub-regions, which do not overlap on one hand and which totally cover the whole analyzed dipole on the other hand (see fig. **4.4A.2b**).
   Performing discretization, boundary conditions have to be kept in mind. In our situation, x-component of current density J_x has to be zero at edges $x = 0$, $x = A$ because ends of the dipole can be understood as open ends of the microstrip transmission line.
   Exploiting piecewise constant approximation of current density components, boundary conditions can be met simply. The discretization mesh is extended behind the end of the dipole for one half of the discretization segment. Then, we enforce those extended segments to represent zero value of current (see fig. **4.4A.2b**).

2. The sought function $J_x(x, y)$ is approximated exploiting known basis functions (they are of unitary value over surface of a single discretization element and of zero value over the rest of elements) and unknown approximation coefficients

$$\tilde{J}_x^{(n)}(x, y) = \sum_{nx=1}^{Nx} J_x^{(nx)} \Pi^{(nx)}(x, y). \tag{4.4A.7}$$

In the above-given relation, $\tilde{J}_x^{(n)}$ denotes approximation of x-component of current density over n-th discretization element, $J_x^{(nx)}$ is a *sample* of exact value of this component in the middle of n-th element, $\Pi^{(nx)}$ denotes basis function, which is unitary over the element nx and which is zero elsewhere, and Nx is total number of elements in the discretization mesh.

3. Approximation (4.4A.7) is substituted into the initial equation (4.4A.6). Since the approximation (4.4A.7) does not meet the initial equation (4.4A.6) exactly, we have to respect this fact adding the residual function $R_x(x, y)$ to (4.4A.6) together with the approximation (4.4A.7)

$$R_x(x, y) = E^S(x, y) - \sum_{nx=1}^{Nx} \left\{ J_x^{(nx)} \iint_S \left[G_A^{xx}(x, y | x', y') + \overline{G}_V(x, y | x', y') \right] dx' dy' \right\}. \tag{4.4A.8}$$

In these relations, G_A^{xx} is x diagonal term of dyadic Green function and using $\overline{G}_V$, partial derivative of scalar Green function is expressed.

Since basis functions Π are unitary over a respective element and are zero elsewhere, they do not have any representation in the above-given relations. Next, integration and summation were swapped, and the approximation coefficients J_x^{nx} were moved in front of the integral thank so their constant character.

4. Residual function $R_x(x, y)$ is going to be minimized. Lower values of the residual function are, closer our solution to the exact solution is. Minimization is done by the method of weighted residuals

$$\iint_S \{W_x(x, y)R_x(x, y)\}dxdy = 0. \qquad (\,4.4\mathrm{A}.9\,)$$

Here, R_x denotes the residual function computed according to (4.4A.8). The symbol W_x represents properly elected weighting functions. Weighting is done by Dirac pulses in order to eliminate one of integrations thanks to the filtering property

$$\iint_S \{\delta(x - x_m, \, y - y_n)R_x(x, y)\}dxdy = R_x(x_m, \, y_n). \qquad (\,4.4\mathrm{A}.10\,)$$

Accuracy of the method, which performs weighting by Dirac pulses, cannot be very good because the error is not minimized globally in the whole analyzed region but only in points, where Dirac pulses are of non-zero value (we operate with *moments* of residual function only).

5. Using the same number of weighting functions as the number of unknown approximation coefficients is, we obtain the set of N linear equations for N unknown coefficients. Solving this set of equations, we obtain unknown values of approximation coefficients, and therefore, approximation of the current distribution on the microstrip dipole can be composed. Considering known current distribution, the desired technical parameters of the antenna (input impedance, gain or directivity pattern) can be computed.

The whole algorithm described by the above points is given in detail in the layer B and in [10], [11]. Replacing all the derivatives by central differences in the above-given algorithm, we obtain a matrix equation

$$\mathbf{U}_x = \mathbf{Z}_{xx}\,\mathbf{I}_x. \qquad (\,4.4\mathrm{A}.11\,)$$

In this equation, $\mathbf{U}_x$ is column vector of voltages in the direction x on discretization elements. This voltage is computed by multiplying x-component of electric field intensity by x-size of the discretization element

$$U_x(m, n) = E_x(m, n)a. \qquad (\,4.4\mathrm{A}.12\,)$$

The symbol a denotes the height of the discretization element (i.e. the size in the direction x).

Since the microstrip dipole is supposed to be fabricated from perfect electric conductor (voltage on this conductor is zero), the vector of voltages is filled in by zeros only except of elements relating to the excitation gap.

Next, $\mathbf{I}_x$ is column vector of currents in the direction x, which is unknown this moment for us. Elements of $\mathbf{I}_x$ are related to the component of current density J_x as follows

$$I_x(m, n) = J_x(m, n)B \qquad (\,4.4\mathrm{A}.13\,)$$

(B is the width of the dipole, and consequently the width of the discretization element). Impedance matrix $\mathbf{Z}_{xx}$ describes contribution of currents $\mathbf{I}_x$ and contribution of charge densities ρ (expressed from continuity equation (4.4A.4) using x-components of current density J_x on elements) to voltages $\mathbf{U}_x$ on those elements. Elements of the impedance matrix $\mathbf{Z}_{xx}$ are known (see the layer B)

$$Z_{xx}(m, n) = \frac{j\omega a}{B}\,\Gamma_A^{xx}(m, n) + \frac{1}{j\omega aB}\left[\Gamma_V(m^+, n^+) - \Gamma_V(m^-, n^+) - \Gamma_V(m^+, n^-) + \Gamma_V(m^-, n^-)\right]. \qquad (4.4\mathrm{A}.14\,)$$

In order to evaluate the impedance matrix $\mathbf{Z}_{xx}$, values of integrals of Green functions Γ_A^{xx} and Γ_V over the surface of the discretization element have to be computed for various distances between the source elements (over its surface, current distribution and charge one are integrated) and the observation one (on its surface, electric intensity is computed. Description of the numeric computation of those integrals in Matlab is given in the layer D.

Matlab program, which performs analysis of the microstrip dipole by the moment method, is described from user's point of view in the layer C. Here, we provide only illustration results obtained by the program.

For simplicity, the substrate between the dipole and the reflector is assumed to be of the same parameters as vacuum. If the dipole length equals to one half of wavelength and if the dipole width is $B = \lambda/1000$, following values of input impedance are obtained (in the first row of the table, the number of discretization elements to which antenna is subdivided is given):

Tab. 4.4A.1 *Input impedance for different number of discretization elements*

N	10	20	30	40
R_{vst} [Ω]	95.2	97.3	98.0	98.4
X_{vst} [Ω]	72.4	73.6	74.7	75.5

The results show the method to exhibit good stability with respect to the number of discretization elements. Moreover, the input impedance of the microstrip dipole with the reflector computed by moment method, is close to the results of analytical computations (sinusoidal current distribution J_x a wire dipole over an infinite planar reflector assumed). For a single wire dipole, radiation resistance of the antenna (related to the input) equals to $R_\Sigma = 85.6\ \Omega$.
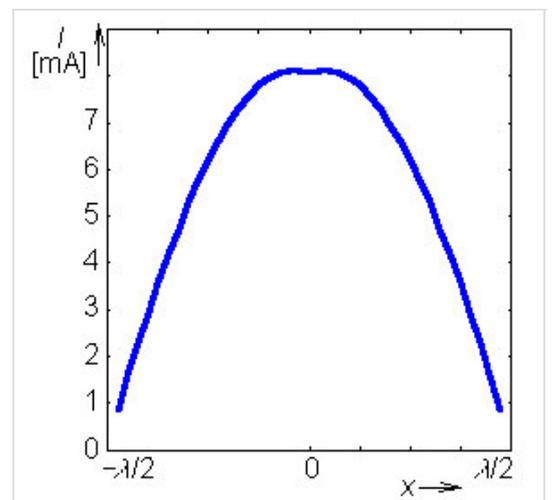
Investigating approximation of the distribution of *x*-component of the current along the microstrip dipole, we can show that depicting approximation coefficients from the vector $\mathbf{I_x}$ into a chart. That way, we obtain a course which is close to a sinusoidal current distribution (see fig. **4.4A.3**).

Fig. 4.4A.3 *Current distribution on half-wavelength symmetric microstrip dipole with planar reflector in the distance of one quarter of wavelength. Distribution computed for 40 cells.*

4.4 Microstrip dipole

Advanced theory

As a microstrip dipole, we denote an antenna consisting of two narrow microstrip arms, which are fed by a symmetric source in the center. The antenna is placed on an upper side of the dielectric substrate. The bottom side of the substrate is fully covered by a metallic layer and I of zero potential (fig. **4.4B.1**).

Effects of currents flowing on the antenna can be described by vector potential
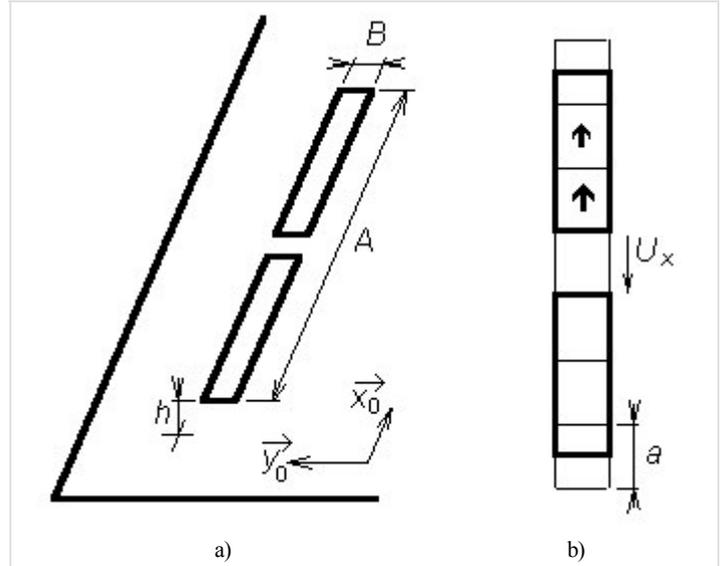
Fig. 4.4B.1 *Microstrip dipole plus reflector.*
a) global view
b) discretization net for x component of current density

$$\mathbf{A}(\mathbf{r}) = \iint_S \{ G_A(\mathbf{r}|\mathbf{r}_0) \cdot \mathbf{J}(\mathbf{r}_0) \} dS_0.$$ (4.4B.1a)

Effects of charges on the antenna can be described by scalar potential

$$V(\mathbf{r}) = \iint_S \{ G_V(\mathbf{r}|\mathbf{r}_0) \rho(\mathbf{r}_0) \} dS_0.$$ (4.4B.1b)

In the above-given equations, $\mathbf{A}(\mathbf{r})$ denotes vector potential in the point $\mathbf{r}$, $\mathbf{J}$ is current density vector in the point $\mathbf{r}_0$, $\mathbf{G}_A$ is dyadic Green function and G_V denoted scalar Green function (see the layer D). The parameter of Green functions $\mathbf{r}|\mathbf{r}_0$ tells us that we compute the contribution of a current (charge) from the point $\mathbf{r}_0$ to potentials in the point $\mathbf{r}$. The symbol r denotes charge density.

Current density and charge one are mutually associated by the continuity equation

$$-j\omega\rho = \nabla \cdot \mathbf{J}.$$ (4.4B.1c)

If both the vector potential and the scalar one are expressed on the surface of the microstrip dipole, we can evaluate electric intensity of a wave, which is radiated by the antenna.

$$\mathbf{E}^S = -j\omega \mathbf{A} - \nabla V.$$ (4.4B.1d)

Applying (4.4B.1) to the microstrip dipole from fig. **4.4B.1** and substituting from continuity equation to (4.4B.1b), we obtain the following relations

$$A_x(x_m, y_n) = \iint_S \{ G_A^{xx}(x_m, y_n | x', y') J_x(x', y') \} dx' dy',$$ (4.4B.2a)

$$V(x_m, y_n) = -\frac{1}{j\omega} \iint_S \{ G_V(x_m, y_n | x', y') \rho(x', y') \} dx' dy',$$ (4.4B.2b)

$$E_x^S(x_m, y_n) = -j\omega A_x(x_m, y_n) - \frac{\partial V(x_m, y_n)}{\partial x}.$$ (4.4B.2c)

Here, A_x denotes x-component of vector potential and V is scalar potential, G_A^{xx} denotes x diagonal element of dyadic Green function and G_V is scalar Green function, J_x is x-component of sought vector of current distribution and E_x is x-component of the radiated electric field intensity. Details are given in the layer A.

Substituting vector potential (4.4B.2a) and scalar one (4.4B.2b) to the relation (4.4B.2c), we obtain the initial equation for moment analysis of the dipole

$$E_x^S(x_m, y_n) = -j\omega \iint_S \{ G_A^{xx}(x_m, y_n | x', y') J_x(x', y') \} dx' dy' + \frac{1}{j\omega} \iint_S \frac{\partial}{\partial x} \left\{ G_V(x_m, y_n | x', y') \left[\frac{\partial J_x(x', y')}{\partial x} \right] \right\} dx' dy'.$$ (4.4B.3)

Magnitude of electric field intensity on the surface of the microstrip dipole can be computed using a boundary condition if perfect conductivity of metallic parts is assumed (except of the excitation gap, the intensity is zero). Therefore, x-component of current density J_x is the only unknown in (4.4B.3).

In the first step, surface of the dipole is divided into discretization elements. Center of the first discretization element is denoted by **1**, center of the second element by **2**, etc. The upper bound of the discretization element is denotes by the same number completed by superscript "+" (plus), the lower bound by superscript "-" (minus); see fig. **4.4B.1**.

Next, points in the center of discretization elements are used for computing x-component of electric field intensity. Since contribution of the current to the magnitude of electric field intensity is described by vector potential without presence of derivatives (see eqn. 4.4B.2), vector potential is computed in the center of elements.

On the contrary, contribution of charges to electric field intensity is described by scalar potential performing two derivations according to x. In numerical computations, derivations are replaced by central differences. Values of scalar potential V, which derivatives are used for determining contributions of charges to the electric intensity of the radiated wave, have to be known at the borders of discretization elements so that the result of central differentiating appears in the center of the element. Values of charge density are computed from continuity equation deriving components of current density in centers of discretization elements (derivatives are replaced by central differences again). In order to obtain values of charge density on the border of elements, we have to differentiate components of current distribution in the center of elements. This fact suits us very well because the computed values of current density are valid just in these points.

Finally, values of current density components are computed in the center of elements and values of charge density have to be evaluated on borders of the elements. Therefore, value of components of vector potential and value of components of electric field intensity are valid for the center of elements, and value of scalar potential for border of elements.

In the next step, we substitute piecewise constant approximation of current distribution to initial relations and we replace partial derivatives by central differences. Exploiting continuity equation, charge density on the upper edge and on the low one of the discretization element is expressed

$$\rho\left(m_x^+, n_x\right) = \frac{-1}{j\omega} \frac{J_x\left(m_x+1, n_x\right) - J_x\left(m_x, n_x\right)}{a}, \qquad (\,4.4B.4a\,)$$

$$\rho\left(m_x^-, n_x\right) = \frac{-1}{j\omega} \frac{J_x\left(m_x, n_x\right) - J_x\left(m_x-1, n_x\right)}{a}. \qquad (\,4.4B.4b\,)$$

Symbol a height of the discretization element (see fig. **4.4B.1**), $J_x(m_x, n_x)$ corresponds to a constant value of x-component of current density vector on the surface of the element containing the central point m_x, n_x) and ω is angular frequency.

On the basis of equations (4.4B.4a) and (4.4B.4b), we compute contribution of charges, which are represented by charge density ρ, to the x-component of electric field intensity vector. Considering charge densities of the upper edge $\rho(m_x^+, n_x)$ and on the low one $\rho(m_x^-, n_x)$, scalar potential on those edges can be computed. Moreover, substituting partial derivatives according to x by central differences, we obtain contribution of charges to x-component of electric field intensity.

Now, charge densities on borders of discretization elements are known and we assume that those values are valid not only on the borders but too over the whole surface of charge elements (they are of the same size as discretization elements but they are shifted so that borders of discretization elements can be in the center of charge ones as depicted in fig. **4.4B.1**). Then, charge densities can be described by the following piecewise constant functions

$$\rho(x, y) = \rho(x_m^+, y_n)\Pi(x_m^+, y_n|x, y) \quad x \in (x_m, x_{m+1}),\ y \in (y_n^-, y_n^+), \qquad (\,4.4B.5a\,)$$

$$\rho(x, y) = \rho(x_m^-, y_n)\Pi(x_m^-, y_n|x, y) \quad x \in (x_{m-1}, x_m),\ y \in (y_n^-, y_n^+). \qquad (\,4.4B.5b\,)$$

In these relations, $\Pi(x_m^+, y_n|x, y)$ denotes a function, which is unitary in the rectangular region with the center in (x_m^+, y_n), a is height of the element and B is its width. The similar situation is for points (x_m^-, y_n). Values of charge density $\rho(x_m^+, y_n)$ and (x_m^-, y_n) in the center of this rectangular region are given by relations (4.4B.5).

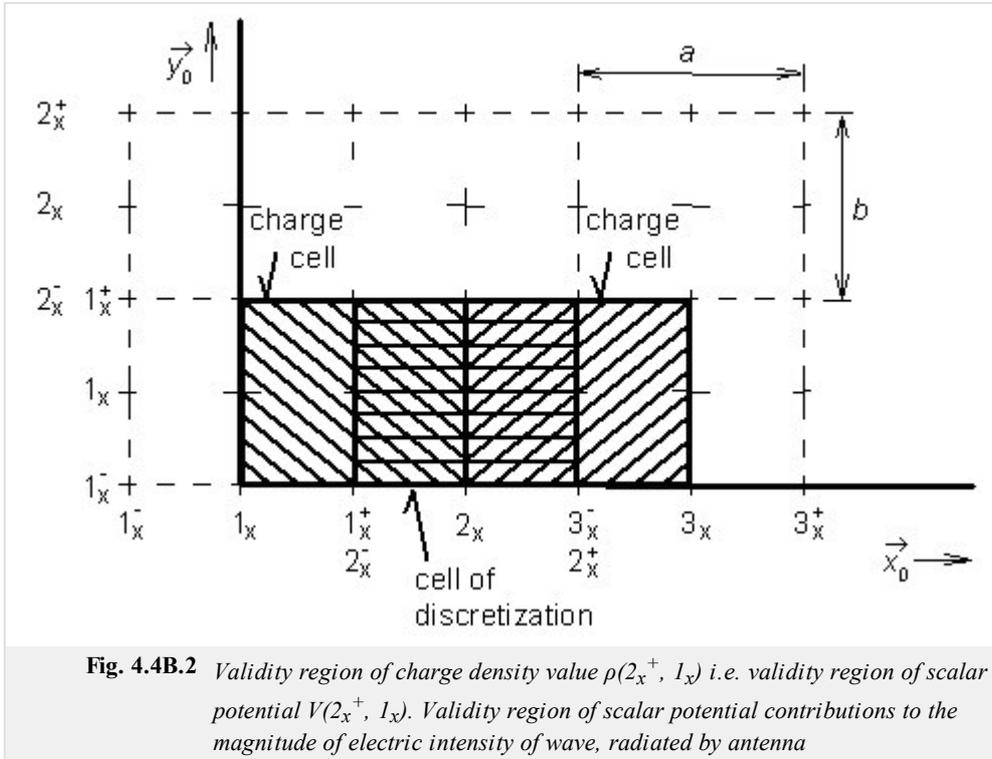If charge density on the microstrip dipole is known, we can substitute the distribution to

$$V(x, y) = -\frac{1}{j\omega} \iint_S \left\{ G_V\left(x, y|x', y'\right)\rho\left(x', y'\right) \right\} \mathrm{d}x'\mathrm{d}y'$$

and we can compute scalar potential on respective charge elements. For the charge element, which center lies on the upper edge of the element (m_x, n_x), we get

$$V\left(m_x^+, n_x\right) = -\frac{1}{j\omega} \iint_S \left\{ G_V\left(m_x^+, n_x|x', y'\right) \sum_{p,q} \left[\rho\left(p_x^+, q_x\right)\Pi\left(p_x^+, q_x|x', y'\right)\right] \right\} \mathrm{d}x'\mathrm{d}y' =$$

$$(\,4.4B.6\,)$$

$$= -\frac{1}{j\omega} \sum_{p,q} \left\{ \rho(p_x^+, q_x) \int\limits_{-B/2}^{+B/2} \int\limits_{-a/2}^{+a/2} G_V(m_x^+ - p_x^+, n_x - q_x | x', y') \mathrm{d}x' \mathrm{d}y' \right\}$$

and similarly for $V(m_x^-, n_x)$.

Fig. 4.4B.2 *Validity region of charge density value $\rho(2_x^+, 1_x)$ i.e. validity region of scalar potential $V(2_x^+, 1_x)$. Validity region of scalar potential contributions to the magnitude of electric intensity of wave, radiated by antenna*

Before continuing, let us discuss the relation (4.4B.6). Rearranging it, we swapped integration and summation, and the integral of the product of the unitary pulse Π and the scalar Green function G_V over the whole dipole was substituted by the integral of the single scalar Green function over the surface of this charge element, where π is non-zero.

Dealing with indexes, (m, n) determines position of the *observation* element, over which the value of scalar potential is computed, and indexes (p, q) specifies position of the *source* element, whose charges contribute to the scalar potential of the element (m, n).

Scalar Green function G_V is the only continuous function in (4.4B.6), and therefore, the integral has to be evaluated for this function only. Evaluating this integral for various distances between the source element and the observation one, position of the observation element is changed only, and the source element stays in the origin of the coordinate system. Therefore, integration limits stay the same in all the cases (from $-a/2$ to $+a/2$ for the coordinate x' and from $-B/2$ to $+B/2$ for the coordinate y').

Now, constant values of scalar potential are known over all charge elements. Therefore, we can compute contribution of charges, which are represented by scalar potential, to the values of x-component of electric field intensity

$$^V E_x^S(m_x, n_x) = -\frac{\partial V}{\partial x} \cong -\frac{V(m_x^+, n_x) - V(m_x^-, n_x)}{a}. \qquad (4.4B.7)$$

Here, a denotes height of the element, B is the width of the element, and values of scalar potential V over charge elements are given by the relation (4.4B.6).

As already said, we are aimed to express an approximation of electric field intensity of the radiated wave on the surface of the dipole as a function of current density on this dipole. Hence, we substitute (4.4B.6) for scalar potential V on the border of elements, which enforces (4.4B.7) to be a function of charge density r on edges of elements

$$^V E_x^S(m_x, n_x) = \frac{1}{j\omega a} \left\{ \sum_{p,q} [\rho(p_x^+, q_x) \Gamma_V(m_x^+ - p_x^+, n_x - q_x)] - \sum_{p,q} [\rho(p_x^-, q_x) \Gamma_V(m_x^- - p_x^-, n_x - q_x)] \right\}. \qquad (4.4B.8)$$

In the above-given relation, a denotes height of element and B is its width. Angular frequency ω corresponds to the frequency, on which antenna is analyzed. The symbol $^V E_x^S$ denotes a contribution of scalar potential V to the magnitude of x-component of the vector of electric intensity of the radiated wave. Finally, G_V represents integral of scalar Green function G_V over the surface of the element

$$\Gamma_V(m-p, n-q) = \int\limits_{-B/2}^{+B/2} \int\limits_{-a/2}^{+a/2} G_V(m-p, n-q|x', y')\mathrm{d}x'\mathrm{d}y'. \tag{4.4B.9}$$

As the next step in expressing electric field intensity as a function of current density, values of charge density from (4.4B.5) are substituted to (4.4B.8)

$$
{}^V E_x^S(m_x, n_x) = +\frac{1}{\omega^2 a^2} \sum_{p,q} \left\{ \left[J_x(p_x+1,q_x) - J_x(p_x, q_x) \right] \Gamma_V(m_x^+ - p_x^+, n_x - q_x) \right\} -
$$
$$
- \frac{1}{\omega^2 a^2} \sum_{p,q} \left\{ \left[J_x(p_x, q_x) - J_x(p_x-1,q_x) \right] \Gamma_V(m_x^- - p_x^-, n_x - q_x) \right\}. \tag{4.4B.10}
$$

The relation describes the contribution to the component of the vector of electric field intensity using unknown values of the component of current distribution vector J_x and known coefficients Γ_V, given by (4.4B.9). From the point of view of scalar potential, the aim was reached, and therefore, the attention is turned to the vector potential.

In order to evaluate the contribution of the current to the electric field intensity, we have to compute vector potential substituting piecewise constant approximation of current density to (4.4B.2a)

$$
A_x(m_x, n_x) = \iint\limits_S \left\{ G_A^{xx}(m_x, n_x|x', y') \sum_{p,q} \left[J_x(p_x, q_x) \Pi(p_x, q_x|x', y') \right] \right\} \mathrm{d}x'\mathrm{d}y'
$$
$$
= \sum_{p,q} \left\{ J_x(p_x, q_x) \int\limits_{-B/2}^{+B/2} \int\limits_{-a/2}^{+a/2} G_A^{xx}(m_x - p_x, n_x - q_x|x', y')\mathrm{d}x'\mathrm{d}y' \right\} \tag{4.4B.11}
$$
$$
= \sum_{p,q} \left\{ J_x(p_x, q_x) \Gamma_A^{xx}(m_x - p_x, n_x - q_x) \right\}
$$

In the above-given relation,

$$
\Gamma_A^{xx}(m-p, n-q) = \int\limits_{-B/2}^{+B/2} \int\limits_{-a/2}^{+a/2} G_A^{xx}(m-p, n-q|x', y')\mathrm{d}x'\mathrm{d}y'. \tag{4.4B.12}
$$

Next, a is height of the discretization element and B is its width. G_A^{xx} denotes x diagonal component of dyadic Green function. Function $\Pi(p_x, q_x|x',y')$ is unitary over an element with the center in the point (p_x, q_x) and is zero elsewhere. Values J_x represent piecewise constant current density in the element mesh (p_x, q_x).

During derivation, integration and summation were swapped, and integral over the whole dipole surface S was replaced by the integral over a single element (due to multiplying by Π the integrand is non-zero over a single discretization element only).

Finally, substituting vector potential to (4.4B.3) and replacing derivatives of scalar potential by contributions (4.4B.10), we obtain the final equation

$$
E_x^S(m_x, n_x) = -j\omega \sum_{p,q} \left\{ J_x(p_x, q_x)\Gamma_A^{xx}(m_x - p_x, n_x - q_x) \right\} +
$$
$$
+ \frac{1}{\omega^2 a^2} \sum_{p,q} \left\{ \left[J_x(p_x+1,q_x) - J_x(p_x, q_x) \right]\Gamma_V(m_x^+ - p_x^+, n_x - q_x) \right\} - \tag{4.4B.13}
$$
$$
- \frac{1}{\omega^2 a^2} \sum_{p,q} \left\{ \left[J_x(p_x, q_x) - J_x(p_x-1,q_x) \right]\Gamma_V(m_x^- - p_x^-, n_x - q_x) \right\}.
$$

Since eqn. (4.4B.13) is rather complicated, we rewrite it into a matrix form

$$\mathbf{U}_x = \mathbf{Z}_{xx}\,\mathbf{I}_x. \tag{4.4B.14}$$

Here, $\mathbf{U}_x$ is column vector of voltages in the direction x over elements. Voltages are computed by multiplying x-component of electric field intensity by x-dimension of the discretization element

$$U_x(m, n) = E_x(m, n)a. \tag{4.4B.15}$$

The symbol a denotes height of the discretization element (i.e. the element dimension in the direction x).

Since the microstrip dipole is assumed to be fabricated form perfect electric conductor, the vector of voltages consists of zeros only (except of excitation elements).

Next, $\mathbf{I}_x$ is column vector of unknown current in the direction x. Elements of $\mathbf{I}_x$ are related to current density J_x by the equation

$$I_x(m, n) = J_x(m, n)B \qquad (4.4B.16)$$

(B is width of the dipole, and therefore, even the width of the discretization element). Impedance matrix $\mathbf{Z}_{xx}$ describes contributions of currents $\mathbf{I}_{xx}$ and contributions of charge densities r to voltages $\mathbf{U}_x$ over elements. Single elements of the impedance matrix $\mathbf{Z}_{xx}$ are obtained comparing (4.4B.13) to (4.4B.16)

$$Z_{xx}(m, n) = \frac{j\omega a}{B} \Gamma_A^{xx}(m, n) + \frac{1}{j\omega a B} \left[\Gamma_V(m^+, n^+) - \Gamma_V(m^-, n^+) - \Gamma_V(m^+, n^-) + \Gamma_V(m^-, n^-) \right]. \qquad (4.4B.17)$$

In order to evaluate impedance matrix $\mathbf{Z}_{xx}$, we have to know values of integrals of Green functions over the surface of the discretization element for different distances between source elements and observation ones. Numerical computation of those integrals is described in the layer D.

The Matlab program for the analysis of the microstrip dipole by the moment method is described in the layer C from user's point of view.

4.4 Microstrip dipole

Matlab program

The program, which accompanies the paragraph dealing with moment analysis of the microstrip dipole, is of simple user's interface. First the path of Matlab has to be redirected to the folder **M_dipol**. Then in the command window of Matlab, the basic m-file `mdipol.m` has to be run, and an appearing form has to be filled in. In the form, the operation frequency of the antenna, length a width of the dipole, height and relative permittivity of the dielectric substrate have to be given. Then, the program prints the input impedance of the dipole, and in an independent window, current distribution on the dipole is drawn.

The program uses a fixed number of discretization elements (N = 33), the excitation gap is positioned to the center of the dipole.

4.4 Microstrip dipole

Developing Matlab

Since we became familiar with theoretical basis of the moment analysis of the microstrip dipole in the layer A (and the layer B), we can turn our attention to the software implementation of the analysis in Matlab.

First, we take care for the numerical evaluation of integrals of Green functions.

For simplicity, we introduce a variable, which expresses the distance between the actual position of the source of waves (x', y') and the actual observation point (x_m, y_n) on the surface of the dipole

$$\delta = \sqrt{(x_m - x')^2 + (y_n - y')^2}. \qquad (\ 4.4D.1\ )$$

Using the distance δ the x diagonal component of dyadic Green function is of the form **[12]**

$$G_A^{xx}(\delta) = \frac{\mu}{4\pi} \left\{ \frac{\exp[-jk_0 r_0(\delta)]}{r_0(\delta)} - \frac{\exp[-jk_0 r_1(\delta)]}{r_1(\delta)} \right\}, \qquad (\ 4.4D.2\ )$$

where

$$k_0 = \omega \sqrt{\mu_0 \varepsilon_0} \qquad (\ 4.4D.3\ )$$

denotes wave-number in the vacuum (μ_0 and ε_0 are permeability and permittivity of the vacuum and ω is angular frequency of the wave we are working with). Next, distances r_0 and r_1 can be described by the relation

$$r_i^2(\delta) = \delta^2 + (2ih)^2, \qquad (\ 4.4D.4\ )$$

where δ is the distance between the *source* and *observation point* on the surface of the antenna element and h is the height of the substrate.

Scalar Green function can be described by the following relation **[12]**

$$G_V(\delta) = \frac{1-\eta}{4\pi\varepsilon} \left[\frac{\exp(-jk_0 r_0)}{r_0} - (1+\eta) \sum_{i=1}^{\infty} \frac{(-\eta)^{i-1} \exp(-jk_0 r_i)}{r_i} \right], \qquad (\ 4.4D.5\ )$$

where

$$\eta = \frac{\varepsilon_r - 1}{\varepsilon_r + 1} \qquad (\ 4.4D.6\ )$$

and the rest of symbols is of the same meaning as described in layers A and B. Let us notice the infinite sum in the relation (4.4D.5). If the dielectrics between the ground plane and the microstrip dipole is replaced by the vacuum, the sum consists of a single non-zero summand (substituting $\varepsilon_r = 1$ to (4.4D.6) yields $\eta = 0$, and only the zero power of zero differs from zero). Increasing the value of the permittivity, the coefficient h approaches one, and the value of summands decreases relatively slowly (the decrease is caused by the increasing value of r_i). Therefore, the number of summands has to be increased when relative permittivity rises. The following table illustrates the situation:

Tab. 4.4D.1 *Number of summands for different permittivity values*

ε_r	2	4	6	9
i	9	16	23	33

Here, i denotes an index of the second addend in the summation (4.4D.5), which modulus is lower than 0.01 (considered parameters: $f = 3$ GHz, $\delta = 0$ mm, $h = 1.5$ mm)

Green functions (4.4D.2) and (4.4D.5) are not of general validity. They can be used for a microstrip structure consisting of a ground plane, of a homogeneous dielectric substrate and of a microstrip antenna element only. Moreover, those relations can be utilized for a very low distance between sources of EM field and observation points, where electric field intensity is computed. Since we compute electric field intensity on the surface of the microstrip antenna element, the distance between sources and observation points is satisfactorily short for exploiting (4.4D.2) and (4.4D.5).

Now, when mathematic description of Green functions is known to us, we can start their integration over the whole surface of a discretization element. Since analytic solution of those integrals is unknown, we perform that in a numeric way. The numeric integration does not cause any problems if the source element differs from the destination one. In the opposite case (when computing contributions of currents and charges on the element to the electric field intensity on the same element), the problem of singularity of the integrated function appears. If the source point and the destination one are identical then their distance δ is zero (4.4D.1). As obvious from (4.4D.4), even the distance r_0, which is present in

the denominators of Green functions (4.4D.2) and (4.4D.5), is zero. As explained in **[13]**, the singularity is weak, and therefore, we can eliminate it. The singularity elimination consists in dividing Green function G to the singular part G_H (which has to be carefully handled) and to the regular one $(G - G_H)$, which can be numerically integrated a usual way. Hence

$$G = G_H + (G - G_H). \qquad (4.4D.7)$$

In our case, only the first term of Green functions (4.4D.2) and (4.4D.5) is singular. On the basis of (4.4D.7), the singular term can be divided to the singular part and to the regular one by the following way:

$$\frac{\exp(-jk_0 r_0)}{r_0} = \frac{1}{r_0} + \frac{\exp(-jk_0 r_0) - 1}{r_0}. \qquad (4.4D.8)$$

Whereas the singular part of Green function is represented by the first addend in (4.4D.8), the second addend in (4.4D.8) corresponds to the regular difference of Green functions. Regularity of the second addend can be proven by computing its limit for the distance between the source and observation point tending to zero

$$\lim_{r_0 \to 0} \frac{\exp(-jk_0 r_0) - 1}{r_0} = \lim_{r_0 \to 0} \frac{-jk_0 \exp(-jk_0 r_0)}{1} = -jk_0. \qquad (4.4D.9)$$

Evaluating this limit, l'Hospital rule was used.

The integral of the singular part of Green function over the whole surface of the discretization element can be computed analytically **[12]**

$$\int_{-a/2}^{+a/2} \int_{-B/2}^{+B/2} \frac{\mathrm{d}x\mathrm{d}y}{\sqrt{x^2 + y^2}} \cong 2a \ln\left[\tan\left(\frac{\alpha}{2} + \frac{\pi}{4}\right)\right] - 2b \ln\left[\tan\left(\frac{\alpha}{2}\right)\right]. \qquad (4.4D.10)$$

Here, a denotes the height of the discretization element, b is its width and a is given by

$$tg\,\alpha = \frac{B}{a}. \qquad (4.4D.11)$$

The practical algorithm of computing the singular term of Green functions (4.4D.2) and (4.4D.5) is illustrated by fig. (4.4D.1). The regular part of the singular term (the second addend in eqn. 4.4D.8) is numerically integrated from $-a/2$ to $a/2$ with respect to x and from $-B/2$ to $-n$ with respect to y (see the left-hand part of the element). The number n approaches zero (-10^{-5}, e.g.), which enables us to integrate up to a near distance from the origin (here, the integration is troubled by zero denominator). Since the integrated function behaves well near the origin (as shown in eqn. (4.4D.9), its value is about jk_0 here), only a small inaccuracy is produced.

The integral of the regular part of the singular Green function over the right-hand part of the element does not need to be evaluated because its numerical value is the same as the left-hand side integral.

Fig. 4.4D.1 *Integrating singular parts of Green function over discretization cell*

The complete integral of Green function is computed by taking twice the left-hand side integral and by adding the integral of the singular term, which is obtained by evaluating (4.4D.10).

Now, the moment analysis can be implemented in Matlab.

Programming the moment method, we start to write m-files containing Green function. Therefore, we have to write m-files implementing relations (4.4D.2) and (4.4D.5).

The first m-file of Green function is devoted to the case when computing contribution of charges and currents on the discretization element to potentials on the same element. Moreover, we concentrate to the term, where the distance r_0 is present (we compute effects of *currents* on the antenna element to *observation points* on the same element). Performing the described integration, singularity appears.

```
r0 = sqrt( x.*x + y^2);
out = ( exp( -j*k0*r0) - 1)./r0;
```

The second file of Green function is used to computing contributions of charges and currents on the discretization element to potentials on other elements (i.e., terms containing r_0 for non-zero distance between *sources* and *observation points* are handled) and to computing effects of the ground plane (i.e., terms containing r_i are handled). Therefore, singularity does not appear when integrating.
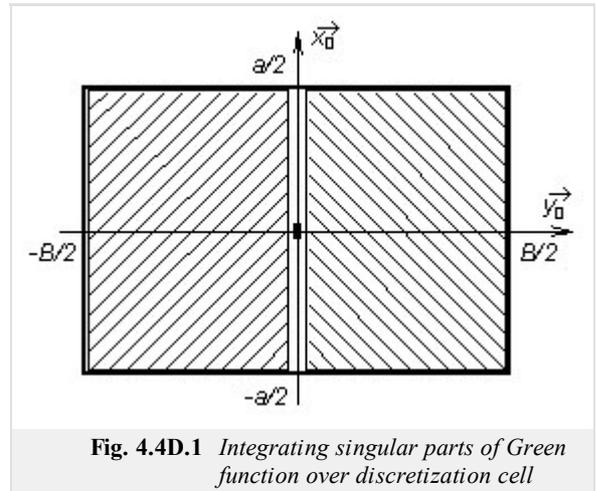
```
r1 = sqrt( (x+xdis).*(x+xdis) + y^2 + (2*i*h)^2); out = exp( -j*k0*r1)./r1;
```

In the above-given source code, `x` and `y` denote the position of a *source*. Sources are assumed on an element, which center corresponds to the origin of the coordinate system (0, 0). The distance between the center of the *observation* element and the center of the *source* element in *x*-direction is given by the parameter `xdis`. The parameter `h` denotes the distance between the dipole and the reflector, and `k0` is wave-number in vacuum (where the antenna is placed).

Integrating Green function, parameters `x` and `y` vary over the whole surface of the observation element. Parameters `xdis`, `h` and `k0` are constant during integration. Nevertheless values of constant parameters have to be passed on m-files of Green functions. Passing on is implemented by declaring `xdis`, `h` and `k0` as global variables. Then, content of these variables is *visible* from any m-file of the program. Global variables are declared by typing

```
global h k0 xdis i
```

both at the beginning of the main m-file and at the beginning of all m-files, which exploit these variables.

Finally, integration variables `x` and `y` are passed on m-files as parameters in their headers:

```
function out = regular( x, y)
```

Now, Green functions are ready to integration over the whole surface of the element, and therefore, the main m-file can be written. After initial declarations (determining frequency, antenna dimensions, and the discretization mesh), a cycle over all the possible distances of discretization elements is programmed. For each distance, we evaluate integrals of Green functions and we store them to the array `psi`. The first index of `psi` indicates the distance between the *source element* and the *observation* one in x-direction, the second index can be up to the value `Ne` (integral of the term of Green function, which contains the distance r_n). The whole cycle can be of the following form:

```
psi = zeros( Nx+2, Ne);
for m=1:(Nx+2)
    xdis = (m-1)*a; % horizontal distance of element centers
    if m==1
        psi(m,1) = 2*dblquad('singular', -a/2, +a/2, 1e-10, B/2, 1e-3,
        'quad');
        psi(m,1) = psi(m,1) + stt;
    else
        i = 0;
        psi(m,1) = dblquad('regular', -a/2, +a/2, -B/2, B/2, 1e-3, 'quad');
    end
    for i=1:Ne
        psi(m,i+1) = dblquad('regular', -a/2, +a/2, -B/2, B/2, 1e-3,
        'quad');
    end
end
```

The above-given source code shows that the singular function is integrated in such a case only, when the *source* discretization element and the *observation* one are identical (`m` = 1) and when the effect of the reflector is not handled.

As already explained, the singular Green function is numerically integrated up to the close distance of the singularity (i.e., up to the distance 10^{-10} meters), and then, the value of the analytically computed integral of the singular part (4.4D.10) is added

```
alp = atan( b/a);
stt = 2*a*log( tan( alp/2+pi/4)) - 2*b*log( tan( alp/2));
```

The double integral is numerically evaluated using the standard Matlab function `dblquad`. The first parameter is the string containing the name of the m-file, where he integrated function is implemented. Integration limits of the first integration variable are the second parameter and the third one (the variable has to be the first parameter in the header of the m-file containing the integrated function). Integration limits of the second integration variable are the fourth parameter and the fifth one (the variable has to be the second parameter in the header of the m-file containing the integrated function).

The sixth parameter of `dblquad` can be used for determining accuracy of numeric integration, and the last parameter can specify the algorithm of numeric integration.

When integrals of Green functions are evaluated for all the possible distances between elements, the impedance matrix $\mathbf{Z}_{xx}$ (4.4A.14, 4.4B.17) can be composed. Composing can be performed by the following source code:

```
G = ( psi(:,1) - psi(:,2))/(4*pi);
cf1 = (1-eta)/(j*omg*ep0*epr*a*B);
cf2 = j*omg*mi0*a/B;
```

```matlab
% relation betw. currents and voltages

ix = 0;
Zxx = zeros( Nx, Nx);

for m=1:Nx % over observation elements in x
    ix = ix + 1; iy = 0;
    for o=1:Nx % over source elements in x
        dx = abs( m-o); % distance source - observation
        iy = iy + 1;
        hlp = 2*psi( 1+dx,1) - psi( 1+abs(dx+1),1) - psi( 1+abs(dx-1), 1);
        for i=1:Ne
            inc = ( 2*psi(1+dx,i+1)-psi(1+abs(dx+1),i+1)-psi(1+abs(dx-
            1),i+1)); hlp = hlp - (1+eta)*(-eta)^(i-1)*inc;
        end
        Zxx(ix,iy) = cf1*hlp/(4*pi) + cf2*G(1+dx,1);
    end
end
```

The source code corresponds to relations (4.4A.14) and (4.4B.17).

Inverting the impedance matrix, admitance matrix is obtained. If excitation voltage 1 V is assumed, elements of the admitance matrix directly equal to currents on the microstrip dipole. Inverting the element of the admitance matrix, which corresponds to the excitation element, input impedance of the antenna is obtained.

4.4 Microstrip dipole

Java applet

Output / Výstup dat	Module Graph	Phase Graph / Fázi Graf

Input / Vstup dat

Enter the frequency of the microstrip dipole antenna:

Frequency / Kmitočet [GHz]
6

Enter the length of the microstrip dipole antenna:

Length of dipole / Délka dipólu [mm]
20

Enter the width of the microstrip dipole antenna:

Width of dipole / Šiřka dipólu [mm]
1

How many segments does the antenna have?

# of Segments / Počet segmentů (max. 250)
33

Enter the height of the substrate:

Height of substrate / Výška substrátu [mm]
10

Enter the relat. permitivity of the substrate:

Rel.permitivity of substrate/Rel.permitivita substrátu
2

Calculate / Vypočti   Help

4.4 Microstrip dipole

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**10** of **10**:

Question 1

Microstrip dipole etched on a microwave substrate …

Possible answers for question 1:

- ○ … behaves the same way as a wire dipole completed by a plane reflector in vacuum.
- ○ … exhibits rather complicated behavior due to the wave reflections on the boundary of the dielectric substrate and the vacuum.
- ○ … behaves the same way as a wire dipole completed by a plane reflector in free space, which permittivity equals to the arithmetic average of the permittivity of vacuum and the permittivity of the substrate.
- ● I won't answer this question

Question 2

Green functions describe how …

Possible answers for question 2:

- ○ … currents contribute to vector potential and charges contribute to scalar potential.
- ○ … current distribution on the antenna influences the directivity pattern.
- ○ … current distribution contributes to input impedance of the antenna.
- ● I won't answer this question

Question 3

Dyadic Green function …

Possible answers for question 3:

- ○ … and Green function are two names for the same thing.
- ○ … is 3 x 3 matrix of functions relating contributions of current density components to vector potential components.
- ○ … is 3 x 3 matrix of functions for anisotropic non-linear substrates only.
- ● I won't answer this question

Question 4

Continuity theorem …

Possible answers for question 4:

- ○ … relates charges and currents on the dipole.
- ○ … expresses the fact that electromagnetic wave is continuously radiated by the antenna.
- ○ … explains the continuity of current distribution.
- ● I won't answer this question

Question 5

If the frequency of the microstrip dipole antenna decreases, the phase interval of the current distribution of the microstrip dipole antenna …

Question 6

If the heigth of the substrate (of the microstrip dipole antenna) is decreased below its default value 10 [in the applet on layer E], then the module of the current distribution of the microstrip dipole …

Question 7

At the ends of a half-wave dipole the …

Question 8

A circular microstrip antenna element has a radius of 60 mm with ε_r = 2.0 and thickness t = 2.0 mm. The lowest mode giving a maximum in the radiation pattern in the broadside direction (normal to the plane of the disc) has a frequency f_r of …

Question 9

In the design described in the above question, the patch is excited by a 50Ω coaxial probe feed at a point along a radial line where the calculated R_{in} = 50Ω. The inner conductor of the coaxial feed has a radius of 1 mm. The required capacitive element C in series with the probe to make the antenna resonant is approximately …

Question 10

The microstrip antenna is in resonance when the length of the microstrip element equals to …

○ … the double of the wavelength on the widened transmission line.

○ … the wavelength on the widened transmission line.

○ … the half of the wavelength on the widened transmission line.

● I won't answer this question
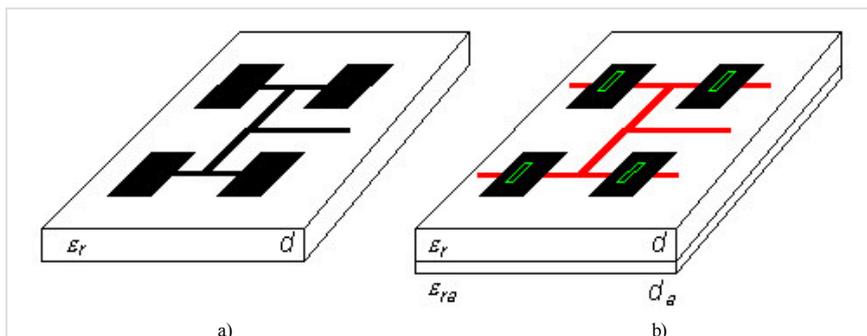
see result

4.5 Microstrip antenna

Basic theory

We become familiar with the patch antenna in the paragraph describing the analysis of the wire dipole (par. 4.4) already. Nevertheless, we left this antenna and concentrated on the one-dimensional moment method for the analysis of a dipole. Since patch antennas play a significant role in today's wireless communication systems, we pay our attention to its analysis in this special part of the textbook.

In the introductory part, architecture and properties of patch antennas are described. Then, two ways of the antenna analysis are explained. Finally, accuracy and computational demands of those two methods are compared. In the appendix, an overview and basic features of commercial programs for the analysis of patch antennas are given.

I. Introduction

First notes on patch antennas appeared in the literature at the beginning of fifties of 20th century. This time, planar antennas, which can meet the surface of airplanes and military equipment, were demanded. Later, patch antennas were intensively exploited even in the civic are (in communication especially).

These days, a plenty of various types of patch antennas are at our disposal. These antennas differ in the shape of antenna elements (rectangular, circular), in polarization they operate with, in feeding (in par. 4.4, microstrip and coaxial feeding were described), etc. In our description, we concentrate on rectangular patch antennas fed by a microstrip transmission line. We subdivide path antennas into two groups depending if the feeding microstrip is placed on the same side of the substrate as the antenna element (classical patch antenna, fig. **4.5A.1a**) or not. The second (newer) case is going to be investigated more intensively.

Within last 15 years, indirect feeding of patch antennas is frequently used. For this purpose, small coupling slots in the ground plane are utilized. These slots are usually

Fig. 4.5A.1 *Classical patch antenna.*
a) aperture-fed patch antenna
b) feeding microstrip at the bottom side of the substrate is drawn in red, apertures in the ground plane in green.

positioned in the center of antenna elements. Below the slot, where the feeding microstrip is placed, a certain (inductive) impedance can be observed. This impedance physically represents an antenna element, which *exhausts* energy from the microstrip transmission line. The microstrip crosses the slot and continues further (see overlaps in fig. **4.5A.1**) in order to compensate the reactive part of the above-mentioned impedance by the shunt. Antennas, which are fed by the described way, are for simplicity called aperture-fed patch antennas.

Both groups of antennas exhibit some similarities and some differences (see tab. **4.4A.1**).

Small impedance bandwidth is a common property of patch antennas. For classical patch antennas, the bandwidth is about 2 to 3%, for aperture-fed antennas about 4 to 6%. High quality factor of patches is the reason. Patches behave as a resonator of a quality factor of several tens.

Antenna efficiency is another important parameter. Radiation efficiency (computed for lossless antenna) is given by the ratio of radiated power (integrating Poynting vector in the far region over the whole half-space) and real power on the input port of the antenna (element).

Energy, which is not radiated by the antenna, is taken away in the form of surface waves along the infinite dielectric substrate (in real situation, the substrate is limited, of course).

From the practical point of view, the total efficiency is important. Except of losses in dielectrics, surface-wave losses contribute to its value. Surface-wave losses are extremely undesired. Lossy dielectrics guiding surface waves can be imagined as a lossy transmission line. Exploiting a limited substrate, standing waves are formed in the substrate. Higher losses are related to higher standing wave ratio. Increase of losses due to the increase of standing wave ratio is called surface-wave losses.

The total efficiency of patch antennas is for 1 to 2 dB worse that the efficiency of reflector antennas.

Patch antennas exhibit very good cross-polarization. This parameter is not important for antennas operating linearly polarized waves. Designing a dual antenna exhibiting high polarization isolation, this parameter is of high importance.

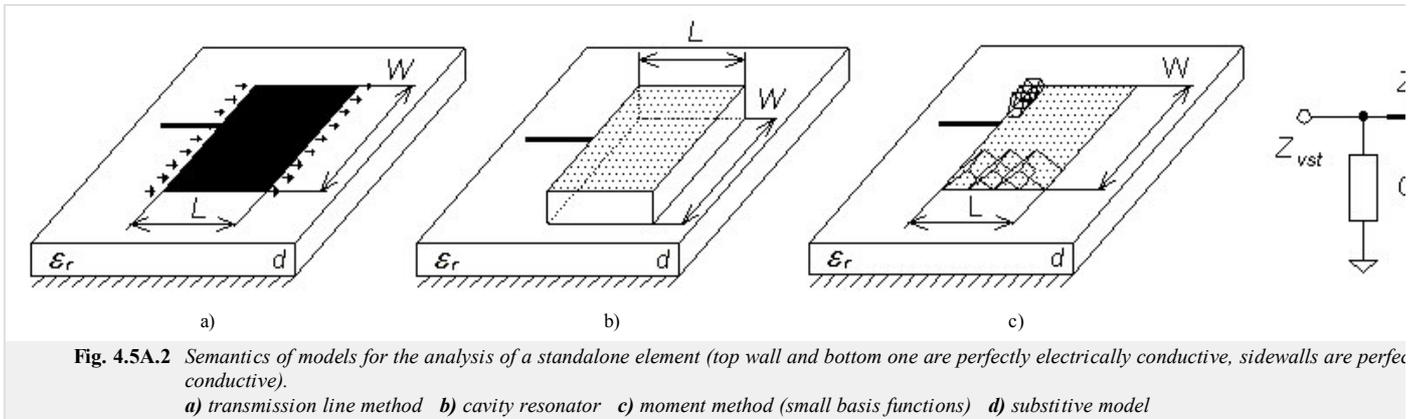|	classical antenna	aperture-fed antenna	note
Tab. 4.5A.1 *Overview of parameters of patch antennas*			
impedance bandwidth	2 to 3 %	4 to 6 %	in % of resonant frequency
radiation efficiency	80 %	80 %	surface-wave losses only
total efficiency	40 to 60 %		considering losses in dielectrics
cross-polarization	35 to 40 dB		
patch coupling	indispensable	small	
feeder radiation	middle (*)	small	(*) reducible by feeding topology

II. Analysis of stand-alone antenna element

There are many methods for the analysis of a stand-alone antenna element. We are going to describe three methods, which are of the most frequent use. Two of the are going to be explained in detail.

The first method of the analysis understands the antenna element as a wide microstrip transmission line. As the radiation source, fringe electric fields are considered (fig. **4.5A.2a**).

The second method of analysis consists in replacing the space below the antenna element by a lossy resonator. Then, input impedance can be computed using the theory of cuboidal cavity resonators **[14]**. In that case, zero normal component of electric field intensity is assumed on sidewalls of the resonator (see fig. **4.5A.2**). Hence, sidewalls can be understood as perfect magnetic conductors.

Numerical analysis of the antenna element is the third method. The numerical analysis is based on the moment method, which is formulated in the spectral domain or in the spatial one (this formulation was introduced in par. 4.1, par. 4.3 and par. 4.4). Even here, we concentrate on the moment method in the spatial domain.

Fig. 4.5A.2 *Semantics of models for the analysis of a standalone element (top wall and bottom one are perfectly electrically conductive, sidewalls are perfe̶ conductive).*
a) transmission line method   b) cavity resonator   c) moment method (small basis functions)   d) substitive model

II.I. Analysis of a stand-alone element by the transmission line method

The method excels in simplicity on one hand, and enables to reach a good accuracy on the other hand. Accuracy of computing input impedance of the antenna element depends of the choice of antenna model. In this paragraph, the simplest model considering radiating slots between two neighboring antenna elements only is assumed (tangential fringe components of electric field intensity are considered as a single source of radiation). The other effects (mutual coupling among radiation sources, radiation of marginal edges) are neglected.

Designing the antenna, the admitance of the radiating slot Y_S (arrows in fig. **4.5A.2**) has to be known **[14]**

$$Y_S = \frac{\pi W}{\lambda_0 Z_0} \left\{ 1 + j\left[1 - \ln(2)\ln(k_0 w)\right] \right\}$$

(4.5A.1)

Here, W denotes length of the slot [m], w width of the slot [m] (it equals to the height of the substrate d approximately), λ_0 denotes free-space wavelength [m], k_0 is free-space wave-number ($k_0 = 2\pi/\lambda_0$) [m^{-1}] and Z_0 is free-space impedance [Ω].

In the relation, influence of the permittivity of the substrate is not included because the constant magnitude of electric field intensity is considered along the whole slot.

Designing the antenna, the length of the antenna element is chosen $(0.48 \div 0.49) \lambda/2$, where $\lambda = c/[f(\varepsilon_{eff})^{1/2}]$ and where the effective permittivity of the substrate ε_{eff} is computed using quasi-static relations or numerically (computing effective permittivity or characteristic impedance of microstrip transmission line, a proper computer program has to be chosen (several programs do not properly model the dispersion of the transmission line on higher frequencies; hence we recommend Txline from AWR, or TRL85 in Serenade 8.5SV).

The length of the antenna element does not equal exactly to the half of the wavelength because the capacitive impedance of the right-hand slot has to be transformed to the inductive impedance so that the capacitive impedance of the first slot on the frequency f can mutually compensate the inductive one.

In the second step, such width W of the antenna element is chosen so that the desired resonant resistance of the antenna element is obtained. This resistance can be expressed as $R_{rez} = (2 G_S)^{-1}$, where G_S is the real part of the slot admitance

$$G_S = \text{Re}\{Y_s\} = \frac{\pi W}{\lambda_0 Z_0}$$

(4.5A.2)

Substituting free-space wavelength λ_0 and free-space characteristic impedance Z_0, we get

$$R_{rez} = 120 \frac{\lambda_0}{2W}.$$

(4.5A.3)

Choosing $W = \lambda_0/2$, the resonant resistance equals to 120 Ω.

The described model of the antenna element is valid for electrically thin substrates only (we assume $d / \lambda \approx 0.01$, where d is the height of the substrate).

If the resonant resistance is required to be higher than 120 Ω, the length of the antenna element W has to be smaller than one half of the wavelength. Resonant resistance cannot be significantly increased because its maximum value cannot be (respecting feasibility of the feeding microstrip) higher than 150 Ω. If impedance transformer is exploited, then the antenna element of the impedance about 240 Ω can be used.

Optimizing the width of the antenna element by extending its length is not recommended due to the spatial reasons and functional ones (parasitic resonance and parasitic lobes can appear).

Satisfactorily small radiation resistance can be achieved by feeding the antenna element near its center instead of the edge feeding.

An example of the design of a stand-alone antenna element is given in the layer C including user's guide of an accompanying Matlab program.

II.II. Analysis of a stand-alone element by the moment method

Applicability to the limited number of shapes of the antenna element and to the electrically thin substrates is drawback of the transmission-line method. In order to analyze elements of a general shape and elements on electrically thick substrates, the moment method has to be used.

The planar antenna element (two-dimensional one in contrast to par. 4.4) is assumed to be fed by a microstrip transmission line. Computing the input impedance of the antenna in the position of the microstrip input is the aim of the analysis.

On the antenna element, there is a given current and charge distribution (in reality, a current is distributed over the antenna element). In order to compute the input impedance, currents and charges have to be computed first. Then, another quantity (electric field intensity, magnetic field intensity, input impedance, etc.) can be computed.

Since the moment analysis was in detail described in par. 4.1, par. 4.3 ane par. 4.4, a brief overview is presented here only:

Fig. 4.5A.3 *The mesh of current cells and charge ones*

1. The analyzed structure is divided to charge cells and current cells. The shape of cells is chosen as simple as possible (rectangular one). Next, both current and charge are assumed being constant over the cell (their value can differ from a cell to a cell). We concentrate on current cells because the charge can be computed using continuity equation. The mesh of current cells is shifted for one half of a cell with respect to charge cells.

2. We formulate a relation among electric field intensity in the center of i-th current cell and currents over all elements (including i-th one). Tangential components of electric field intensity in the center of elements are required to be zero except for source elements (the structure is assumed to be fed by flowing ideal voltage sources). Explanation of the connection of the source is given in point 4.

3. As a result, a set of simultaneous linear equations is obtained. Solving it, currents over all cells are obtained.

4. On the basis of computed current distribution over the antenna element, input impedance can be computed. Input impedance cannot be computed directly because the structure is fed by sources in the horizontal direction, which contacts the feeding microstrip by the first pole, and the auxiliary shunt by the second pole. The described voltage source is placed to the slot, which is denoted by one in fig. **4.5A.3**. The shunt is depicted by a dotted line. The ratio of a voltage and a current in the slot does not equal to the correct value of input impedance, and several additional steps have to be performed (we call them de-embedding, see layer B).
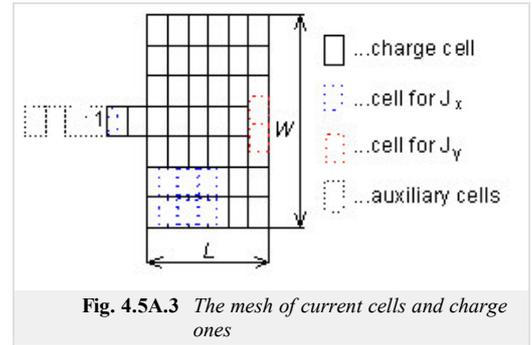
A detailed description of the analysis is rather complicated, and therefore we refer to the layer B. Further information is given in **[15]**. An example of moment analysis of 2D structures is given in the layer C.

Finally, we provide an overview of commercial programs, which are based on moment method.

Tab. 4.5A.2 *Commercial programs based on moment method*					
Product	**IE3D**	**Ensemble**	**Momentum**	**Sonnet**	**Microwave Office**
Producer	Zeland Software	Ansoft	Agilent Technologies	Sonnet	Applied Wave Research
Version	v. 7 (2000)	v.7 (2000)	ADS (1999)	v. 6	v. 3.22 (1999)
Domain [2])	3D/spatial	2.5D/spatial	2.5D/spatial	2.5D/spectral	2.5D/spectral
Cells	Triangle/Rectangle	Triangle	Triangle/Rectangle	Rectangle	Rectangle
Equidist. mesh	NO	NO	NO	YES	YES
Current approximation	Linear	Linear (RWG)	Linear (rooftops)	Linear (rooftops)	Linear (rooftops)
G. function for PEC box	NO (v. 8 YES)	NO	NO	YES	YES
Modeling open problems	Exact	Exact	Exact	Approximate	Approximate
Aperture feeding	YES	YES	YES	NO	NO
Conductor thickness	YES	NO	NO	YES	YES
Limitations	- - -	- - -	- - -	[1])	[1])

[1]) The transition between layers can be modeled by a single current layer only (i.e., it has to be much shorter than a wavelength).

[2]) Determination whether the program is 2.5-dimensional or 3-dimensional is rather problematic. Therefore, a simplified criterion is applied here: if the analyzed structure can contain conductors, which are oriented in an arbitrary angle with respect to the ground plane, then the program is denoted as 3-dimensional.

In some aspects, the above-given programs are similar, in the other ones, they differ. A planar formulation of the moment method in frequency domain (conductive facets are discretized) is a common feature. The basic difference consists in openness or closeness of the formulation (structure in a conductive box). Formulation of a closed problem is simpler because Green functions for a conductive box filled by a layered medium can be composed more easily compared to open structures.

Exploitation of basis functions, which are much smaller than wavelength, is another common feature of programs. In order to imagine how memory demands [MB] for the storage of matrices of coefficients rise, the following table was built:

Tab. 4.5A.3 *Memory demands [MB] for different number of unknowns*			
Unknowns	**100**	**1000**	**10000**
IE3D, Ensemble, Momentum	0,16	16	1600
Sonnet, Microwave Office	0,16 (0,08)*	16 (8)*	1600 (800)*

* lossless structure

A significant difference among programs is hidden in the type of the domain, where the problem is formulated. Whereas the programs, which are based on the spatial version of the moment method, can use rectangular cells, the spectral-domain programs allow a homogeneous mesh of rectangular elements only, which leads to an enormous size of resultant matrices. Therefore, *Sonnet* and *Microwave Office* utilize the technique of sub-domains, which are formed by summing contributions of elementary cells. Programs *Sonnet* and *Microwave Office* exploit FFT algorithm to accelerate double summations, which replace computation of double integral in the spectral method. The difference between *Sonnet* and *Microwave Office* is in the type of FFT: *Microwave Office* admits the total number of elements to be a product of prime numbers, *Sonnet* is limited to the power of two due to the use of standard FFT.

Programs *Sonnet* and *Microwave Office* are advantageous (compared to space-domain programs) in real matrices of coefficients for lossless structures (twice larger problems can be solved, and moreover, the problem is solved twice more quickly thanks to the use of real arithmetic).

All the above-described programs provide similar possibilities from the point of view of fullwave analysis. Of course, there are differences is the speed of programs. In general, programs based on the spatial moment method exhibit higher speed than those based on the spectral version of the method. More complicated development of programs is a price which has to be paid for that.

Simple and well-throughout user's interface is another very important feature of programs. From that point of view, *Microwave Office* is very popular.

4.5 Microstrip antenna

Advanced theory

In this paragraph, two-dimensional moment analysis of a patch antenna is presented. In a fact, the approach from par. 4.4 is extended for another dimension.

As already said, electric field intensity has to be expressed as a function of currents and charges over cells. For that purpose, vector potential **A** and scalar potential φ are used. Moreover, the positive convention $\exp(+jkr)$ is considered

$$\mathbf{E} = -j\omega\mu\,\mathbf{A} - \nabla\,\varphi. \tag{4.5B.1}$$

Vector potential **A** and scalar potential φ of elementary planar currents and charges physically represent contributions of sources to the electric field in a given point.

From the mathematical point of view, potentials are the solution of Sommersfeld integrals. For simple situations, potentials can be composed on the basis of the physical point of view. This approach is used even here.

Assume a two-dimensional antenna element over the ground plane (fig. **4.5B.1**). Relative permittivity of the dielectrics is considered (for this moment) to be one, and contribution of an elementary current facet and charge one in the position **r'** to the intensity in **r** is computed. Using Coulomb (resp. Biot-Savart) low and respecting influence of ground plane (mirror principle), we get

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\left[\frac{\exp(-jkr_0)}{r_0} - \frac{\exp(-jkr_1)}{r_1}\right]dr',$$

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\left[\frac{\exp(-jkr_0)}{r_0} - \frac{\exp(-jkr_1)}{r_1}\right]dr',$$

where the distances r_0 and r_1 are

$$r_0 = \sqrt{(x-x')^2 + (y-y')^2},$$

$$r_1 = \sqrt{(x-x')^2 + (y-y')^2 + (2h)^2}.$$

Fig. 4.5B.1 *Coordinate system for computing contributions of elementary current (resp. charge) elements*

Here $\mathbf{r} = (x, y)$, $\mathbf{r'} = (x', y')$ and h denotes the height of the substrate.

Including the influence of the dielectrics is more complicated. The physical notion tells us that wave can propagate to the observation point **r** by the infinite number of paths:

$$\varphi(\mathbf{r}) = \frac{1-\eta}{4\pi\varepsilon}\left[\frac{\exp(-jk_0 r_0)}{r_0} - (1+\eta)\sum_{i=1}^{\infty}\frac{(-\eta)^{i-1}\exp(-jk_0 r_i)}{r_i}\right],$$

where

$$\eta = \frac{\varepsilon_r - 1}{\varepsilon_r + 1}, \qquad r_i = \sqrt{x^2 + y^2 + (2ih)^2}$$

and ε_r is relative permittivity of the substrate.

Vector potential **A** stays unchanged.

In the case of sources of finite dimension (rectangular cells), the contribution is expressed in integral way

$$\mathbf{A}(\mathbf{r}) = \iint_{p\check{r}es\,plochu\,bu\check{n}ky} \mathbf{G}_A\big(|\mathbf{r}-\mathbf{r}'|\big)\mathbf{J}(\mathbf{r}')\,dS', \tag{4.5B.2}$$
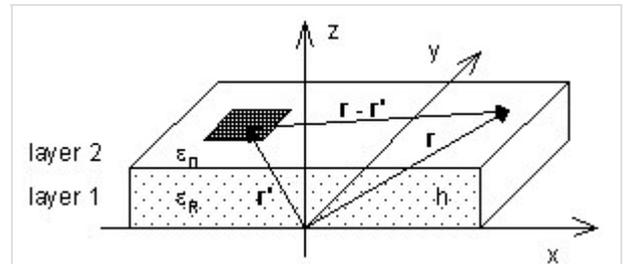
$$\varphi(\mathbf{r}) = \iint_{p\check{r}es\,plochu\,bu\check{n}ky} \mathbf{G}_q\big(|\mathbf{r}-\mathbf{r}'|\big)q(\mathbf{r}')\,dS', \tag{4.5B.3}$$

where $\mathbf{G}_A$ and $\mathbf{G}_q$ are Green functions for magnetic potential and electric one, σ is charge density.

Now, eqn. (4.5B.2) is discretized. In order to meet this task, a contribution of an arbitrary current (charge) cell to the potential in the center of another cell has to be expressed. Therefore, we denote

$$m_{ij}^{Ax} = \iint_{p\check{r}es\,plochu\,i-t\acute{e}bu\check{n}ky} G_{xx}^A\big(|\mathbf{r}_j-\mathbf{r}'|\big)J_x(\mathbf{r}')\,dS', \tag{4.5B.4a}$$
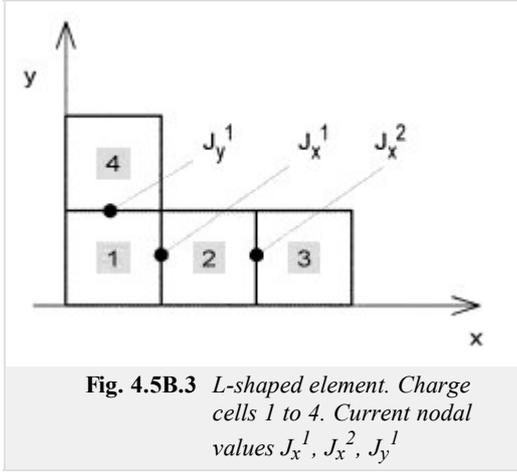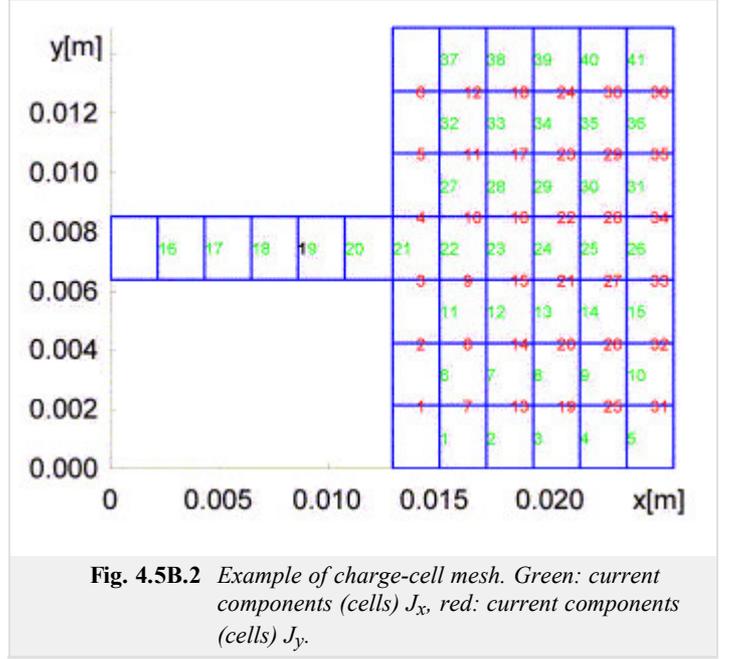
$$m_{ij}^{Ay} = \iint_{p\check{r}es\,plochu\,i-t\acute{e}bu\check{n}ky} G_{yy}^A\big(|\mathbf{r}_j-\mathbf{r}'|\big)J_y(\mathbf{r}')\,dS', \tag{4.5B.4b}$$

$$m_{ij}^{\varphi} = \iint\limits_{p\check{r}es\,plochui-t\acute{e}bu\check{n}ky} G^q\left(\left|\mathbf{r}_j - \mathbf{r}'\right|\right)\sigma(\mathbf{r}')\,dS'. \tag{4.5B.5}$$

Expression m_{ij}^{Ax} tells how j-th current cell for x-component contributes to vector potential in the center of i-th cell. Similar meaning is observed at m_{ij}^{Ay} for J_y and at m_{ij}^{φ} for scalar potential. As shown in the layer A, meshes of current cells and charge ones are mutually shifted for one half of a cell, and hence, a different numbering is used. Moreover, we independently number current cells for x-component and y-component. This fact is illustrated by numbering in fig. **4.5B.2**, which corresponds to the structure in fig. **4.5A.3**.

Fig. 4.5B.2 *Example of charge-cell mesh. Green: current components (cells) J_x, red: current components (cells) J_y.*

Fig. 4.5B.3 *L-shaped element. Charge cells 1 to 4. Current nodal values J_x^1, J_x^2, J_y^1*

Composing equations for unknown current distribution, we have to know a function describing contributions between two arbitrary cells (of the same type). Due to the limited space, we do not describe this function here. We state only that all the contributions are evaluated at the beginning of the program. Moreover, several contributions repeat (e.g., cells 1 and 2 for the component J_x are of the same contribution as cells 3 and 4). The mutual contributions are stored in a moment table (matrix). Calling the function, a contribution on a respective position of the table is read.

Now, the set of equations can be composed. We do not introduce a general relation, but a given simple situation is described. Fig. **4.5B.3** contains a L-shaped antenna element, consisting of four charge cells and three current cells (two of them are related to J_x, one of them to J_y).

Next, equations for E_1^x are built:

$$E_1^x = -j\omega\sum_{j=1}^{3} m_{1j}^{Ax} - \frac{1}{\Delta}\left[-\frac{j}{\omega}\frac{J_1^x - 0}{\Delta x} + \frac{J_1^y - 0}{\Delta y}\left(m_{21}^{\varphi} - m_{11}^{\varphi}\right) - \frac{j}{\omega}\frac{0 - J_2^x}{\Delta x} + \frac{0 - 0}{\Delta y}\left(m_{23}^{\varphi} - m_{13}^{\varphi}\right) - \frac{j}{\omega}\frac{0 - 0}{\Delta x} + \frac{0 - J_2^y}{\Delta y}\left(m_{24}^{\varphi} - m_{14}^{\varphi}\right)\right].$$

The first term represents the contribution of all current cells to the current cell J_x^1. The second term represents the contribution of charge cells; expression $\{\}/\Delta x$ denotes replacing the derivative $\partial\varphi/\partial x$ by the central difference (see par. 4.1). Brackets contain three terms since there are four charge cells. E.g., the first term represents the contribution of the first charge cell to the center of the second cell minus the contribution of the first charge cell to the center of the first cell. I.e., the first index at m denotes the center of the observation charge cell, where the contribution is computed, and the second index denotes the source cell. Charge in each cell is evaluated using nodal current densities and exploiting continuity equation

$$\nabla \cdot \mathbf{J} = -j\omega\sigma, \tag{4.5B.6}$$

where J[A/m] and σ[C/m^2] are current density and charge density.

Similarly, equations for E_2^x and E_1^y can be composed. Next, we turn our attention to feeding the structure.
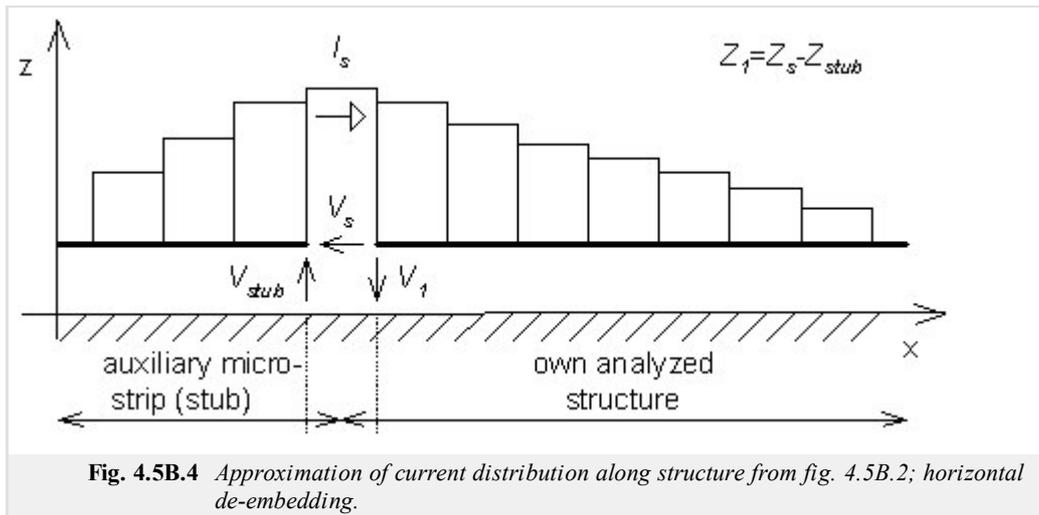
Plane wave feeding is the simplest way of feeding (see par. 4.4). For each current cell, we substitute $E_x = -E_x^I$ and $E_y = -E_y^I$, where E_x^I and E_y^I represent components of incident wave. In our situation, the antenna is assumed to be fed by a voltage source, which is connected between the analyzed structure and an auxiliary microstrip (the incident wave is not considered). In detail, the situation from fig. **4.5B.2** is depicted in fig. **4.5B.4**.

In the position, where 1 denoted the input port in fig. **4.5B.2**, a voltage source of a voltage V_S is considered. The source supplies the structure by the current I_S. One pole of the source is connected to the analyzed structure, the second pole is connected to the auxiliary microstrip (*stub*), which represents a certain impedance Z_{stub} related to ground. Input impedance Z_1 is then given by the relation

$$Z_1 = Z_S - Z_{stub}, \tag{4.5B.7}$$

where the impedance Z_S is determined as the ratio of the voltage V_S (its value is eligible) and the current I_S. The impedance Z_{stub} is computed

analytically as $-Z_0 \cot(\beta \, l_{stub})$.

Fig. 4.5B.4 *Approximation of current distribution along structure from fig. 4.5B.2; horizontal de-embedding.*

If impedance is computed by the above-described way, the result does not correspond to the reality because influence of edge capacitances is not considered (see fig. **4.5B.5**).

Fig. 4.5B.5 *Edge capacitances*
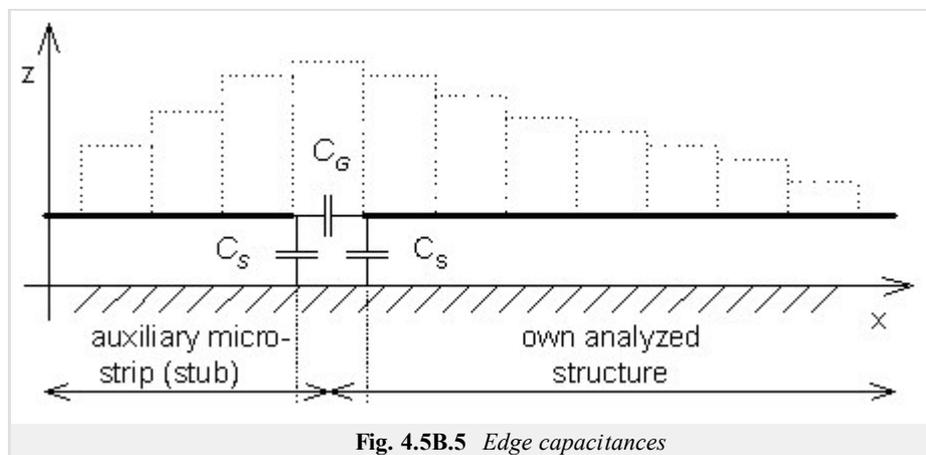
Respecting capacitances, the impedance Z_{stub} has to be added in parallel to $Z_S = 1/j\omega \, C_S$ (C_S is not open-end capacitance, but it is smaller). That way, we get $Z_{stub}' = Z_{stub} \parallel Z_S$. Capacitance C_G is already included in the impedance Z_S and computing impedance $Z_1' = Z_S - Z_{stub}'$, we have to subtract Z_S. The final impedance is $Z_1'' = Z_1' \, Z_S / (Z_S - Z_1')$.

Finally, we give a note about feeding the structure.

We can assume the structure to be fed by a voltage source connected between the microstrip and ground. This approach causes the direct computation of input impedance from the voltage and the current to provide results, which differ from the measured value. The reason of this difference is hidden in the fact that long cables are used for measuring, and therefore, the TEM wave (forward one, backward one) is present in the point of interest. If feeding is implemented by the above-described source connected between the microstrip and ground (localized feeding), then field distribution at the positive pole does not correspond to the field caused by TEM waves, but end effect influences results. In practical life, the result of localized feeding has to be transformed to the long-cable feeding (vertical de-embedding). The current standing wave has to be de-composed to the forward wave and the backward one. De-composition is not performed at the source port, but is shifted to the right. If both the waves are known, then reflection coefficient can be evaluated at the position of measurement. Then, the phase of reflection coefficient is shifted to the computation position.

The set of equations for unknown nodal values was built respecting physical principles. Now, the mathematical approach, which corresponds to the above-described notions, is given:

- Current density is approximated in terms of basic functions

$$J_X(x, y) = \sum_{n=1}^{N_X} J_{X_n} B_{X_n} \quad J_Y(x, y) = \sum_{n=1}^{N_y} J_{Y_n} B_{Y_n}$$

In our approach piecewise constant basis functions are used. As another possibility, *rooftops* can be used (linear in the first direction, constant in the second direction).

- Integral expression of potentials based on Green functions is substituted to (4.5B.1)

$$\left\{ -j\omega \iint\limits_{\Omega} \mathbf{G}_A(x-x', y-y')\mathbf{J}(x', y')\,dx'dy' \right\} + \left\{ -\nabla \left[\iint\limits_{\Omega} \mathbf{G}_q(x-x', y-y')\rho(x', y')\,dx'dy' \right] \right\} + \mathbf{E}^I = 0.$$

Here, W denotes the metallic area and E^I is intensity of incident wave. In most cases, the structure is fed via microwave ports and E^I is zero. Analyzing selective surfaces e.g., plane-wave feeding is assumed.

Charge is substituted from continuity equation and current density is expressed as a combination of basis functions. That way, we obtain

$$\tilde{J}_x^{(n)}(x, y) = \sum_{nx=1}^{N_x} J_x^{(nx)} \Pi^{(nx)}(x, y)$$

$$\left\{ -\frac{1}{j\omega} \frac{\partial}{\partial x} \iint\limits_{\Omega} G^q(x-x', y-y') \left[\frac{\partial}{\partial x} \sum_{n=1}^{N_x} J_{X_n}(x', y') + \frac{\partial}{\partial y} \sum_{n=1}^{N_y} J_{Y_n}(x', y') \right] dx'dy' \right\} + E_X^I = 0,$$

$$\left\{ +j\omega \iint\limits_{\Omega} \mathbf{G}_{yy}^A(x-x', y-y') \sum_{n=1}^{N_y} J_{Y_n}(x', y')\,dx'dy' \right\} -$$

$$\left\{ -\frac{1}{j\omega} \frac{\partial}{\partial y} \iint\limits_{\Omega} G^q(x-x', y-y') \left[\frac{\partial}{\partial x} \sum_{n=1}^{N_x} J_{X_n}(x', y') + \frac{\partial}{\partial y} \sum_{n=1}^{N_y} J_{Y_n}(x', y') \right] dx'dy' \right\} + E_Y^I = 0.$$

- Exploiting Galerkin method, current approximation error is minimized. Then, the following set of equations is obtained:

$$\sum_n J_n^X \left\{ \langle T_{xm}, G_{xx}^A * B_{xn} \rangle + \frac{1}{\omega^2} \langle T_{xm}, \frac{\partial}{\partial x} \left[G_q * \frac{\partial}{\partial x} B_{X_n} \right] \rangle \right\} + \sum_n J_n^Y \left\{ \frac{1}{\omega^2} \langle T_{xm}, \frac{\partial}{\partial x} \left[G_q * \frac{\partial}{\partial y} B_{yn} \right] \rangle \right\} = 0,$$

$$\sum_n J_n^X \left\{ \langle T_{ym}, G_{yy}^A * B_{yn} \rangle + \frac{1}{\omega^2} \langle T_{ym}, \frac{\partial}{\partial y} \left[G_q * \frac{\partial}{\partial y} B_{yn} \right] \rangle \right\} + \sum_n J_n^X \left\{ \frac{1}{\omega^2} \langle T_{ym}, \frac{\partial}{\partial y} \left[G_q * \frac{\partial}{\partial x} B_{xn} \right] \rangle \right\} = 0$$

.

Inner integrals are of convolution nature, and therefore, they are denoted by asterisk. Outer integrals are of correlation nature, and hence, they are denoted by brackets $\Leftrightarrow$. The symbol T denotes weighting functions (in Galerkin method, they are identical with basis functions).

From the mathematical point of view, Dirac pulses play the role of weighting functions in our formulation of moment method. This approach is called collocation method.

- All the integrals are efficiently evaluated and set of equations is composed. Solving it, unknown nodal values of current density are obtained.

- On the basis of known current distribution, other antenna parameters can be computed (reflection coefficient at a given port, input impedance, etc.).

In the layer C, the patch antenna analysis by the moment method is illustrated by an example.

4.5 Microstrip antenna

Matlab program

I. Analysis of standalone antenna element by transmission line method

On the basis of the theory given in the layer A, a microstrip antenna element is designed, which resonates on frequency f = 2.4 GHz. The antenna is going to be built on the substrate of the height d = 1.57mm and of the relative permittivity ε_r = 2.33. At the resonant frequency, input impedance 120 Ω.

The design consists of the following steps:

1. Using relation $W = (120/R_{rez}) \times (\lambda_0/2)$ (the layer A, 4.5A.2), width of the element is computed: W = 62.5 mm.

2. In order to determine the resonant length of the antenna element, effective permittivity has to be computed. Utilizing the program Serenade 8.5SV, we obtain ε_{eff} = 2.26.
   If effective permittivity is known, then the resonant length of the antenna element can be computed: $L = 0.48\lambda_0 \times (\varepsilon_{eff})^{-1/2} = 0.48 \times 0.125 \times (2.26)^{-1/2}$ = 40 mm.

3. Next, characteristic impedance Z_C of a microstrip transmission line of the width W. Tu is evaluated utilizing the program Serenade 8.5SV; we obtain Z_C = 5.2 Ω
   In case of a wide microstrip, characteristic impedance is approximately given by

$$Z_c = Z_0 \frac{d}{W\sqrt{\varepsilon_r}} = 377 \frac{1{,}57}{62{,}5\sqrt{2{,}33}} = 6{,}2\,\Omega. \qquad (\,4.5C.1\,)$$

Exploiting a Matlab program (click here) for download), following results are obtained:
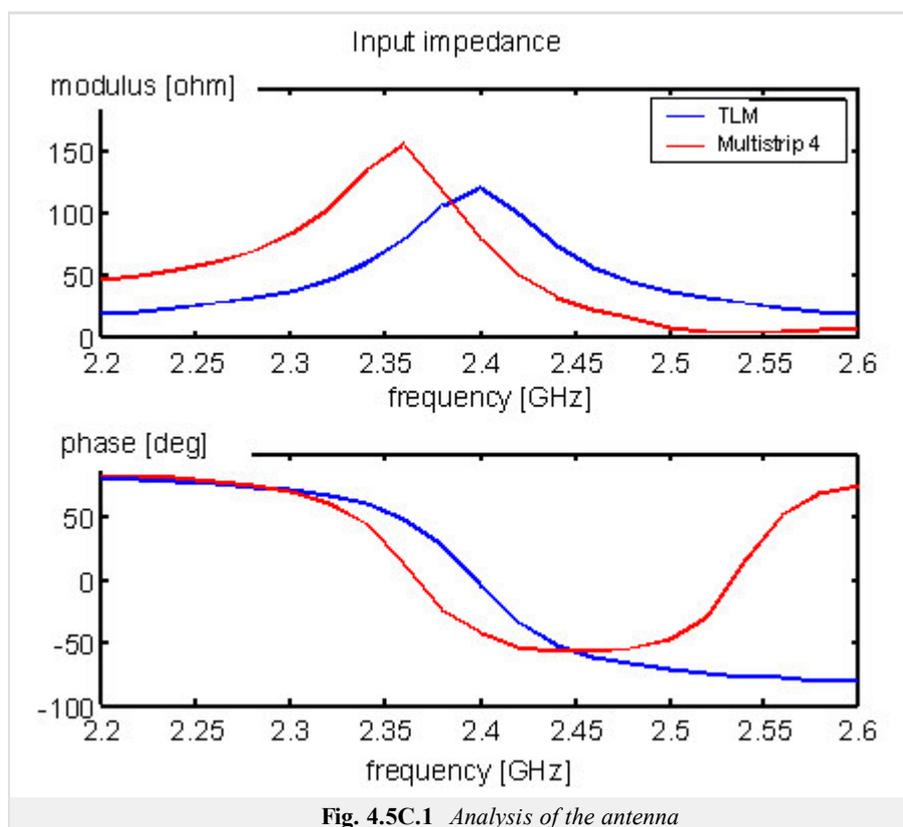
Fig. 4.5C.1 *Analysis of the antenna*

Input impedance, which is produced by the program Multistrip, is depicted by red color in the Figure as a reference. The real resonant frequency is shifted for 2% with respect to the desired one. Magnitude of the resonant resistance is for 25% lower than the desired one. The computed frequency course is not valid for higher frequencies because the simple transmission line model models the basic resonance and its multiples only.

In order to demonstrate the behavior of the antenna element, input impedance is depicted for wider frequency range in the next Figure. Next parallel resonance can be observed on the frequency 3.1 GHz, which corresponds to the double length of the antenna element $2W$ being equal to the wavelength in dielectrics. On the antenna, a vertical current mode is excited.

**Fig. 4.5C.2** *Analysis of the antenna in wider frequency range*
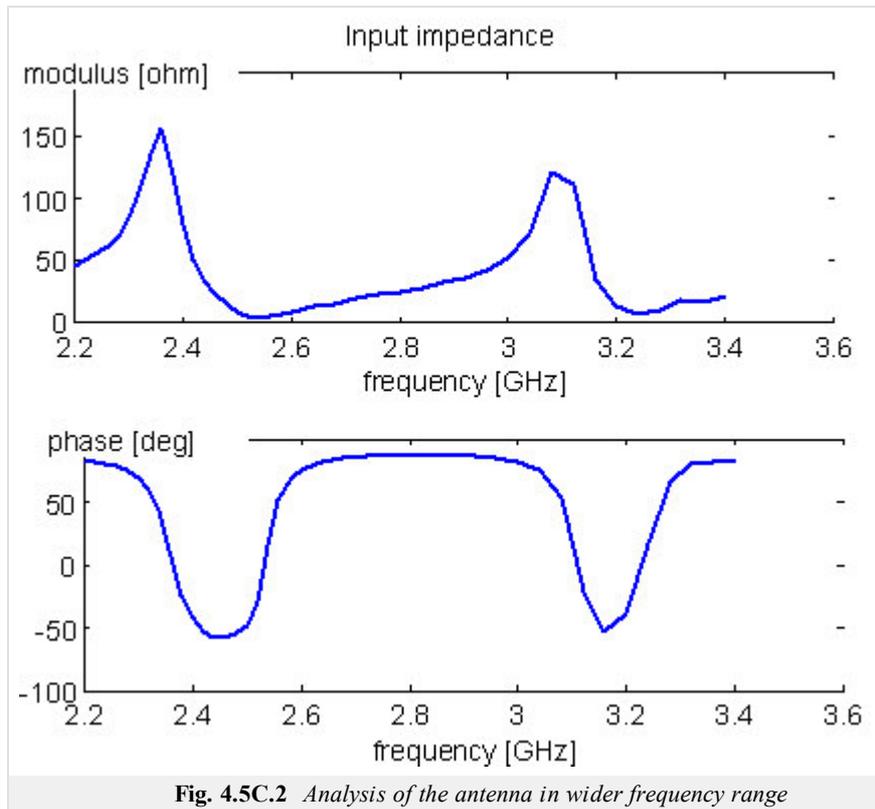
II. Analysis of standalone antenna element by moment method

This paragraph is aimed to the comparison of results provided by moment method (collocation) and reference results. As a reference, the program Multistrip is used.

As testing example, the rectangular antenna element 1.5 × 1.5 mm was used. The mesh of charge cells and comparison of computed reflection coefficient s_{11} with reference one are given in the following Figure. Accuracy of computing the resonant resistance is about 25% due to the approximate model of dielectrics and utilization of the collocation method. Further refinement of the mesh does not cause any increase of accuracy.

Fig. 4.5C.3 *Mesh of charge cells for the analysis of a patch antenna*
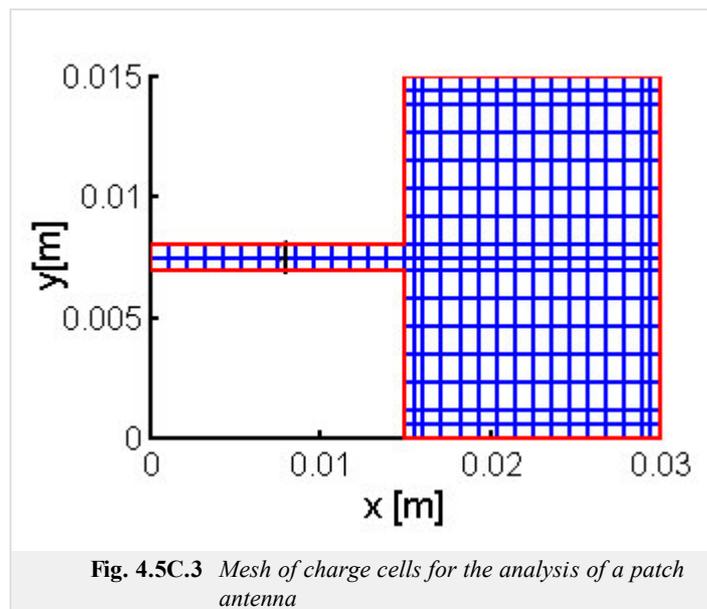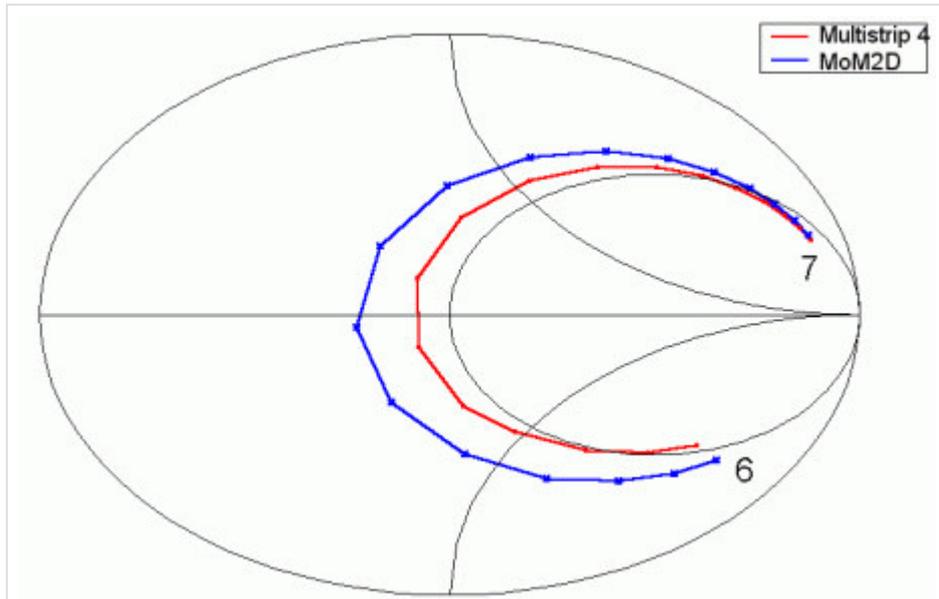
Fig. 4.5C.4 *Reflection coefficient of a patch antenna depicted in Smith chart*

4.5 Microstrip antenna

Java applet

4.5 Microstrip antenna

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**10** of **10**:

Question 1

The aperture-fed patch antenna consists of …

Possible answers for question 1:

- ○ … the patch and the coaxial probe in the role of feeding.
- ○ … the patch and the aperture of the horn antenna in the role of the primarily radiator.
- ○ … the microstrip on the bottom, the slot in the middle, and the patch on the top.
- ◉ I won't answer this question

Question 2

The patch antenna excels in …

Possible answers for question 2:

- ○ … low profile and low fabrication costs.
- ○ … low fabrication costs, and very high efficiency.
- ○ … high gain.
- ◉ I won't answer this question

Question 3

The simplified analysis of the patch antenna …

Possible answers for question 3:

- ○ … considers the patch only – the antenna behaves like a wide dipole.
- ○ … assumes the antenna behaves like a dielectric resonator (non-radiating structure).
- ○ … cannot be performed – full-wave methods have to be applied.
- ◉ I won't answer this question

Question 4

Surface waves degrade antenna parameters …

Possible answers for question 4:

- ○ … in case the dielectric constant of the substrate is high, and the frequency is high also.
- ○ … in case the height of the substrate is negligible compared to the wavelength.
- ○ … any time – they cannot be satisfactorily suppressed.
- ◉ I won't answer this question

Question 5

Increasing the dielectric constant of the substrate …

⊙ … the antenna is smaller, and its parameters are not influenced.

⊙ … the antenna is smaller, and the bandwidth is narrower.

⊙ … the bandwidth is narrower, the dimensions are not influenced.

◉ I won't answer this question

Question 6

Microstrip feeding of the microstrip antenna …

Possible answers for question 6:

⊙ … is hard to be fabricated.

⊙ … does not influence the radiation pattern of the antenna.

⊙ … simplifies the construction of antenna arrays.

◉ I won't answer this question

Question 7

Dealing with the dimensions of the rectangular patch …

Possible answers for question 7:

⊙ … the width should equal to half wavelength on the substrate, and the length is unimportant.

⊙ … both the width and the length should equal to half wavelength on the substrate.

⊙ … the length should equal to half wavelength on the substrate, and the width is unimportant.

◉ I won't answer this question

Question 8

Green functions for layered media …

Possible answers for question 8:

⊙ … can be exactly expressed in the closed form.

⊙ … have to be approximated (limited validity) for the practical use.

⊙ … are not used when analyzing patch antennas.

◉ I won't answer this question

Question 9

Moment method analysis of the patch antenna …

Possible answers for question 9:

⊙ … can be performed in two dimensions when feeding is mapped to the plane of the patch.

⊙ … has to be carried out as the three-dimensional problem.

⊙ … is naturally a two-dimensional problem.

◉ I won't answer this question

Question 10

In the moment method analysis of the patch antenna …

Possible answers for question 10:

⊙ … a single mesh is used both for x-components and for y-components.

⊙ … no meshing is needed.

⊙ … two meshes have to be used (the first one for x-components, the second one for y-components).

◉ I won't answer this question

see result

4.6 Multi-band patch antennas

## Basic theory

Introduction

Telecommunication technology is one of technical fields with great interest in military, professional and also civil sphere. Due to commercial exploitation of complex electrical devices (e.g. mobile phones), new requirements on their properties appear all the time. Engineers are compelled to develop novel solutions they match the increasing demands on both their technical parameters and physical dimensions. Miniature antennas of modern communication devices have to be able to receive and transmit electromagnetic waves in different frequency bands related to different communication services.

The characteristics and analysis of single-band patch antenna was discussed in the previous chapter of the textbook. Now we turn our attention to patch antennas with multi-band properties. First, basic practices for obtaining multi-band resonances of patch antennas are introduced. Second, the modal analysis in the design of planar antennas is discussed. In layer B, example of a simple dual-band patch antenna and its computational model in Ansoft HFSS and CST Microwave Studio is presented. Via this chapter, we get a short overview about multi-band patch antennas and learn about their design. Beside this, we become familiar with two well-known simulation softwares.

Basic types of multi-band patch antennas

There are countless numbers of approaches how to obtain multi-band resonance of planar antennas. As example we can name the stacked patch configuration, patches with slots and cuts of different shapes, antennas with shorting pins or PIFAs (planar inverted-F antennas). The mentioned types of antennas are depicted in fig. **4.6A.1**.

Considering the polarization, multi-band antennas can have parallel or orthogonal polarization in working frequency ranges. The orthogonal polarization is preferred in systems with polarization diversity.
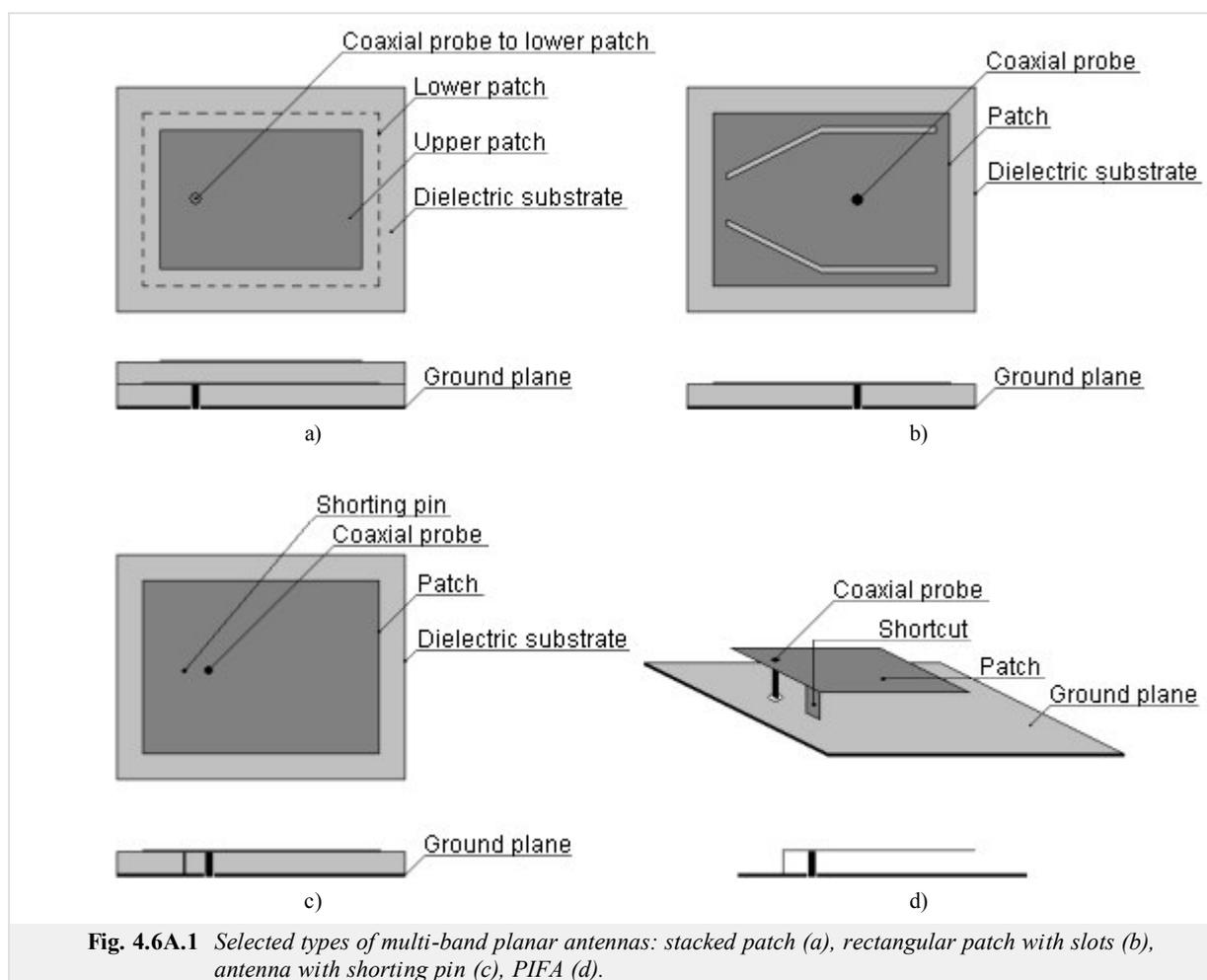
Fig. 4.6A.1 *Selected types of multi-band planar antennas: stacked patch (a), rectangular patch with slots (b), antenna with shorting pin (c), PIFA (d).*

Let us now describe the working principle of antennas from fig. **4.6A.1**.

In the case of stacked patch, the lower metallic patch is fed directly via connector (obviously coaxial probe) and the upper metallic patch is in parasitic coupling with the lower one. The configuration has two resonant frequencies – the lower one is given by the dimensions of the lower

(larger) metallic patch and the higher one is given by the dimensions of the upper (smaller) metallic patch. The basic operating mode of rectangular patch is TM_{10}.

The current distribution on rectangular patch can be influenced using slots and cuts of different shapes. As a result we obtain different antenna modes at the resonant frequencies. In the case of the structure in fig. **4.6A.1b** beside of the basic TM_{10} mode appears another resonant mode TM_{x0} ($1 < x < 2$) which radiation efficiency is the highest for band angle from 15 deg to 30 deg. The ratio of patch resonant frequencies f_1/f_2 extends from 1.29 to 1.60.

In the case of antenna with shorting pin good impedance matching can be obtained for the first two modes positioned along the resonant length of the antenna. If the shorting pin is placed in the centre of the patch, the resonant frequency f_1 of the basic TM_{10} mode remains not affected, whereas a new resonance emerges at the frequency $f_2 = 0.5 f_1$. That means such an antenna can works as half-wave and quarter-wave at the same time. Changing the position of the shorting pin, the ratio f_1/f_2 changes in interval of <2.0; 3.6>

PIFAs are widely used multi-band antennas mounted in a broad range of communication devices. Space between the ground plane and the patch is filed with dielectric. As in the previous case, good impedance matching in the required frequency bands is achieved using shorting pins and also cuts and slots (at different frequencies different portions of the patch are in resonance). PIFAs can be used also for three- and four-band applications, but the design of such antennas is rather than complicated.

Detailed description of the above-mentioned antennas and other variants can be found in the literature [33], [34]. To get fundamental physical insight into the working principle of multi-band planar antennas, the basics of the modal analysis is discussed in the next section. The modal analysis describes the field distribution of particular antenna modes. It simplifies the design process and final tuning because we are able to do a good estimation about the shape and dimensions of the metallic patch to achieve resonances at required frequencies.

Modal analysis

In the case of the modal analysis the planar antenna is modeled as a resonant cavity with perfect electric walls at the top and bottom. The main restriction is that the thickness of the dielectric substrate must be much smaller than the wavelength. Due to this, the variance of the electric field in the plane perpendicular to the patch (axis z) is negligible. Because of the applied boundary conditions the electric field is non-zero only along the axis z, whereas the magnetic field has non-zero components only in the x-y plane. Along the boundaries of the metallic patch the Neumann condition for electric field is valid. It expresses the fact that the derivative of the electric field intensity vector normal to the perfect magnetic conductor has to be zero ($\partial \mathbf{E}/\partial \mathbf{n} = 0$). It implies the boundaries of the patch can be considered as perfect magnetic conductor. The component of the electric field intensity perpendicular to the plane of the patch satisfies the scalar wave equation

$$\Delta E_z + k^2 E_z = 0. \qquad (4.6A.1)$$

The magnetic field vector **H** which is proportional to the current density **J** on the patch surface, can expressed as the gradient of the electric field intensity E_z using the equation
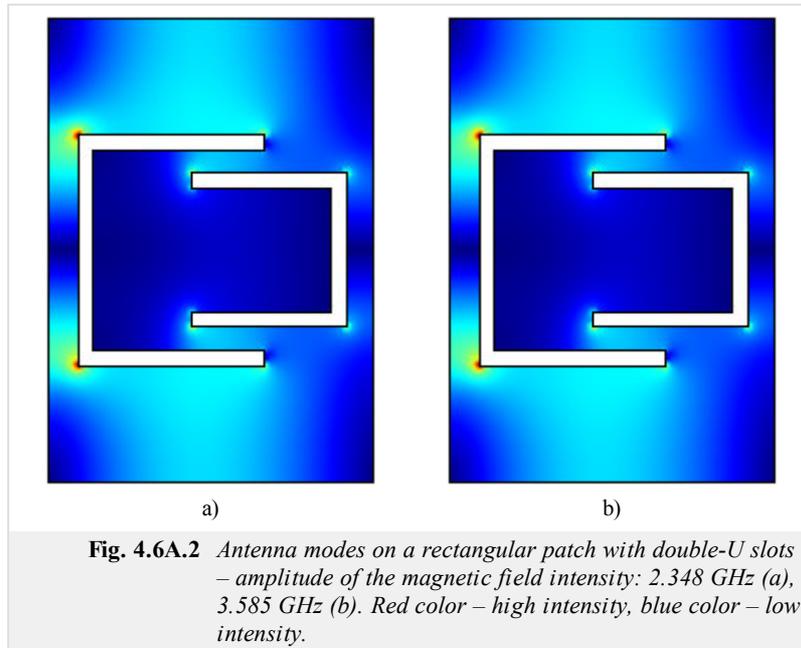
$$\mathbf{H} = \frac{1}{j\omega\mu_0}\, \mathbf{z}_0 \times \nabla\, E_z. \qquad (4.6A.2)$$

To solve the wave equation (4.6A.1), the finite element method software Comsol Multiphysics can be exploited. The patch is modeled as homogenous dielectric waveguide with the permittivity the same as the permittivity of the antenna substrate. The boundaries of the waveguide are considered to be perfect magnetic conductor. The cutoff frequencies of the computed transverse magnetic modes correspond to the antenna resonance frequencies (note that for this case it is necessary to set the value of the propagation constant near to zero, for example 0.001 rad.m^{-1}).

In fig. **4.6A.2** amplitude of the magnetic field intensity for a dual-band planar antenna with rectangular patch and double-U slot is depicted. The antenna is designed for frequency bands 2.45 GHz and 3.60 GHz. The working principle is as follows: at lower frequency resonates the whole patch, whereas at higher frequency only the smaller inner portion. In tab. **4.6A.1** the calculated resonant frequencies are summarized. Clearly, the third and fifth resonant frequencies are very close to the required ones. Parameters of the designed antenna are introduced in layer B together with the simulation results from Ansoft HFSS and CST Microwave Studio.

Tab. 4.6A.1 *Resonant frequencies of the antenna from fig.* **4.6A.2** *– results of the modal analysis.*

Resonance number	1.	2.	3.	4.	5.
Frequency [GHz]	1.055	1.100	**2.348**	2.985	**3.585**

Fig. 4.6A.2 *Antenna modes on a rectangular patch with double-U slots – amplitude of the magnetic field intensity: 2.348 GHz (a), 3.585 GHz (b). Red color – high intensity, blue color – low intensity.*

Using the above described modal analysis, basic characteristics of the designed antenna as resonator can be obtained in a very short time. That is because in Comsol Multiphysics we draw the longitudinally homogenous waveguide (i.e. our antenna) as a two-dimensional structure. This fact significantly reduces the computational time in comparison to three-dimensional model.

4.6 Multi-band patch antennas

Computer simulation

In this layer we are going to learn how to model planar antennas in the modern commercial softwares Ansoft HFSS and CST Microwave Studio. The first one, similarly to Comsol Multiphysics, is based on finite element method (FEM), and the second one uses the finite integration technique (FIT). The FIT is very similar to finite difference time domain (FDTD) method. The main difference is that FDTD solves differential Maxwell's equations, while FIT uses their integral form [35], [36].

Let us consider a planar antenna with the metallic patch from fig. **4.6B.2**. The patch is placed on dielectric substrate Arlon AD600 with relative permittivity $\varepsilon_r = 6.15$, thickness $h = 1.575$ mm and dimensions 50x50 mm^2. The antenna is fed by coaxial probe. A simple Matlab code for calculating the approximate size of patches is introduced in layer C. Parameters of the antenna after final tuning in Ansoft HFSS are summarized in fig. **4.6B.2**. The antenna was tuned manually and the surface current distribution was monitored. In such a way we successfully obtained good impedance matching ($S_{11} < -10$ dB) near to required frequencies 2.45 GHz and 3.60 GHz. Together with the results from Ansoft HFSS, computer simulations of the antenna also from CST Microwave Studio are introduced and mutually compared in the next section.

In fig. **4.6B.2** the computer model (together with the applied boundary conditions) of the antenna under investigation in the selected two softwares is depicted. In both cases, finite substrate and finite ground plane are considered.
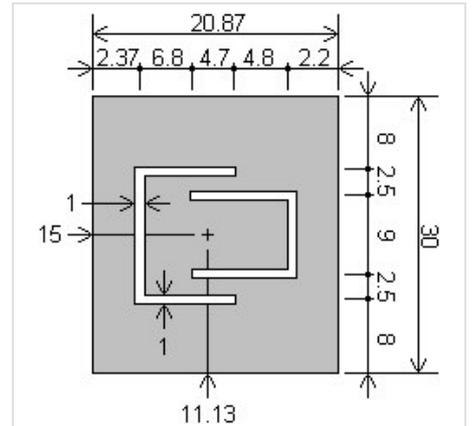

**Fig. 4.6B.1** *Parameters of the metallic patch of the designed dual-band antenna.*

a)   b)
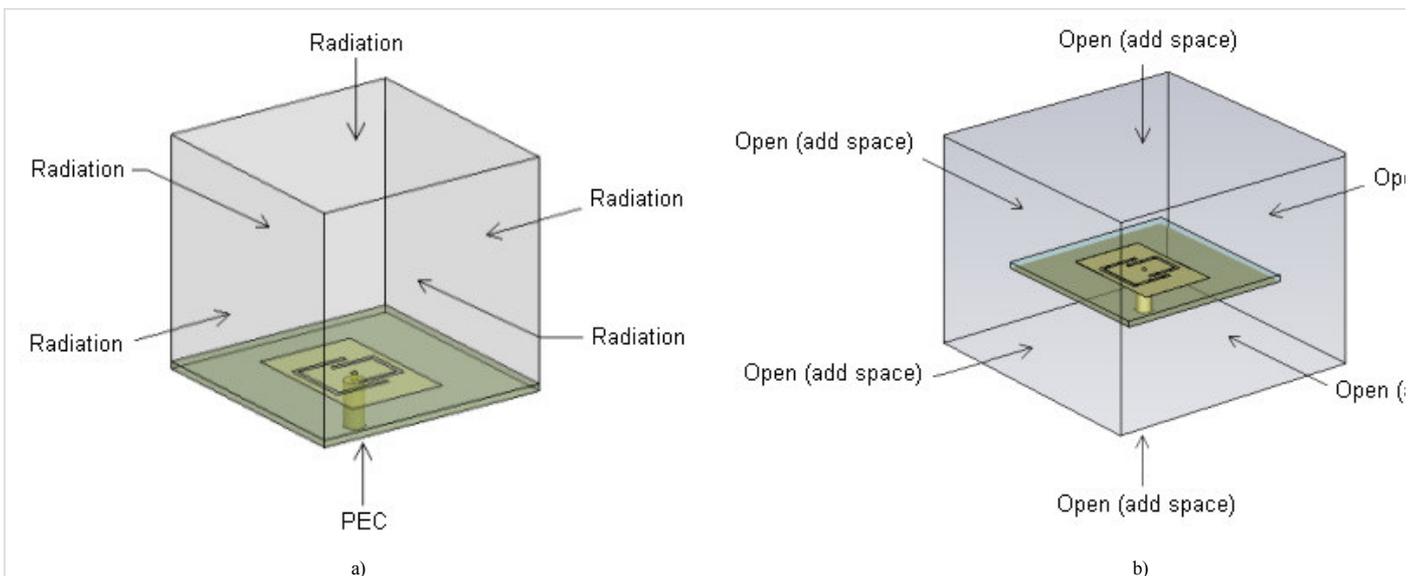
Fig. 4.6B.2 *Computer model of probe-fed planar antenna in Ansoft HFSS (a) and CST Microwave Studio (b). With arrows are show applied boundary*

Let us now investigate the frequency dependence of the input impedance of the antenna (fig. **4.6B.3**). Clearly, resonant frequencies computed by Ansoft HFSS are f_1 = 2.439 GHz and f_2 = 3.600 GHz, and the ones calculated by CST Microwave Studio are f_1 = 2.452 GHz and f_2 = 3.588 GHz. Because of the resonant frequencies obtained by different programs are in good agreement, our computer models seem to be correct.

Fig. **4.6B.3** shows the surface current density on the metallic patch of the antenna. The results are the same as computed in Comsol Multiphysics – at the lower resonant frequency currents flow on the whole patch, whereas at the higher resonant frequency only in the area bounded by the U slots.
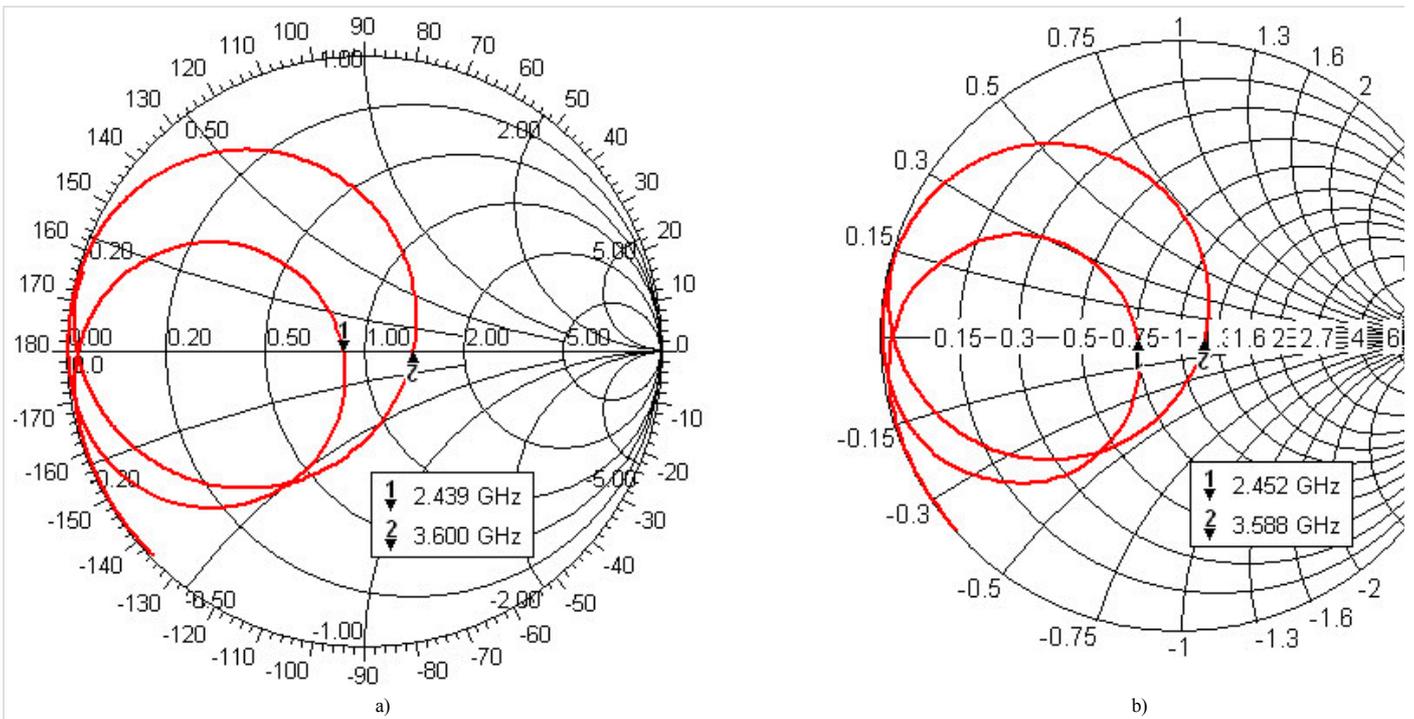
Fig. 4.6B.3 *Input impedance of the designed dual-band patch antenna in the Smith chart (from 2 GHz to 4 GHz): Ansoft HFSS (a), CST Microwave St...*
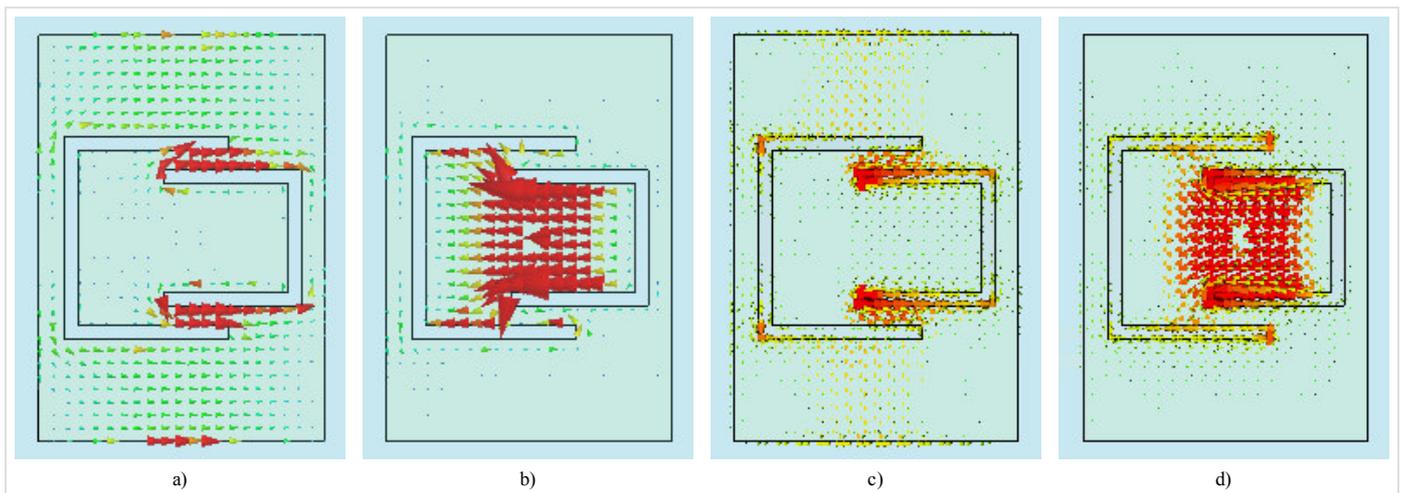
Fig. 4.6B.4 *Surface current density on the designed dual-band antenna: Ansoft HFSS, f_1 = 2.439 GHz (a), Ansoft HFSS, f_2 = 3.600 GHz (b), CST Microwave Studio, f_1 = 2.452 GHz (c), CST Microwave Studio, f_2 = 3.588 GHz (d). Red color – high density, blue color – low density.*

Radiation patterns of the designed antenna are shown in fig. **4.6B.5**.
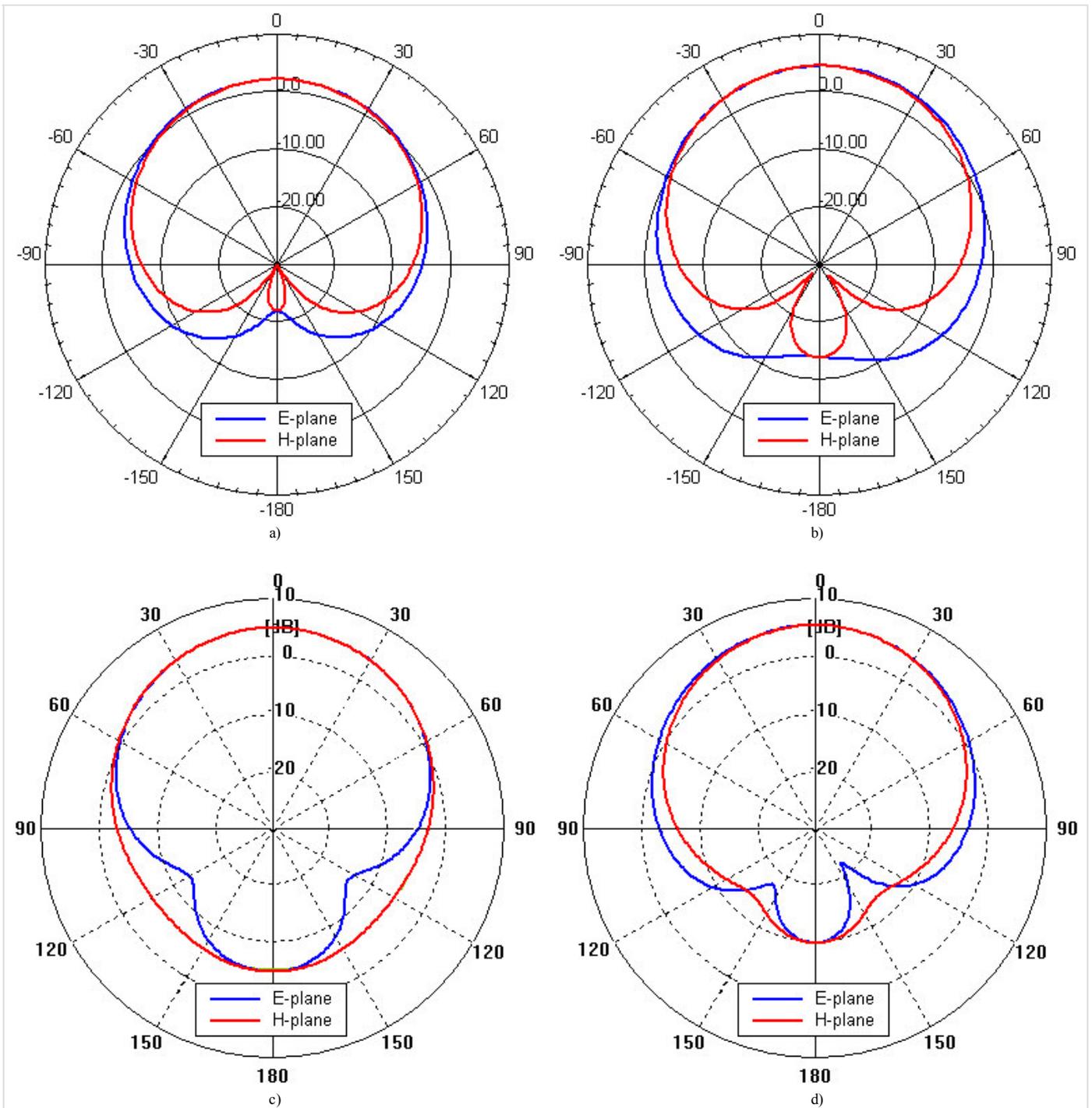
Fig. 4.6B.5 *Radiation pattern of the designed dual-band antenna: Ansoft HFSS, f_1 = 2,439 GHz (a), Ansoft HFSS, f_2 = 3,600 GHz (b), CST Microwave Studio, f_1 = 2,452 GHz (c), CST Microwave Studio, f_2 = 3,588 GHz (d).*

Now the properties of our antenna are described completely. Comparison of the results from the two softwares follows.

A simple dual-band planar antenna was modeled in Ansoft HFSS and CST Microwave Studio. The first of the softwares solves Maxwell's equations using FEM, the second one uses FIT. In both cases the antenna was placed on finite dielectric slab and finite ground plane with dimensions of 50x50 mm². Resonant frequencies obtained by different programs agree well and the surface current density corresponds to the one computed in Comsol Multiphysics. Small differences in radiation patterns computed by the two programs are probably caused by slightly different boundary condition setup as seen in fig. **4.6B.2**. Comparing the radiation patterns for E-plane and zero elevation at 2.45 GHz and 3.60 GHz, larger beamwidth at the higher resonant frequency can be observed. This indicates the existence of surface waves. At higher frequencies when thickness of the substrate becomes comparable with the wavelength, strong excitation of surface waves can lead to their diffraction at edges of finite dielectric slab and deform the radiation pattern.

4.6 Multi-band patch antennas

Matlab program

Program (patchant.m) calculates the approximate size of metallic patch and position of coaxial probe for planar antenna. The input parameters are the frequency f [GHz], thickness of the dielectric slab h [mm], relative permittivity eps [-] and input impedance Ri [Ohm].

4.6 Multi-band patch antennas

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

Polarization of an antenna is given by the orientation of …

Possible answers for question 1:
- ○ … E-plane.
- ○ … H-plane.
- ○ … Poynting vector.
- ◉ I won't answer this question

Question 2

If the transmitter antenna is polarized vertically and the receiver antenna is polarized horizontally, then the signal level at the connector of the receiver antenna (considering zero cross-polarization) is …

Possible answers for question 2:
- ○ … the same as for parallel polarized antennas.
- ○ … 3 dB less than for parallel polarized antennas.
- ○ … zero.
- ◉ I won't answer this question

Question 3

Dirichlet condition says that …

Possible answers for question 3:
- ○ … the component of electric field intensity tangential to the perfect electric conductor (PEC), is zero.
- ○ … the component of electric field intensity perpendicular to the perfect electric conductor (PEC), is zero.
- ○ … the component of magnetic field intensity tangential to the perfect electric conductor (PEC), is zero.
- ◉ I won't answer this question

Question 4

In the Smith chart, a circle with centre [1, 0j] and point [0.5, 0j] on its perimeter circumscribes the area where …

Possible answers for question 4:
- ○ … the input impedance of the antenna is purely real.
- ○ … the antenna is in resonance.
- ○ … the standing wave ratio at the antenna connector is smaller or equal to 2.
- ◉ I won't answer this question

Question 5

Diffraction of surface waves at the edges of dielectric slab results in …

◯ … enhancement of antenna efficiency.

◯ … enhancement of antenna bandwidth.

◯ … deformations of directivity pattern, deterioration of front-to-back ratio, lower antenna gain, etc.

◉ I won't answer this question

see result

◯ … enhancement of antenna efficiency.

◯ … enhancement of antenna bandwidth.

◯ … deformations of directivity pattern, deterioration of front-to-back ratio, lower antenna gain, etc.

◉ I won't answer this question

4.7 Planar ultra wideband antennas

Basic theory

Ultra wide band technology

Systems with extreme bandwidth are one of the prospective systems of modern high-capacity radio communications. Originally, this technology has been designed for radar sensing. Due to the wide frequency band radar had the opportunity to "see" beyond barriers such as wooded sections, etc. For their resistance to interference, security against eavesdropping and small power intensity began to be used for data transmission.

Today, a wide range of applications from the replacement of cables between multimedia devices such as camcorders, digital cameras and portable MP3 player through the networking of computers and peripherals, high-speed wireless universal serial bus (WUSB) to replace the cables in the third generation mobile phones. With the use of UWB technology are calculated primarily for very fast networks with short range (10m), where the need for large data throughput.

Speed data transmission in broadband technology can deliver hundreds of Mb/s. Another advantage of UWB technology is the absence of the intermediate frequency modulator and demodulator and low power levels, which allows coexistence with other technologies in the same frequency band. With low power level is reduced to detection of these signals, allowing widespread use in military applications, high security transmission, and interception virtually impossible confidentiality presence of wireless devices.

With regard to the bandwidth UWB (Ultra Wide Band) technology will interfere in the frequency bands other systems and it is necessary to define the spectral mask of UWB and to minimize interference with other systems. UWB technology often uses frequencies from 3.1 GHz to 10.6 GHz. This is the largest contiguous region with the highest power allowed in the defined spectral mask for UWB designated by the FCC (Federal Communications Commission). Each radio channel can have bandwidth more than 500 MHz, depending on its center frequency. Effective use of occupied frequency spectrum ensures the method of the overlay model or implementation of the ad hoc network access between nodes WPAN. Regulation of the frequency spectrum UWB due to the use the same spectrum of other radio services is not yet fully solved and in this dealt the standardizing group of IEEE 802.15.3 and IEEE 802.15.4 [40].

As already mentioned above, for this technology is characterized by a large bandwidth. In absolute terms we are talking about a minimum width of 500 MHz, in relative terms, the minimum width is 20% [39]:

$$\frac{B_f}{f_c} > 0,2 \qquad\qquad (\,4.7A.1\,)$$

where B_f is bandwidth for the observed decrease in quantity of 10dB and f_c is central frequency of the bandwidth.

Unlike other advanced radio technologies, broadband technologies are not harmonic and carrier information is encoded in a sequence of very short pulses (0.2 to 1.5 ns) [43]. Gaussian and Hermitian pulses are particularly used.

Gaussian pulse is described by the relation [39]

$$g_2(t) = K_3 \frac{-2}{\tau^2}\left(1 - \frac{2t^2}{\tau^2}\right) e^{\left(\frac{t}{\tau}\right)^2} \qquad\qquad (\,4.7A.2\,)$$

In relation (4.7A.2) K_3 denotes amplitude constant and τ is constant for changing pulse width.

The sequence of pulses, which is encoded, usually modulates on the carrier. The most commonly used modulation is PPM (Pulse Position Modulation), PAM (Pulse Amplitude Modulation), two-phase modulation, amplitude keying modulation and orthogonal [39]. This type of modulation practically eliminates distortion reflection, respectively, and multipath receiving.

Planar ultra wideband antennas

Planar antennas have significant advantages (small size, low profile and easy integrability into planar microwave circuits, low manufacturing cost in serial manufacturing), but when used in UWB applications, it is necessary to solve their shortcomings, which include mainly narrow impedance bandwidth, low polarization purity, low radiation efficiency. Increase bandwidth can be achieved by many different techniques, as will be shown below.

The main feature of broadband antennas is the minimal variation of the electrical parameters in a relatively wide frequency band. When designing an antenna, the emphasis is put on the stability of input impedance (the standing wave ratio should not exceed 2).

Antenna characteristics depend on the relative dimensions of the antenna to the wavelength. Theoretically, if the length of the antenna is infinite, then the antenna bandwidth can be infinite too. Since the implementation of such antennas is not possible, it is necessary to find another solution.

The most commonly used types of planar antennas for UWB technology:

- **patch antenna:** consists of conductive patch on one side of dielectric substrate and ground plate on the other. Basic shapes often used in practice are shown in fig. **4.7A.3a** Their radiation characteristics are symmetrical. These antennas usually have gain between 5 to 6 dB. Other forms patches (fig. **4.7A.3b**) are used for special applications **[39]**.

- **planar dipole:** differs from the rectangular patch antennas by its length to width ratio. Width of the dipole is usually less than 0.05 λ_0 (wavelength in free space). Radiation characteristics of dipole and patch are identical at the same longitudinal distribution of current. However, radiation resistance, bandwidth and level of cross polarization are different. Microstrip dipole antennas are attractive due to its properties, which are small size and linear polarization. Dipoles are very suitable for higher frequency and achieve significant bandwidths. As an example, a symmetric printed folded dipole can be given, which is composed of the dipole combined with another identical dipole (mirror image) and creates a symmetrical structure. This structure can be regarded as a rectangular patch with a slot in shape H. By $VSWR = 2$, the bandwidth of the dipole is around 16% **[39]**.

- **printed slot antenna:** consist of the slot board grounded in the underlying substrate. Slot can have practically any cross section. Theoretically, most of microstrip patch shapes can be realized in the form of printed slot. In practice, we used a few basic shapes

Fig. 4.7A.1 *Basic types of printed slot antennas.*
a) rectangular slot with microstrip feed,
b) rectangular slot with coplanar feed,
c) angular slot with microstrip feed,
d) angular slot with coplanar feed,
e) rectangular ring slot,
f) tapered slot [39].

of the slots only: a rectangular slot, a circular slot, a rectangular ring and the tapered conical slit (fig. **4.7A.1**). Like microstrip patch antennas, slot antennas can be coupled with microstrip or coplanar waveguide feed. Slot antennas are omnidirectional radiators (emitting of both sides of the slot). Radiation to one half-space can be obtained by using reflective plate on one side slits **[39]**.

- **travelling wave antennas:** can consist of concatenated sections of periodic microstrip lines or long segments of sufficient band width to support the propagation of transversely electric (TE) waves. End of the antenna with a traveling wave is terminated by resistive loads preventing the generation of standing waves on the antenna. Various configurations of antennas with a traveling wave are depicted in fig. **4.7A.2** **[39]**.

- **complementary (dual) antenna:** complementary pair created by two such antennas, which seized conductor position for the first antenna, is by the shape and dimensions identical by flat cut from the second antenna wire **[38]**.

**Fig. 4.7A.2** *Basic shapes of the travelling wave antennas [39].*
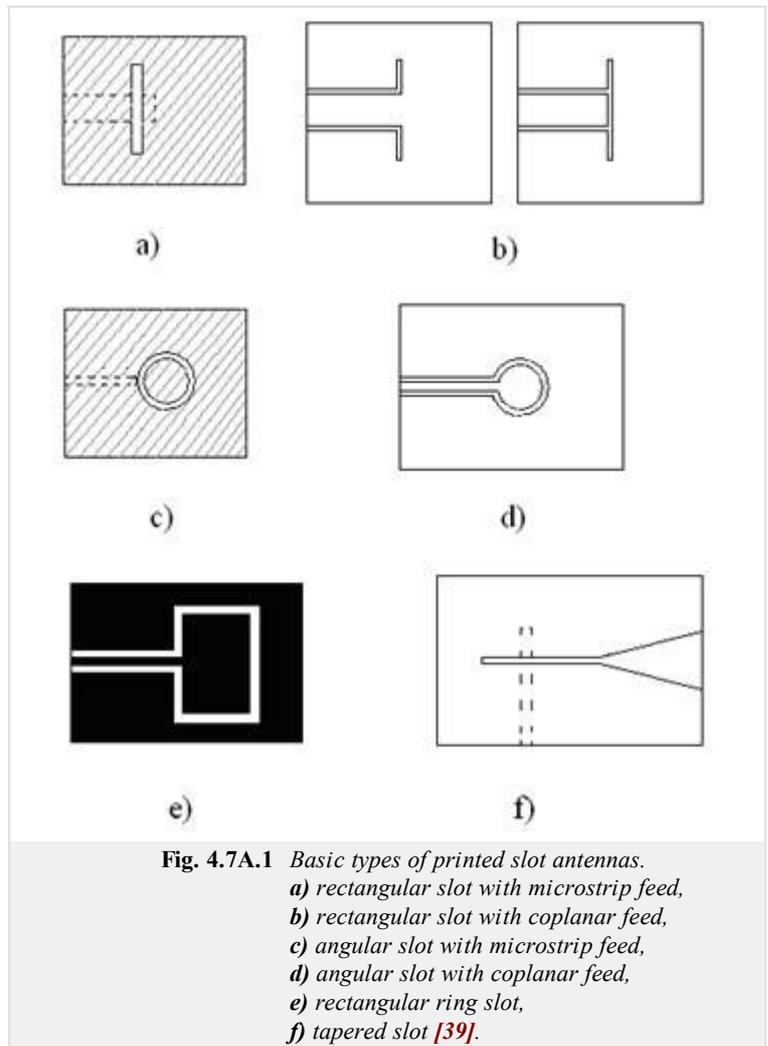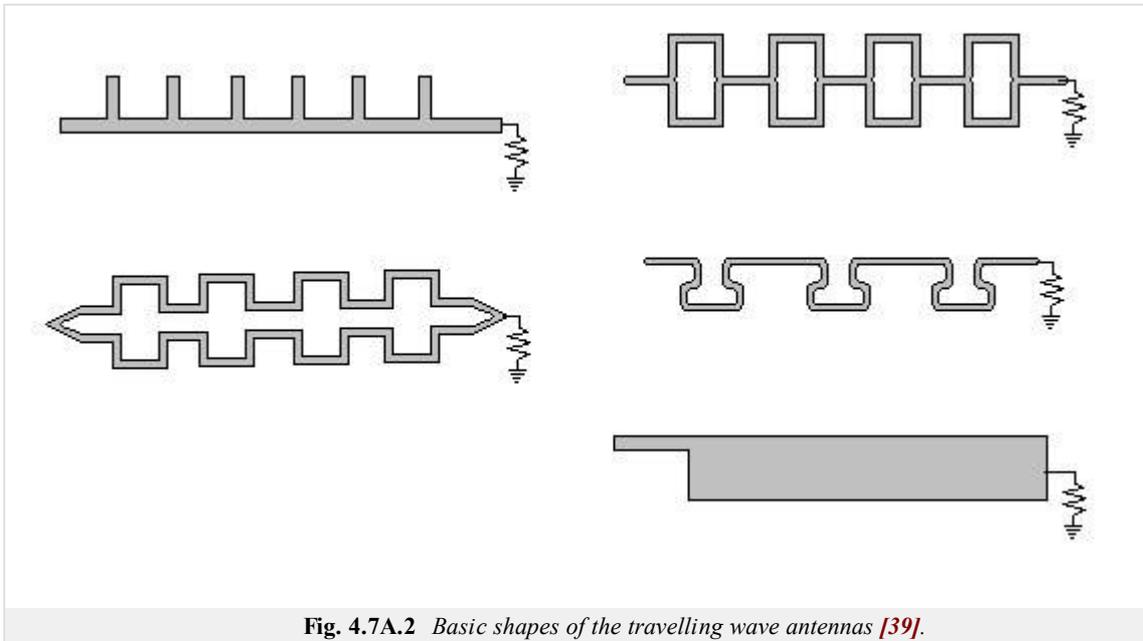
The definition of bandwidth for UWB planar antenna:

- **impedance bandwidth:** the range of frequencies where the antenna has good impedance matching and can be heard by standing wave ratio (*VSWR* 2 or 1.5) or module of the reflection coefficient (s_{11} -10 dB and -15 dB).

- **radiating bandwidth:** radiation pattern is the most important indicator of operating modes of antennas [41]. Radiation characteristics depend on the distribution of current on the antenna patch.

- **polarization and axial ratio bandwidth:** is defined as the maximum level of cross-polarization, or levels of axial ratio. Control of polarization of the antenna depends on the control of orthogonal modes excited in the linear and circularly polarized antennas. The difference between orthogonal modes sets the level of cross-polarization, or the level of axial ratio [41]. The level of axial ratio is influenced by the type of feed and the quality factor of the antenna.

Generally, the bandwidth of microstrip antennas depends on the shape of emitters (patch), on the properties of the substrate, on the feed type, on the value of quality factor of the antenna, on the excited multiple resonance and on the impedance matching. This is valid for broadband planar antennas also. Furthermore, the effects of the basic parameters of the planar antennas on the broadband behavior will be given.

Shape of the patch

A metallic patch of a given shape is etched on a dielectric substrate. A metallic layer should be well-conductive. Patch length is equal to half the wavelength on the substrate. Patch shape affects the distribution of current on the antenna and thus the radiation characteristics of antennas. Basic shapes of the patch antennas are shown in fig. **4.7A.3a**, the other used shapes of the patch antennas are shown in fig. **4.7A.3b**.

Fig. 4.7A.3a *Basic shapes of the patch antennas [39].*

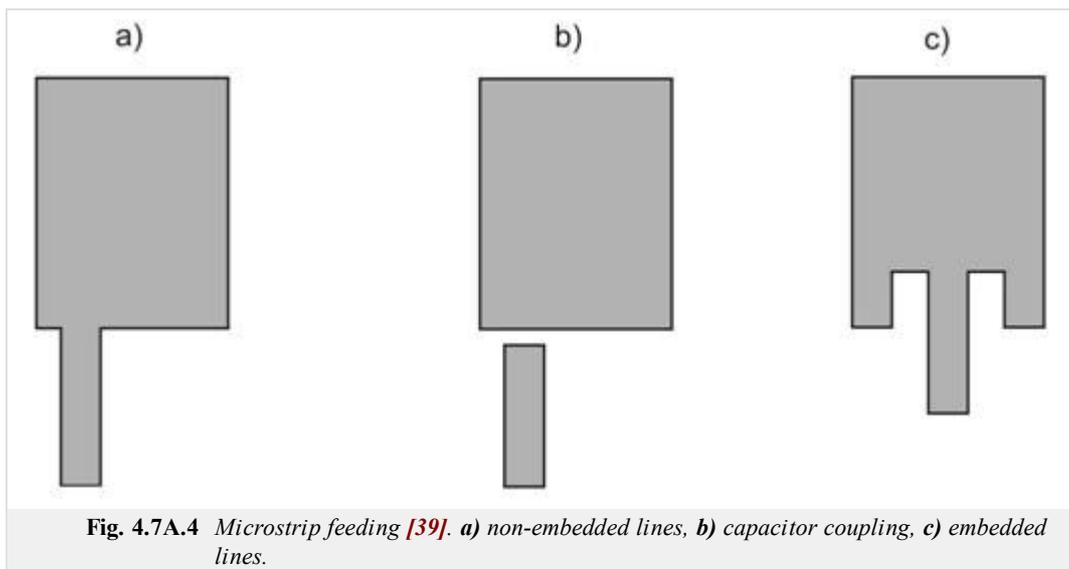**Fig. 4.7A.3b** *Less common shapes of the patch antennas [39].*
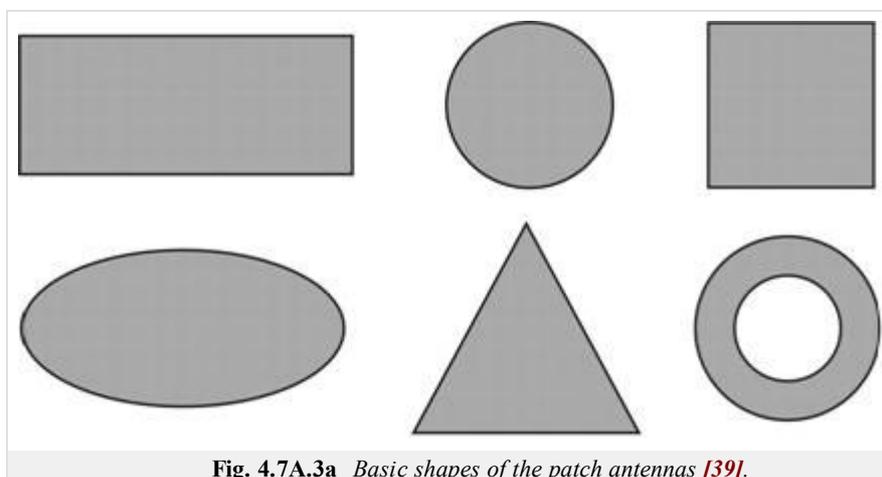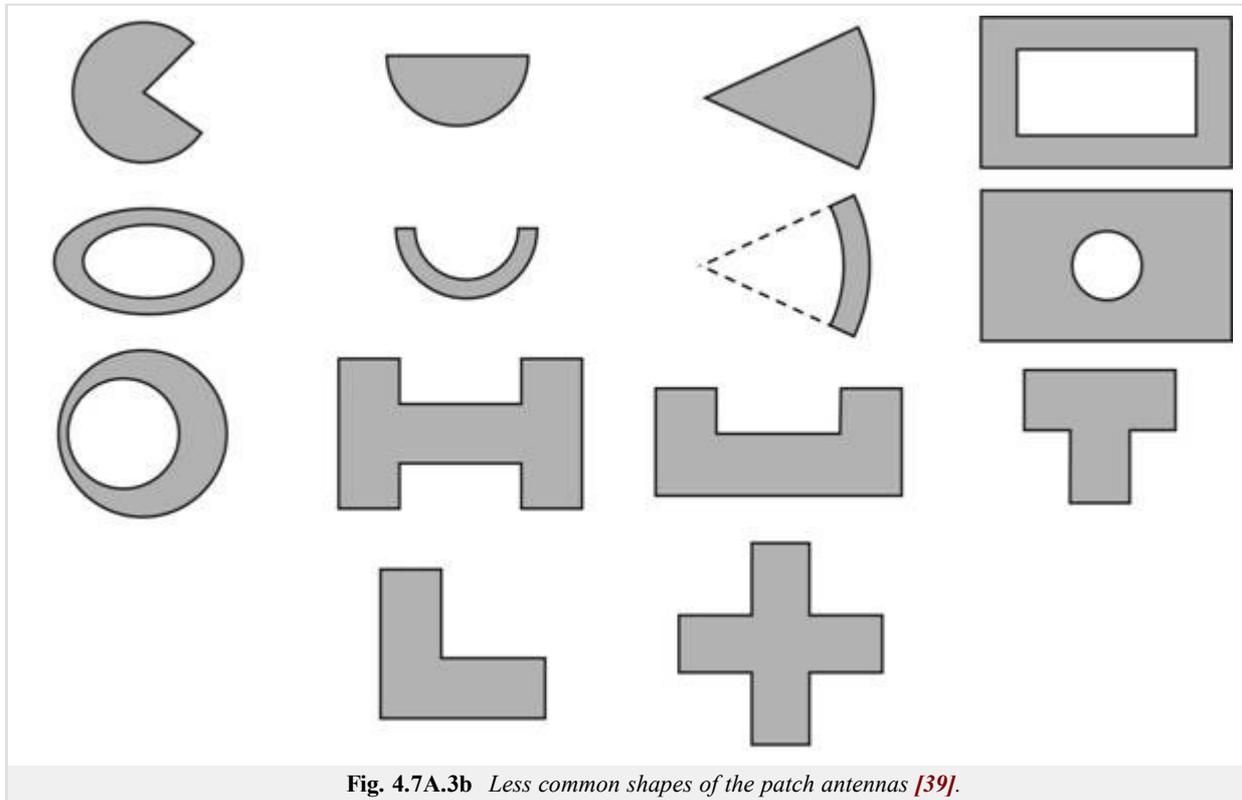
Dielectric substrate

Dielectric substrate used for microwave antennas is often of a relative permittivity of $2.2 \le \varepsilon_r \le 16$ and loss factor $0.0001 \le \text{tg } \delta \le 0.06$. High permittivity ε_r has usually the effect of reducing the radiation efficiency of the antenna. Thickness of the substrate is much smaller than the wavelength.

Feeding

Feeding structure affects the impedance matching, operating modes, parasitic radiation, the radiation of surface waves, and radiated power. The most commonly used feeding is coaxial probe (outer conductor is connected to ground plane, an internal wire is connected with antenna element) and microstrip feeding (fig. **4.7A.4**).

Fig. 4.7A.4 *Microstrip feeding [39]. **a)** non-embedded lines, **b)** capacitor coupling, **c)** embedded lines.*

In some applications, galvanic-separated coupling is used: an antenna is excited by field of the microstrip line (fig. **4.7A.5a**) or through the slot (fig. **4.7A.5b**) by the so-called aperture coupling.
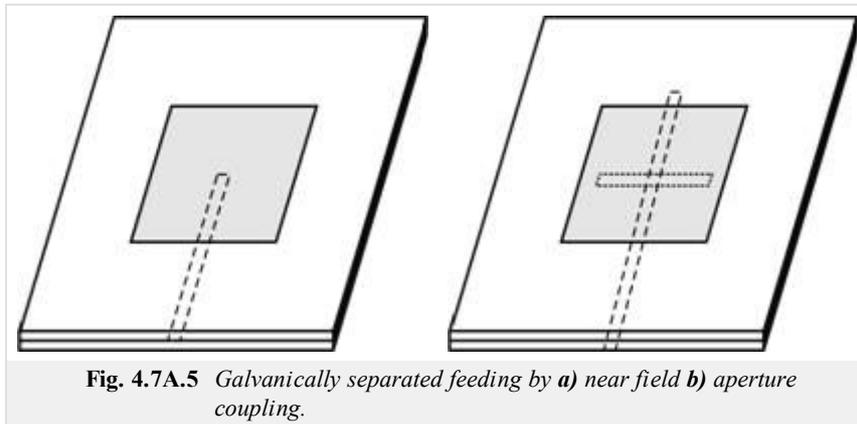
Fig. 4.7A.5 *Galvanically separated feeding by **a)** near field **b)** aperture coupling.*

By requirements for high bandwidth, the antennas are often coupled with coplanar feeding (fig. **4.7A.6**). Connection between patch and CPW can be inductive (fig. **4.7A.6a**) or capacitive one (fig. **4.7A.6b**). Backward radiation can be reduced by substitution of long direct slot over a circular loop (fig. **4.7A.6c**), the loop is located under the center of the patch **[39]**.
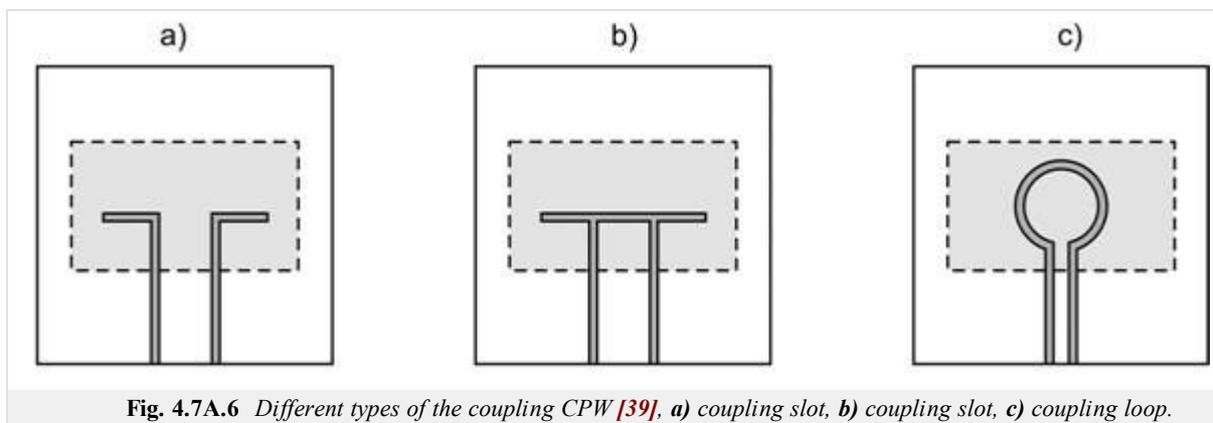

Fig. 4.7A.6 *Different types of the coupling CPW **[39]**, **a)** coupling slot, **b)** coupling slot, **c)** coupling loop.*

Decreasing of the quality coefficient

The broadband planar antennas can be considered as a circuit of a high quality factor. Increasing the bandwidth, quality factor of the antenna has to be reduced. Quality factor can be reduced by selecting appropriate patch shape (this affects operating modes, and thus radiation efficiency of antennas **[41]**), and a suitable choice of dielectric substrate (thick substrate with low relative permittivity extending the bandwidth).

Impedance matching

Feeder exhibits frequency-stable characteristic impedance; the input impedance of the antenna is frequency dependent. The contradiction can be solved by the insertion of a separate adaptive circuit - quarter wave impedance transformer (fig. **4.7A.7a**), tuning stubs (fig. **4.7A.7c,d**), combinations of them (fig. **4.7A.7b**) or adapting the patch shape by slots and notch (fig. **4.7A.8**).
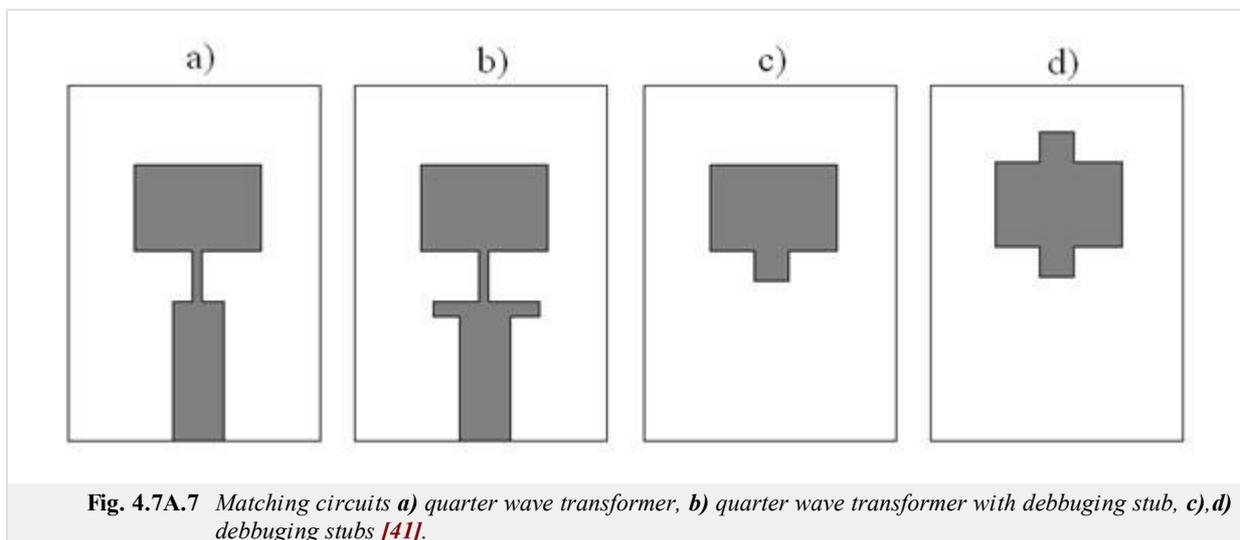
Fig. 4.7A.7 *Matching circuits **a)** quarter wave transformer, **b)** quarter wave transformer with debbuging stub, **c),d)** debbuging stubs **[41]**.*
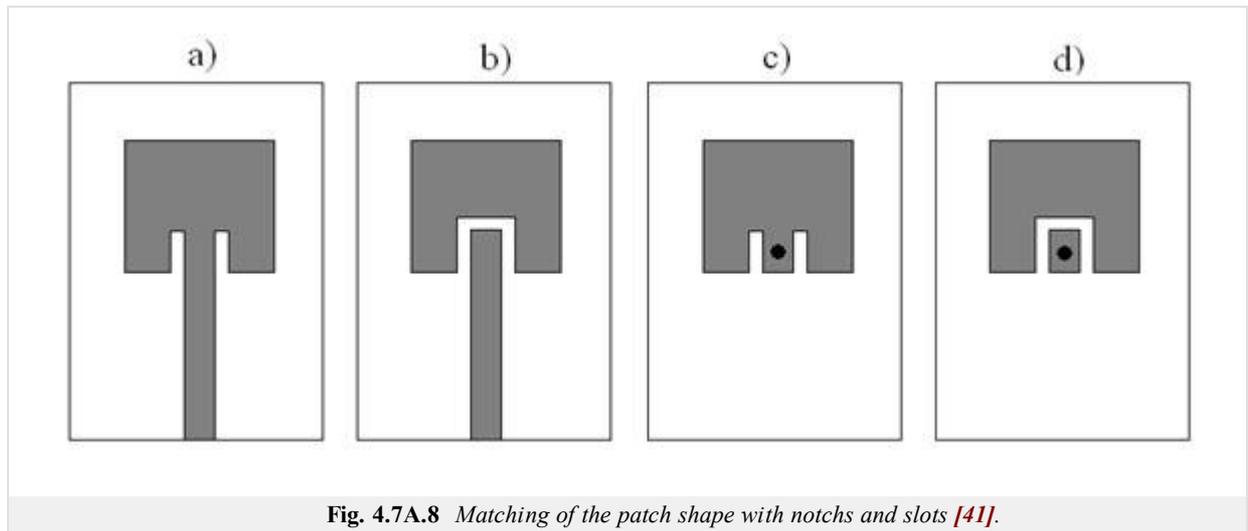
Fig. 4.7A.8 *Matching of the patch shape with notchs and slots [41].*

A well-matched antenna should cover the desired operating frequency range when keeping the defined level of the parameters which should vary marginally only (*VSWR* < 2, module of the reflection coefficient s_{11} < -10 dB, stable value of gain, the main lobe beam width and radiation pattern over the desired bandwidth).

In the layer B, a concrete example of the calculation of broadband planar antennas is mentioned. We show the influence of the patch shape and his size, the choice of coupling and choice of substrate on the characteristics of the antenna.

4.7 Planar ultra wideband antennas

Advanced theory

Design of broadband planar rectangular U-slot microstrip patch antenna

Planar UWB antennas are nowadays often used in high-frequency and high-speed data communications. Planar antennas are easily positioned in small mobile devices, and easy and cheap fabrication is their great advantage. Classical patch antennas have unfortunately very narrow operating frequency band and therefore it is necessary use the techniques to increase the frequency bandwidth. One of the techniques mentioned in the layer A uses notches in the original patch shape. This method is suitable for use in antenna arrays, because they do not increase the antenna size.

Fig. 4.7B.1 *Layout of the microstrip U- shape antenna [42].*
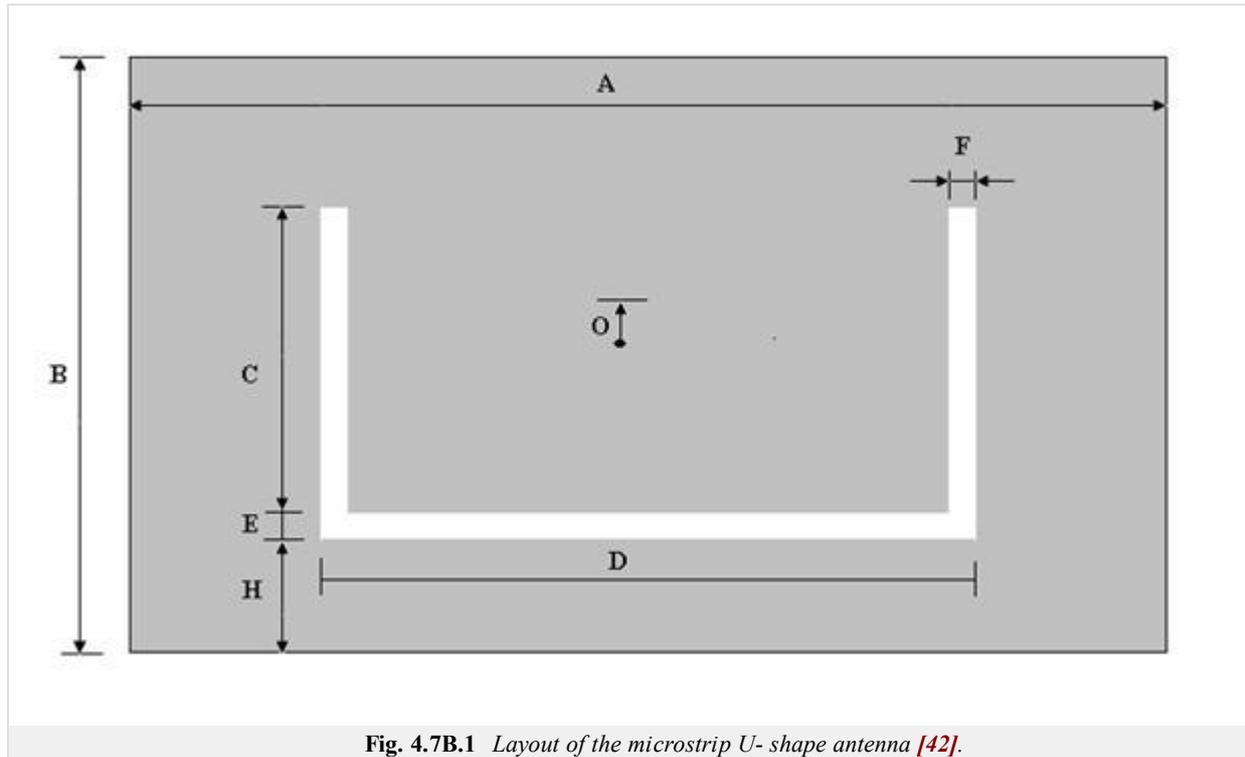
Step 1:

Specify the central frequency f_{res3} and desired frequency bandwidth (f_{res2} - f_{res4}) [42]. Antenna is designed to cover both the PCS and WLAN communication from 1.8 GHz to 2.5 GHz. The central frequency is 2.15 GHz.

f_{res2} = 1.8 GHz

f_{res3} = 2.15 GHz

f_{res4} = 2.5 GHz

Step 2:

Select a substrate permittivity (ε_r) and substrate thickness (T). For broadband antennas, the following rule should be satisfied:

$$T \geq 0.06 \frac{\lambda_{res3}(air)}{\sqrt{\varepsilon_r}},$$
(4.7B.1)

where λ_{res3} is wavelength in vacuum.

T = 6.35 mm

ε_r = 2.2

Step 3:

Estimate the quantity (B)

$$B + 2\Delta B \approx \frac{c_0}{2\sqrt{\varepsilon r} f_{res3}} \qquad (4.7B.2)$$

$B + 2\Delta B \approx 47.04$

Step 4:

Calculate of the width of the patch (A)

$$A = 1.5(B + 2\Delta B) \qquad (4.7B.3)$$

$A = 70.56$ mm

Step 5:

Calculate of the effective permitivity (ε_{eff}) and parameter $2\Delta B$

$$\varepsilon_{eff} = \frac{\varepsilon r + 1}{2} + \frac{\varepsilon r - 1}{2}\sqrt{\left(1 + \frac{12T}{A}\right)} \qquad (4.7B.4)$$

$$2\Delta B = 0.824T\,\frac{\left(\varepsilon_{eff} + 0.3\right)\left(\frac{A}{T} + 0.262\right)}{\left(\varepsilon_{eff} - 0.258\right)\left(\frac{A}{T} + 0.813\right)} \qquad (4.7B.5)$$

$\varepsilon_{eff} = 2.465$

$2\Delta B = 6.249$

Step 6:

Back calculate of the length of the patch (B)

$$B = \frac{c_0}{2\sqrt{\varepsilon_{eff}} f_{res3}} - 2\Delta B \qquad (4.7B.6)$$

$B = 38.19$ mm

Step 7:

Select a starting value of slot thickness (E, F)

$$E = F = \frac{\lambda_{res3}(air)}{60} \qquad (4.7B.7)$$

$E = F = 2.33$ mm

Step 8:

Calculate width of the U-slot (D)

$$D = \frac{c_0}{\sqrt{\varepsilon_{eff}} f_{res2}} - 2\left(B + 2\Delta B - E\right) \qquad (4.7B.8)$$

$D = 21.93$ mm

Step 9:

Select C such that

$$\frac{C}{A} \geq 0.3, \; \frac{C}{D} \geq 0.75 \qquad (4.7B.9)$$

$\frac{C}{A} \geq 0.33, \; \frac{C}{D} \geq 0.9$

Step 10:

Calculate the effective permittivity and effective length extension of the pseudo-patch of the fourth resonance with effective patch width as

D-2F

$$\varepsilon_{eff(PP)} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2} \sqrt{\left(1 + \frac{12T}{D - 2F}\right)} \qquad (4.7B.10)$$

$$2\Delta_{B-E-H} = 0.824T \frac{\left(\varepsilon_{eff(PP)} + 0.3\right)\left(\frac{D-2F}{T} + 0.262\right)}{\left(\varepsilon_{eff(PP)} - 0.258\right)\left(\frac{D-2F}{T} + 0.813\right)} \qquad (4.7B.11)$$

$\varepsilon_{\text{eff(PP)}} = 2.996$

$2\Delta_{\text{B-E-H}} = 5.32$

Step 11:

Estimate of the position U-slot

$$H \approx B - E + 2\Delta_{B-E-H} - \frac{1}{\sqrt{\varepsilon_{eff(PP)}}}\left(\frac{c_0}{f_{res4}} - (2C + D)\right) \qquad (4.7B.12)$$

For $C = 23.28$ $H \approx 11.41$

Step 12:

Check that sum $C + E + H$ is less than B. If not adjust C by changing the rations in step 9 and value of H until the design is physically realizable.

$$C + E + H \leq B \qquad (4.7B.13)$$

$23.28 + 2.33 + 11.41 \leq 38.19$

Tab. 4.7B.1 *Approximate values of the parameters calculated using the antenna design relations [42].*								
A [mm]	**B [mm]**	**C [mm]**	**D [mm]**	**E [mm]**	**F [mm]**	**O [mm]**	**T [mm]**	ε_r
70,56	38,19	23,28	21,93	2,33	2,33	0	6,53	2,2

4.7 Planar ultra wideband antennas

Matlab program

For an approximate design, parameters of a broadband antenna with a rectangular slots and U-shaped notch from the layer B are available here. The program runs in a command window by calling function

```
U_antena(f_res2, f_res3, f_res4, ε_r, T)
```

For the parameters of the function `U_antena` we substitute:

f_{res2} the lower limit of the desired frequency range [Hz]

f_{res3} central frequency in [Hz], $f_{res3} = f_{res2} + \frac{1}{2}\left(f_{res4} - f_{res2}\right)$

f_{res4} the higher limit of desired frequency range [Hz]

$ε_r$ desired permittivity substrate

T desired substrate thickness [mm]

Outputs are the approximate dimensions of the antenna with a rectangular slots and U-shaped notch as shown in figure **4.7B.1**.

4.7 Planar ultra wideband antennas

Developing Matlab

This program is designed to determine the approximate parameters of a broadband antenna. Mathematical relationships are given in layer B.

Select a substrate permittivity ε_r and substrate thickness T (4.7B.1)

```
T_test=0.06*(lambda_res3/sqrt(eps_r));

if T>= T_test
else
    T_test=T
end
```

Calculate of the width of the patch A (4.7B.2), (4.7B.3)

```
% Estimate the quantity B+2deltaB
B_dve_delta_B = c0/(2*(sqrt(eps_r))*fres3);

% Calculate of the width of the patch A
A=1.5*(B_dve_delta_B)
```

Back calculate of the length of the patch B (4.7B.4), (4.7B.5), (4.7B.6)

```
% Calculate of the effective permitivity eps_eff and parameter 2deltaB
eps_eff=((eps_r+1)/2)+((eps_r-1)/2)*sqrt(1+((12*T)/A));
dve_delta_B=0.824*T*((eps_eff+0.3)*((A/T)+0.262)/((eps_eff-0.258)*
((A/T)+0.813)));

% Backcalculate of the length of the patch B
B=(c0/(2*sqrt(eps_eff)*fres3))-dve_delta_B
```

Select a starting value of slot thickness E, F (4.7B.7), (4.7B.8)

```
% Select a starting value of slot thickness E and F
E=lambda_res3/60
F=E

% Calculate width of U-slot D
D=(c0/(sqrt(eps_eff)*fres2))-(2*(B+dve_delta_B-E))
```

Select C such that is be satisfied rule (4.7B.9)

```
% Select C such that
C1=0.33*A;
C2=0.9*D;
```

Estimate of the position U-slot (4.7B.10), (4.7B.11), (4.7B.12)

```
% Calculate the effective permittivity and effective length extension of
the pseudopatch of the fourth resonance with effective patch width as D-2F
eps_eff_PP=((eps_r+1)/2)+((eps_r-1)/2)*sqrt(1+((12*T)/(D-2*F)));
dve_delta_B_E_H=0.824*T*((eps_eff_PP+0.3)*(((D-2*F)/T)+0.262)/((eps_eff_PP-
0.258)*(((D-2*F)/T)+0.813)));

% Estimate of the position U-slot
H=B-E+dve_delta_B_E_H-(1/sqrt(eps_eff_PP))*((c0/fres4)-(2*C1+D))
```

For broadband antennas must be the condition (4.7B.13)

```
% Check that sum C+E+H is less than B. If not adjust C by changing the
rations in step 9 and value of H until the design is physically realizable.
if C1+E+H<=B
else
    C1=C2
    H=B-E+dve_delta_B_E_H-(1/sqrt(eps_eff_PP))*((c0/fres4)-(2*C1+D))
end
C=C1
```

4.7 Planar ultra wideband antennas

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

For what distance is calculated mainly with using of the broadband technology?

Possible answers for question 1:
- ○ tens of meters
- ○ hundreds of meters
- ○ tens of kilometers
- ● I won't answer this question

Question 2

What are the main advantages of broadband access?

Possible answers for question 2:
- ○ high speed data transfer, low consumption, high transmission security, resistance to interference
- ○ secure data transmission to long distances
- ○ using yet unused frequency bands
- ● I won't answer this question

Question 3

What are the basic advantages of planar antennas?

Possible answers for question 3:
- ○ usability for large transmission power
- ○ small size, low profile, easy integrability into planar microwave circuits
- ○ high polarization purity
- ● I won't answer this question

Question 4

What frequency bands are most often used for UWB technology?

Possible answers for question 4:
- ○ 2,5 – 6,7 GHz
- ○ 40 GHz and higher
- ○ 3,1 – 10,6 GHz
- ● I won't answer this question

Question 5

What are requirements to the dielectric substrate used in antenna technology?

Possible answers for question 5:

- ○ high permittivity
- ○ low permittivity
- ○ There aren't put any requirements. It does not matter.
- ● I won't answer this question

see result

Chapter 5: Time-domain modeling of antennas

Introduction

Numerical modeling of antennas in the frequency domain is effective if the antenna parameters are desirable in the narrow band of frequencies. If the wideband analysis is desirable, it is better to use time-domain analysis. However, this one is suitable only for the structures with low Q-factor, otherwise, the computed response is too long, and the analysis is time consuming.

For modeling in the time domain, an analyzed structure is excited by a desired pulse **[37]**, and a transient response is computed. From the excitation pulse and the computed response, we can obtain properties of the analyzed structure at those frequencies which the excitation pulse contains. It is wideband modeling which can be used for the wideband analysis of antennas, scatterers, or in the area of the electromagnetic compatibility (EMC), or electromagnetic interference (EMI).

The time-domain modeling has several advantages in the comparison to the frequency-domain one:

- The wideband response is obtained by one run of an analysis.
- The frequency-domain analysis can not be used for the time variant or nonlinear systems.
- By the time domain-analysis, only the beginning of the transient responses can be computed, and the rest of the responses can be ignored.

Due to the above advantages of the time domain modeling, the basic principles of the time-domain modeling will be described in this chapter. Again, as in the frequency domain, the formulations which come out from Maxwell's equations in the differential or integral form are solved.

5.1 Time-domain modeling of wire antennas by method of moments

Basic theory

In the chapters 4.1, 4.4, and 4.5 the wire dipole, microstrip dipole and patch antenna are modeled by method of moments in the frequency domain. The integral equation is transformed to the system of algebraic equations and the current distribution on the antenna is computed at the steady harmonic state at the given frequency. The antenna parameters can be computed from the current distribution.

Numerical modeling of antennas in the frequency domain is effective if the antenna parameters are desirable in the narrow band of frequencies. If the wideband analysis is desirable, it is better to use time-domain analysis. However, this one is suitable only for the structures with low Q-factor, otherwise, the computed response is too long, and the analysis is time consuming.

The basic principles of the time-domain modeling of wire antennas will be described in this chapter. Again, as in the frequency domain, the formulations which come out from Maxwell's equations in the differential or integral form are solved. In this chapter we focus our attention on the integral formulation, as in the chapters 4.1, 4.4 and 4.5.

The integral formulation in the time domain is more complicated than in the frequency domain due to the time derivative of the vector potential, thus, the integro-differential formulation is solved instead of the pure integral formulation.

Due to the education purpose of this chapter, we focus our attention only on the solution of the symmetric wire dipole, because the solution of more complicated structures is, from the point of the derivation and implementation, more demanding.

Time-Domain Modeling of Wire Dipole by Method of Moments

For modeling a symmetric wire dipole in the time domain we come out from a similar situation as in case the frequency domain. Let's suppose that the symmetric wire dipole is made from the perfect electric conductive material, and placed in the vacuum. The dipole axis is identical with z axis of the cylindrical coordinate system, and the electromagnetic plane wave incidents on it (fig. **5.1A.1**). The intensity of the electric field is parallel to z axis, and the waveform is arbitrary. Further, the wire radius a is much smaller than the wavelength.

The incident plane wave $\mathbf{E}^I$ induces a current I_z (the dipole is placed along z axis) on the dipole surface. Since at open ends of the antenna the electric currents do not have a chance to flow, the electric charge σ accumulates. In the next instants the accumulated charge flows off and in, depending on the phase of each spectral component of the incident wave. The induced currents and charges become the sources for the radiation of the secondary wave. The influence of the current along the z axis can be described by the vector potential [37]
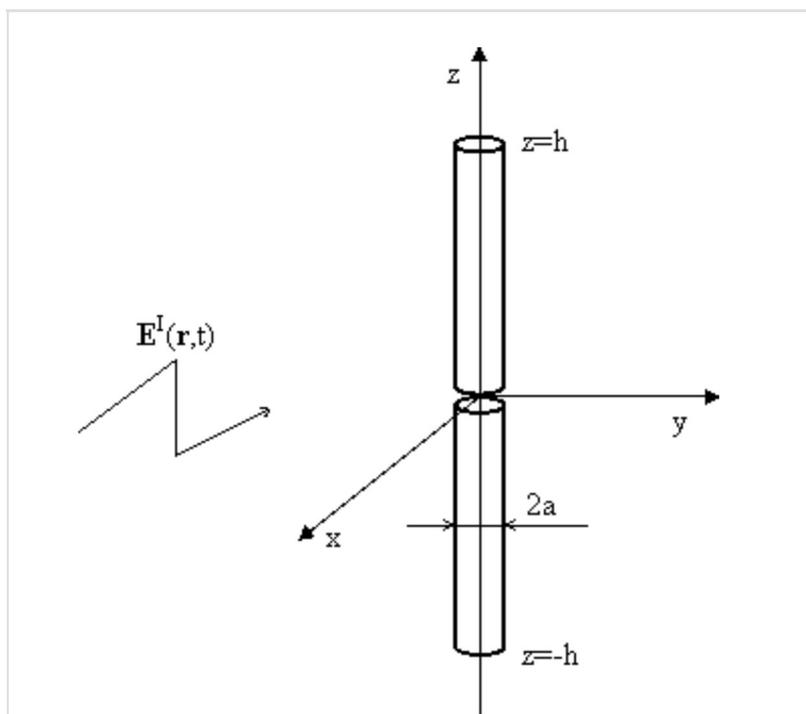

Fig. 5.1A.1 *Wire dipole placed along the z axis and excited by electromagnetic wave.*

$$A_z(z, t) \;=\; \frac{\mu}{4\pi} \int_{\xi=-h}^{h} \frac{I_z\big(\xi, \, t - R(z, \xi)/c\big)}{R(z, \xi)} \, d\xi. \qquad (\,5.1A.1\,)$$

The influence of the charge density can be described by the scalar potential

$$\varphi(z, t) \ = \ \frac{1}{4\pi\varepsilon} \int\limits_{\xi=-h}^{h} \frac{\sigma(\xi, t - R(z, \xi)/c)}{R(z, \xi)} d\xi. \qquad (5.1A.2)$$

The value of the vector potential A_z at a arbitrary "target" point z, and instant t can be computed by the addition of contributions of the current components from all "source" points ξ on the antenna axis. The source current has to be retarded about time R/c which the wave needs for traveling the distance between the source ξ point and the target point z

$$R(z, \xi) = \sqrt{a^2 + (z - \xi)^2}. \qquad (5.1A.3)$$

Further, c in (5.1A.1) is speed of light in the given environment. The contribution to the vector potential is not proportional to the distance between the source and target point. The integration has to be carried out along the whole antenna structure. Finally, the value of the integral is multiplied by the permeability of surrounding medium μ (the vacuum).

In case of the scalar potential, the current distribution is replaced by the charge density σ. The result of the integration is, instead of the multiplication by the permeability, divided by the permittivity of the surrounding medium ε. The rest in the equation (5.1A.2) stays the same as in (5.1A.1).

Since the current and the charge are mutually tied, the relations (5.1A.1) and (5.1A.2) has to be tied by the continuity equation

$$\frac{\partial \sigma(z, t)}{\partial t} = -\frac{\partial I_z(z, t)}{\partial z}. \qquad (5.1A.4)$$

If the vector and scalar potentials are known, the intensity of the electric field radiated by antenna E_z^S (the scattered field) can be computed

$$E_z^S(z, \ t) \ = \ - \ \frac{\partial A_z(z, t)}{\partial t} \ - \ \frac{\partial \varphi(z, t)}{\partial z} \ . \qquad (5.1A.5)$$

The derivative of the vector potential with respect to time describes "the dynamic contribution" of the electric charge on the antenna to the radiated wave (the electric currents, charges in motion, are the source for the vector potential). The partial derivative of the scalar potential with respect to the variable z express the "static contribution" of the electric charge on the antenna to radiated field (the static charges, accumulated at ends of the antenna, are the source for the scalar potential)

On the surface of the perfect electric conductor the intensity of the electric field has to satisfied the boundary condition

$$E_z^S(z, \ t) \ + E_z^I(z, \ t) \ = 0 \ . \qquad (5.1A.6)$$

The system of equations (5.1A.1) to (5.1A.6) is solved numerically in two steps. The first one consists in the space discretization of the problem and the second one consists in the application of time algorithms on the result of the first step.

The first step is analogical to the solution in the frequency domain, where the method of moments is used for the "transformation" of the integral equation to the system to algebraic equations. Here, we use again the method of moments for the "transformation" of the integro-differential equation (5.1A.5) to a system of the retarded differential equations (the space discretization of the problem).

For the second step (the application of the time algorithms), it is possible to use either the explicit, or the implicit algorithm. At the explicit one, the length of the time step Δt has to satisfy the following condition: $\Delta t \le R_{min}/c$, where R_{min} is the smallest distance of centers of the discretization elements and c is the speed of the light in the surrounding medium. At the implicit algorithm, the length of the time step can be chosen arbitrary. The larger length of the time step increase the speed of computation, but the accuracy falls down. These approaches belong to the marching on in time (MOT) methods.

The marching on in order method (MOO) is the next method. It is based on the approximation of the transient response by a set of weighted Laguerre polynomials. The advantage of MOO is its unconditional stability, but the disadvantages are the higher demands on the implementation, and less effectiveness in the comparison to MOT.

Due to above facts, we focus our attention on MOT. Although we mentioned that the system of equations (5.1A.1) to (5.1A.6) is solved in two steps (the space discretization and the application of time algorithms), we will merge both these steps. It is appropriate to notify that by MOT the system of retarded differential equations is transformed to the system of difference equations (the derivative of the vector potential with respect to tome (5.1A.5) is replaced by appropriate difference).

The advantage of the explicit algorithm is its easier implementation and there is no need to compute an inverse matrix as in case of the solution in the frequency domain, or as at the implicit approach. However, the explicit approach suffers often from the late time oscillation (the computed response is not steady, but grows exponentially.). Therefore, the attention in this chapter is focused only on the implicit approach.

Excitation pulse

Before, than we focus our attention on the implicit approach, let's choose an appropriate excitation pulse.

If we chose the Dirac pulse, the spectrum of the excitation signal would be equally spread over the infinite frequency domain. Since in the technical practice the solution is desirable over the limited bandwidth and we do not have unlimited computer memory and the power of a computer, thus, the Dirac pulse is not an ideal choice.

Instead of the Dirac pulse it is better to use the Gaussian pulse modulated by a harmonic signal. By change of its parameters, it is possible to analyze only the band of our interest.

The vector of the intensity of the electric field of the Gaussian excitation pulse modulated by harmonic signal is given

$$\mathbf{E}^I(\mathbf{r},\,t) \;=\; \mathbf{E}_0 \; \frac{4}{cT\sqrt{\pi}} \; \exp\left[-\left(\frac{4}{T}\left(t - t_0 - \frac{\mathbf{r}\cdot\mathbf{a}_k}{c}\right)\right)^2\right]\cos(2\pi f_0 t), \qquad (5.1A.7)$$

where $\mathbf{E}_0$ is the vector of the intensity of the electric field of the excitation wave at the origin of the coordinate system, $\mathbf{r}$ is the positional vector of the point where the intensity of the excitation wave is computed, $\mathbf{a}_k$ is the unit vector in the direction of the propagation, c is the speed of the light in the surrounding medium, T is the width of the Gaussian pulse, t_0 is the retardation of its peak, and f_0 is the center frequency of the analyzed band. The example of the Gaussian pulse modulated by the harmonic signal with its spectrum, defined by the following parameters: $|\mathbf{E}_0|=120\pi$ V/m, $T=8$ ns, $t_0=10$ ns and $f_0= 500$ MHz, is depicted in fig. **5.1A.2**.
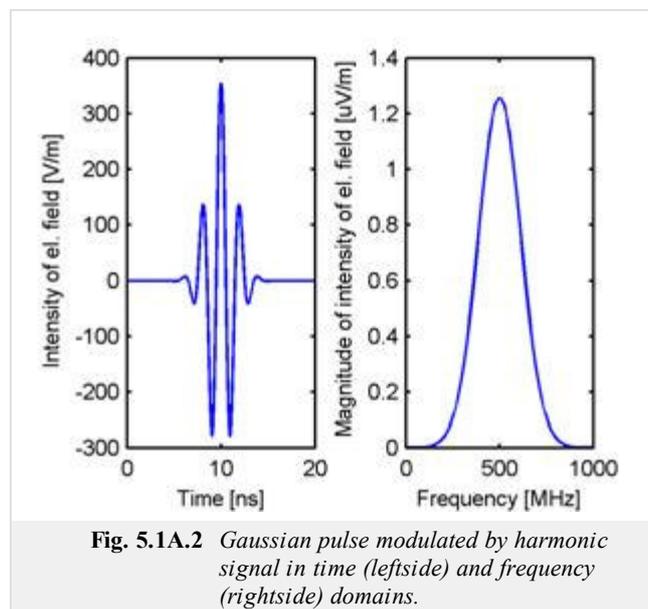

Fig. 5.1A.2 *Gaussian pulse modulated by harmonic signal in time (leftside) and frequency (rightside) domains.*

Implicit algorithm

The derivation of the implicit algorithm is demanding (the system of the equations (5.1A.1) to (5.1A.6) is discretized in space and time), and for the interested reader it is carried out in the layer B. In this layer the properties of the implicit scheme are only discussed.

At the implicit algorithm, opposite to the explicit one, the inverse matrix has to be solved and its implementation on the computer is more demanding than in case of the explicit algorithm. However, its advantage is that the length of the time step can be chosen arbitrarily (we should keep in mind, the length of the time step influences the accuracy), and this algorithm is more stable than the explicit one.

Using the implicit algorithm will be demonstrated on the analysis of the wire dipole of the length 2 m, which is excited at its center by the Gaussian pulse modulated by a harmonic signal (5.1A.7) defined by the following parameters: $E_0=120\pi$ V/m, $T=6$ ns, $t_0=8$ ns, $f_0 = 0$ Hz. Since the center frequency of the analyzed band is 0 Hz, let's further call this pulse as the Gaussian pulse. Let's chose the length of time analysis 400 ns. The body of the dipole is divided to 40 segments (the space discretization of the task, layer B). The analysis is carried out two times for different lengths of the time step. In case of the first analysis the length of the time step is $\Delta t = R_{min}/c = 0.166$ ns, $c\Delta t = 0,05$ m, and in the case of the second analysis the length of the time step is half of the first case, $\Delta t = R_{min}/(2c) = 0.083$ ns, $c\Delta t = 0,025$ m (the time discretization of the task).

The excitation Gaussian pulse is depicted in fig. **5.1A.3**. The computed current responses to this pulse at the center of the dipole are depicted in fig. **5.1A.4**. Comparing both figures it can be said that the computed responses are significantly larger than the excitation pulse and the difference between computed current is very small.

The wire dipole is a narrow band antenna which is able to "transform" the energy to the free space only within the limited range of the frequencies. Since the excitation pulse carries the energy even at the frequencies, where the antenna has this ability significantly smaller, the energy goes from the one arm of the dipole to the other and back, and the energy is gradually radiated.
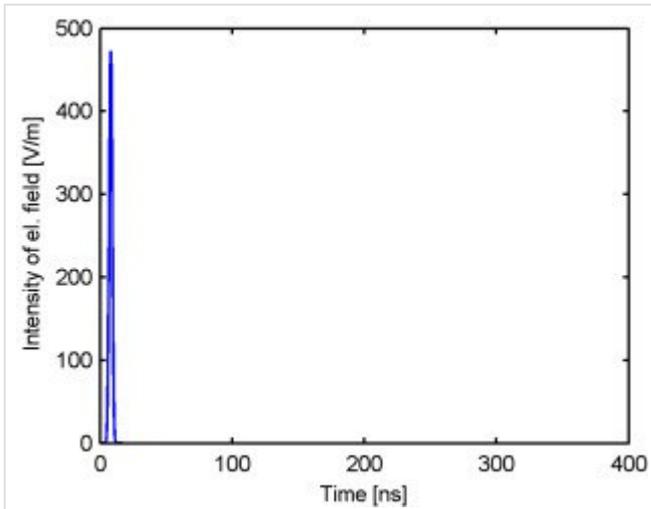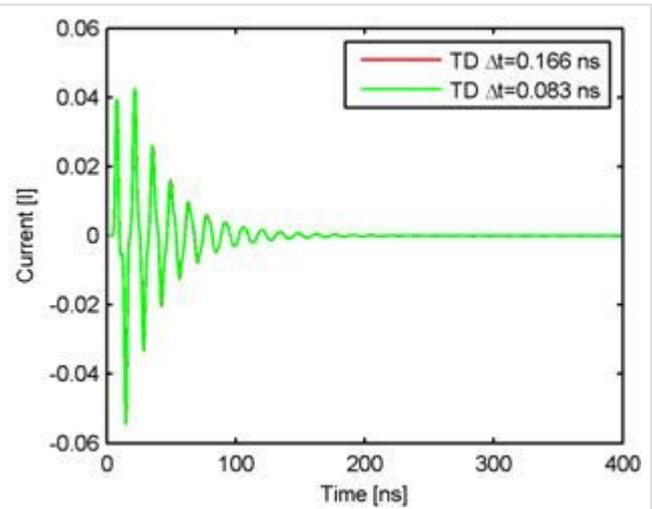
Fig. 5.1A.3 *Excitation Gaussian pulse.*

Fig. 5.1A.4 *Current responses on the Gaussian pulse (obr. 5.1A.3) in the center of the wire antenna for different lengths of the time step*

Since the computed responses are stable, the input impedance of the wire antenna can be computed (fig. **5.1A.5**, denoted by TD). In the same picture the input impedance computed by the method of moments in the frequency domain is depicted (denoted by FD). The solution in the frequency domain is taken as the reference solution because its accuracy is known. Comparing this solution with the results for the implicit algorithm for the length of the time step $\Delta t = R_{min}/c = 0.166$ ns, $c\Delta t = 0,05$ m (red line) we can say that larger differences are apparent at higher frequencies. These differences are mainly caused by the discretization of the system of the equations (5.1A.1) to (5.1A.6) in time. If we decrease the length of the time step to the half of the original value (green line), the differences are smaller (more accurate approximation in the time). The length of the time step influences the accuracy of the algorithm.
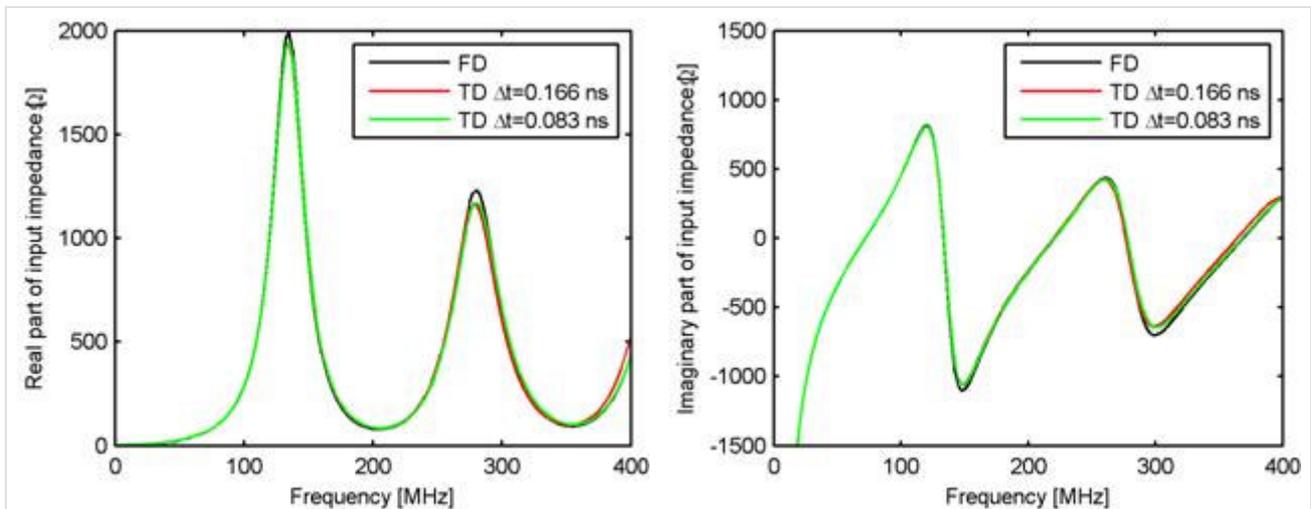

Fig. 5.1A.5 *Real (left) a imaginary (right) part of the input impedance of the analyzed wire dipole.*

The length of the time step influences the stability of the algorithm too. This fact is demonstrated in fig. **5.1A.6**, where the current response of the same dipole antenna for the double length of the time step ($\Delta t = 2R_{min}/c = 0.333$ ns, $c\Delta t = 0,1$ m.) is depicted. The computed response is not stable, but it grows to infinity.
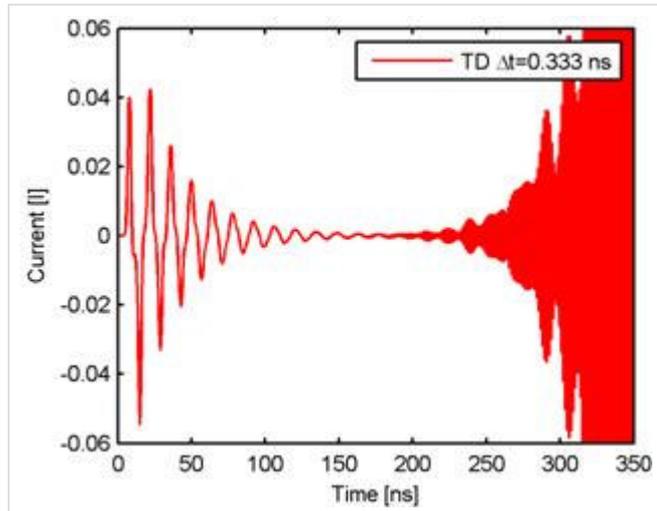
Fig. 5.1A.6 *Unstable current response to the Gaussian pulse (fig. 5.1A.3) in the center of the wire dipole (larger length of the time step).*

For stabilizing the time algorithm different techniques can be used, however, these ones are behind the scope of this chapter.

## 5.1 Time-domain modeling of wire antennas by method of moments

Advanced theory

In this layer the implicit algorithm will be derived by solving the system of the equations (5.1A.1) to (5.1A.6) for the wire dipole which is placed along the z axis (fig. **5.1B.1**). Since only z components of the vector potential and the intensity of electric field are in this system of equations, it is not necessary to distinguish among the components of vectors. Thus, in the following text, the subscript z will be omitted.

The implicit algorithm will be derived from the equation (5.1A.5) which can be rewritten by considering the boundary condition (5.1A.6) to the following form

$$\frac{\partial A(z,\,t)}{\partial t} + \frac{\partial \varphi(z,\,t)}{\partial z} = E^I(z,\,t)\,. \tag{5.1B.1}$$

Now, let's carry out the first step of the solution, the application of the method of moments to (5.1B.1). Let's divide the wire dipole to N segments of the same length Δz and denote the ends of the segments by coordinates $z_0, z_1, \ldots, z_{N+1}$ (fig. **5.1B.1**). Further, let's denote the center of segments by $z_0{}^+, z_1{}^+, \ldots, z_N{}^+$, and $z_1{}^-, z_2{}^-, \ldots, z_{N+1}{}^-$, respectively. For the expansion of the current distribution on the wire dipole in space, let's use the constant basis functions: in the region of a current cell (fig. **5.1B.1**), which is bounded by the coordinates either $z_n{}^+, z_{n+1}{}^+$ or $z_{n-1}{}^-, z_n{}^-$, the constant current distribution is supposed. It is appropriate to remember, that the current at the end of the wire is equal to zero, because it can not flow anywhere. In the region of a charge cell, which is bounded by the coordinates z_n, z_{n+1}, the constant charge density distribution is supposed.

Fig. 5.1B.1 *Wire dipole discretization.*
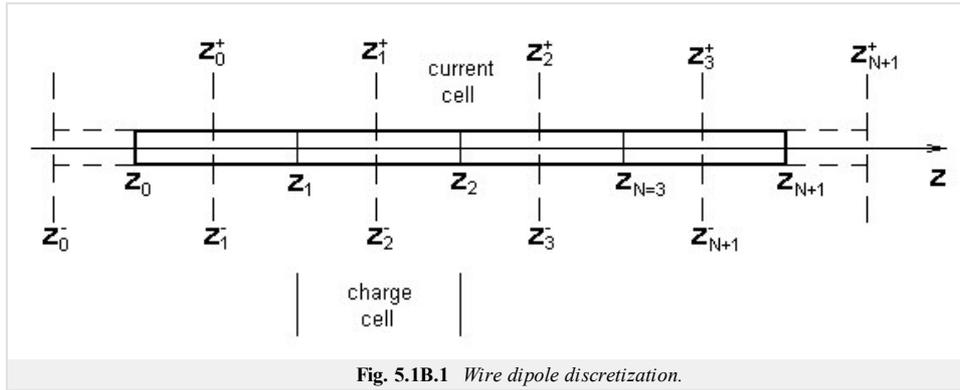
The constant basis functions are in the region of the current cell defined by the following relation

$$f_n = \begin{cases} 1 & z_n - \frac{\Delta z}{2} \le z \le z_n + \frac{\Delta z}{2} \\ 0 & otherwise \end{cases}. \tag{5.1B.2}$$

By these basis functions and time dependant unknown current coefficients $I_n(t)$ the time-space current distribution in (5.1A.1) can be approximated

$$I(z,\,t) \approx \sum_{n=1}^{N} I_n(t) f_n(z). \tag{5.1B.3}$$

Let's discretize our wire dipole in time (the second step). Let's divide the time step to the equal intervals of the length Δt and denote the individual time instant on time axes by $t_k = k\Delta t$ for $k = 0, 1, 2, \ldots, \infty$. At these time instants the current distribution on the wire antenna is be computed.

Substituting (5.1B.3) to (5.1A.1) and (considering (5.1B.2)) the vector potential at the point z_m and the time instant t_k can be written in the following form

$$A(z_m,\,t_k) = \frac{\mu}{4\pi} \int_{\xi=-h}^{h} \frac{I\big(\xi,\,t_n - R(z_m,\,\xi)/c\big)}{\sqrt{a^2 + |z_m - \xi|^2}}\,d\xi \approx \frac{\mu_0}{4\pi} \int_{\xi=-h}^{h} \frac{\sum_{n=1}^{N} I_n(t_n - R(z_m,\,\xi)/c) f_n(\xi)}{\sqrt{a^2 + |z_m - \xi|^2}}\,d\xi = \sum_{n=1}^{N} I_n\big(t_{Rk}(m,\,n)\big)\kappa(m,\,n), \tag{5.1B.4}$$

where

$$\kappa(m,\,n) = \frac{\mu}{4\pi} \int_{\xi=z_n - \frac{\Delta z}{2}}^{z_n - \frac{\Delta z}{2}} \frac{d\xi}{\sqrt{a^2 + |z_m - \xi|^2}} = \frac{\mu}{4\pi}\left\{ \ln\!\left[z_m - z_n + \frac{\Delta z}{2} + \sqrt{\left(z_m - z_n + \frac{\Delta z}{2}\right)^2 + a^2}\right] - \ln\!\left[z_m - z_n - \frac{\Delta z}{2} + \sqrt{\left(z_m - z_n - \frac{\Delta z}{2}\right)^2 + a^2}\right] \right\}, \tag{5.1B.5}$$

$$t_{Rk}(m,\,n) = t_k - R(m,\,n)/c, \tag{5.1B.6}$$

$$R(m,\,n) = \sqrt{a^2 + |z_m - z_n|^2}, \tag{5.1B.7a}$$

$$R(m,\,n) = |z_m - z_n|. \tag{5.1B.7b}$$

For the evaluation of the distance $R(m,\,n)$ in (5.1B.6) the accurate relation (5.1B.7a), or the approximate one (5.1B.7b), which neglects the radius of a wire a, can be used. Using the approximate relation increases the stability of the algorithm, however, it is possible to use it only if the length of the time step multiplied by the speed of light in the surrounding medium is much larger, than the radius of the wire ($c\Delta t \gg a$). Otherwise, the accuracy falls down. Using the approximate relation was introduced in **[37]** and if the mentioned condition is met, it is appropriate to use (5.1B.7b). The interested reader can test the influence of that neglect on the accuracy and stability of the algorithm.

Let's discretize the scalar potential (5.1A.2) where the space-time charge density distribution is the unknown quantity. The charge density can be computed with the continuity equation (5.1A.4) which can be rewritten in the following form

$$\sigma(z, t) = -\int \frac{\partial I(z, t)}{\partial z} dt.$$ (5.1B.8)

Substituting the space-time charge density distribution (5.1B.3) to (5.1B.8) and approxiomating the partial differentiation of the current with respect to variable z by the center difference, the continuity equation (5.1B.8), after the swap of the derivation and integration, can be written

$$\sigma(z, t) \approx -\frac{\partial \sum\limits_{n=1}^{N} \int I_n(t)dt f_n(z)}{\partial z} \approx -\sum\limits_{n=0}^{N} \frac{\int I_{n+1}(t)dt - \int I_n(t)dt}{\varDelta z} f_n^+(z).$$ (5.1B.9)

Since the constant basis functions are used for the space approximation of the current (5.1B.3), the partial derivative with respect to variable z in (5.1B.9) could not be computed straightforwardly (the derivative of the constant function is equal to zero), but the center difference has to be used. To be the differentiating correct, the integral of currents with respect to time at the ends of the wire were included in (5.1B.9), although they are equal to zero. In the last step in (5.1B.9) the basis functions were swaped: from the spacious approximation of the current by the basis functions $f_n(z)$ we got toward the approximation of the charge density by the basis functions $f_n^+(z)$.

Substituting the continuity equation (5.1B.9) to (5.1A.2) the scalar potential at the point z_m^+ and the instant t_k can be evaluated

$$\varphi(z_m^+, t_k) = -\frac{1}{4\pi\varepsilon} \int\limits_{\xi=-h}^{h} \frac{\sigma(\xi, t_k - R(z_m^+, \xi)/c)}{R(z_m^+, \xi)} d\xi \approx -\frac{1}{4\pi\varepsilon} \left. \int\limits_{\xi=-h}^{h} \right[\sum\limits_{n=0}^{N} \frac{\int\limits_{0}^{t_k - R(z_m^+, \xi)/c} I_{n+1}(t)dt - \int\limits_{0}^{t_k - R(z_m^+, \xi)/c} I_n(t)dt}{\varDelta z} f_n^+(\xi) \frac{1}{R(z_m^+, \xi)} d\xi = $$ (5.1B.10)

$$= \sum\limits_{n=1}^{N} \frac{\int\limits_{0}^{t_{Rk}(m^+, n^+)} I_n(t)dt\kappa(m^+, n^+)}{\varDelta z} - \sum\limits_{n=1}^{N} \frac{\int\limits_{0}^{t_{Rk}(m^+, n^-)} I_n(t)dt\kappa(m^+, n^-)}{\varDelta z} = \varphi(z_m^{++}, t_k) - \varphi(z_m^{+-}, t_k)$$

,

where

$$t_{Rk}(m^\pm, n^\pm) = t_k - R(m^\pm, n^\pm)/c,$$ (5.1B.11)

$$R(m^\pm, n^\pm) = \sqrt{a^2 + |z_m^\pm - z_n^\pm|^2},$$ (5.1B.12a)

$$R(m^\pm, n^\pm) = |z_m^\pm - z_n^\pm|,$$ (5.1B.12b)

$$\kappa(m^\pm, n^\pm) = \frac{1}{4\pi\varepsilon} \int\limits_{\xi=z_n^\pm - \frac{\varDelta z}{2}}^{z_n^\pm - \frac{\varDelta z}{2}} \frac{d\xi}{\sqrt{a^2 + |z_m^\pm - \xi|^2}} = \frac{1}{4\pi\varepsilon} \left\{ \ln\left[z_m^\pm - z_n^\pm + \frac{\varDelta z}{2} + \sqrt{\left(z_m^\pm - z_n^\pm + \frac{\varDelta z}{2}\right)^2 + a^2} \right] - \ln\left[z_m^\pm - z_n^\pm - \frac{\varDelta z}{2} + \sqrt{\left(z_m^\pm - z_n^\pm - \frac{\varDelta z}{2}\right)^2 + a^2} \right] \right\}.$$ (5.1B.13)

Here, as in case of the evaluating of the vector potential, it is possible to use for the evaluation of the distance $R(m^\pm, n^\pm)$ in (5.1B.11) either the accurate relation (5.1B.12a), or the approximate relation (5.1B.12b). The reasons and the condition are the same as in case of the evaluation of the vector potential.

Similarly the scalar potential at the point z_m and the time instant t_k can be evaluated z_m^-

$$\varphi(z_m^-, t_k) = \varphi(z_m^{-+}, t_k) - \varphi(z_m^{--}, t_k).$$ (5.1B.14)

To evaluate the relations (5.1B.10) and (5.1B.14), it is necessary to compute the integral of the current with respect to time. The integral can be evaluated according to different numerical integration rules. We choose the trapezoid rule because its implementation is easy and offers sufficient accuracy. By this rule, the integral of the current in the interval from 0 to t_k, in case of the equidistant division of the interval, can be computed

$$\int\limits_{0}^{t_k} I(t)dt \approx \varDelta t \left[\frac{I(t_0 = 0)}{2} + \sum\limits_{l=1}^{k-1} I(t_l) + \frac{I(t_k)}{2} \right].$$ (5.1B.15)

Now, let's go back to (5.1B.1) and discretize it. Let's approximate the first derivative of the vector potential with respect to time by the center difference of the first order. By this step the partial derivative of the vector potential is evaluated at the point z_m and the time instant $t_{k-1/2}$. To be the calculation sufficiently accurate, the partial derivative of the scalar potential with respect to variable z has to be evaluated at the same point and the time instant. This can be reached by using the center differentiation for the scalar potential at the points z_m^+ and z_m^- at two instants t_k and t_{k-1}. The average of these central differences is actually numerically evaluated the derivative of the scalar potential at the point z_m and the time instant $t_{k-1/2}$. Of course, the excitation pulse has to be evaluated at the same point and the instant as the vector and scalar potentials. After this steps the equation (5.1B.1) can be rewritten into the following form

$$\frac{A(z_m, t_k) - A(z_m, t_{k-1})}{\varDelta t} + \frac{1}{2\varDelta z}\left(\varphi(z_m^+, t_k) - \varphi(z_m^-, t_k) + \varphi(z_m^+, t_{k-1}) - \varphi(z_m^-, t_{k-1})\right) = E^I(z, t_{k-1/2}).$$ (5.1B.16)

Rearranging terms in (5.1B.16) we obtain

$$A(z_m, t_k) + \frac{\Delta t}{2\Delta z}\left(\varphi(z_m^+, t_k) - \varphi(z_m^-, t_k)\right) = \Delta t E^I(z, t_{k-1/2}) + A(z_m, t_{k-1})$$
$$- \frac{\Delta t}{2\Delta z}\left(\varphi(z_m^+, t_{k-1}) - \varphi(z_m^-, t_{k-1})\right) \qquad (5.1B.17)$$

The vector potential on the left-side of the equation (5.1B.17), computed according to (5.1B.4), can be transcribed

$$A(z_m, t_k) = A_1(z_m, t_k) + A_2(z_m, t_k), \qquad (5.1B.18)$$

where

$$A_1(z_m, t_k) = \sum_{n=1}^{N} I_n(t_{Rk}(m, n))\kappa(m, n) \quad for \ t_{Rk}(m, n) > t_{k-1}, \qquad (5.1B.19a)$$

$$A_2(z_m, t_k) = \sum_{n=1}^{N} I_n(t_{Rk}(m, n))\kappa(m, n) \quad for \ t_{Rk}(m, n) \le t_{k-1}. \qquad (5.1B.19b)$$

For evaluating $A_1(z_m, t_k)$, only the contributions of currents at delayed time instants $t_{Rk}(m, n) > t_{k-1}$ are considered (the unknowns currents). In case of the evaluating of $A_2(z_m, t_k)$, only the contributions of currents at delayed instants $t_{Rk}(m, n) \le t_{k-1}$ are considered (the known currents). The unknown current at the time instant $t_{Rk}(m, n)$ in the interval from t_{k-1} to t_k can be evaluated

$$I(t_{Rk}(m, n)) = \frac{R(z_m, z_n)}{c}I(t_{k-1}) + \left(1 - \frac{R(z_m, z_n)}{c}\right)I(t_k). \qquad (5.1B.20)$$

Substituting (5.1B.20) to (5.1B.19a), $A_1(z_m, t_k)$ can be expressed by the unknown and known currents at the instants t_k and t_{k-1}

$$A_1(z_m, t_k) = A_{11}(z_m, t_k) + A_{12}(z_m, t_k), \qquad (5.1B.21)$$

where

$$A_{11}(z_m, t_k) = \sum_{n=1}^{N}\left(1 - \frac{R(z_m, z_n)}{c\Delta t}\right)\kappa(m, n)I_n(t_k) = \sum_{n=1}^{N} A_{11C}(m, n)I_n(t_k), \qquad (5.1B.22a)$$

$$A_{12}(z_m, t_k) = \sum_{n=1}^{N}\frac{R(z_m, z_n)}{c\Delta t}\kappa(m, n)I_n(t_k) = \sum_{n=1}^{N} A_{12C}(m, n)I_n(t_k), \qquad (5.1B.22b)$$

The term $A_{11}(z_m, t_k)$ contains only the unknown currents at the instant t_k.

Now, let's focus our attention on the rest of the right-side of the equation (5.1B.17). The scalar potential (5.1B.10) at the point z_m^+ can be evaluated

$$\varphi(z_m^+, t_k) = \varphi(z_m^{++}, t_k) - \varphi(z_m^{+-}, t_k), \qquad (5.1B.23)$$

where the terms on the right-side (5.1B.23) can be expressed as the vector potential (5.1B.18). Let's focus our attention on the first term

$$\varphi(z_m^{++}, t_k) = \varphi_1(z_m^{++}, t_k) + \varphi_2(z_m^{++}, t_k), \qquad (5.1B.24)$$

where

$$\varphi_1(z_m^{++}, t_k) = \sum_{n=1}^{N}\int_{0}^{t_{Rk}(m^+, n^+)} I_n(t)dt\frac{\kappa(m^+, n^+)}{\Delta z} \quad for \ t_{Rk}(m^+, n^+) > t_{k-1}, \qquad (5.1B.25a)$$

$$\varphi_2(z_m^{++}, t_k) = \sum_{n=1}^{N}\int_{0}^{t_{Rk}(m^+, n^+)} I_n(t)dt\frac{\kappa(m^+, n^+)}{\Delta z} \quad for \ t_{Rk}(m^+, n^+) \le t_{k-1}. \qquad (5.1B.25b)$$

The situation is analogical as for the vector potential, however more complicated, because of the integral of the current with respect to time in (5.1B.25). Since the term $\varphi_1(z_m^{++}, t_k)$ contains the unknown current at the time instant t_k, let's express the integral of the current over the interval from 0 to $t_{Rk}(m+, n+)$ in (5.1B.25a) as the sum of two integrals

$$\varphi_1(z_m^{++}, t_k) = \varphi_1'(z_m^{++}, t_k) + \varphi_{13}(z_m^{++}, t_k) \qquad (5.1B.26)$$

where

$$\varphi_1'(z_m^{++}, t_k) = \sum_{n=1}^{N}\int_{t_{k-1}}^{t_{Rk}(m^+, n^+)} I_n(t)dt\frac{\kappa(m^+, n^+)}{\Delta z} \qquad (5.1B.27a)$$

$$\varphi_{13}(z_m^{++}, t_k) = \sum_{n=1}^{N}\frac{\kappa(m^+, n^+)}{\Delta z}\int_{0}^{t_{k-1}} I_n(t)dt = \sum_{n=1}^{N}\varphi_{13C}(m^+, n^+)\int_{0}^{t_{k-1}} I_n(t)dt \qquad (5.1B.27b)$$

The term $\varphi_{13}(z_m^{++}, t_k)$ in (5.1B.27b) can be easily evaluated, since it contains only the known values of the currents. However, the term $\varphi_1'(z_m^{++}, t_k)$ contains the unknown current at the time instant t_k. Therefore we express $\varphi_1'(z_m^{++}, t_k)$ by the trapezoid rule for the numerical evaluation of the integral in the interval from t_{k-1} to $t_{Rk}(m+, n+)$, and the relations (5.1B.18) and (5.1B.20)

$$\varphi_1'(z_m^{++}, t_k) = \sum_{n=1}^{N} \frac{\Delta t}{2} \left(\frac{t_{Rk}(m^+, n^+) - t_{k-1}}{\Delta t} \right) \left(I_n(t_{Rk}(m^+, n^+)) + I_n(t_{k-1}) \right) \frac{\kappa(m^+, n^+)}{\Delta z} =$$

$$= \sum_{n=1}^{N} \frac{\Delta t}{2} \left(1 - \frac{R(z_m^+, z_m^+)}{c\Delta t} \right) \left[I_n(t_{k-1}) \frac{R(z_m^+, z_m^+)}{c\Delta t} + \right.$$

$$\left. + \left(1 - \frac{R(z_m^+, z_m^+)}{c\Delta t} \right) I_n(t_k) + I_n(t_{k-1}) \right] \frac{\kappa(m^+, n^+)}{\Delta z} =$$

$$= \varphi_{11}(z_m^{++}, t_k) + \varphi_{12}(z_m^{++}, t_k)$$

(5.1B.28)

where

$$\varphi_{11}(z_m^{++}, t_k) = \sum_{n=1}^{N} \frac{\Delta t}{2\Delta z} \left(1 - \frac{R(z_m^+, z_m^+)}{c\Delta t} \right)^2 \kappa(m^+, n^+) I_n(t_k) = \sum_{n=1}^{N} \varphi_{11C}(m^+, n^+) I_n(t_k)$$

(5.1B.29a)

$$\varphi_{12}(z_m^{++}, t_k) = \sum_{n=1}^{N} \frac{\Delta t}{2\Delta z} \left(1 - \frac{R^2(z_m^+, z_m^+)}{c^2 \Delta t^2} \right) \kappa(m^+, n^+) I_n(t_{k-1}) = \sum_{n=1}^{N} \varphi_{12C}(m^+, n^+) I_n(t_{k-1})$$

(5.1B.29b)

Now only the term $\varphi_{11}(z_m^{++}, t_k)$ contains the unknown current at the time instant t_k. This one can be easily expressed by (5.1B.29a). The relation (5.1B.24) can be expressed, considering (5.1B.25) to (5.1B.29), as follow

$$\varphi(z_m^{++}, t_k) = \varphi_{11}(z_m^{++}, t_k) + \varphi_{12}(z_m^{++}, t_k) + \varphi_{13}(z_m^{++}, t_k) + \varphi_2(z_m^{++}, t_k).$$

(5.1B.30)

Analogically we can proceed with the evaluation of the term $\varphi(z_m^{+-}, t_k)$

$$\varphi(z_m^{+-}, t_k) = \varphi_{11}(z_m^{+-}, t_k) + \varphi_{12}(z_m^{+-}, t_k) + \varphi_{13}(z_m^{+-}, t_k) + \varphi_2(z_m^{+-}, t_k),$$

(5.1B.31)

or the scalar potential at the point z_m^- and the time instant t_k $\varphi(z_m^-, t_k)$.

$$\varphi(z_m^-, t_k) = \varphi(z_m^{-+}, t_k) - \varphi(z_m^{--}, t_k),$$

(5.1B.14)

where

$$\varphi(z_m^{-+}, t_k) = \varphi_{11}(z_m^{-+}, t_k) + \varphi_{12}(z_m^{-+}, t_k) + \varphi_{13}(z_m^{-+}, t_k) + \varphi_2(z_m^{-+}, t_k),$$

(5.1B.32a)

$$\varphi(z_m^{--}, t_k) = \varphi_{11}(z_m^{--}, t_k) + \varphi_{12}(z_m^{--}, t_k) + \varphi_{13}(z_m^{--}, t_k) + \varphi_2(z_m^{--}, t_k).$$

(5.1B.32b)

The terms on the right-sides of the relations (5.1B.31) and (5.1B.32) can be expressed similarly as the terms on the right-sides of relations (5.1B.24) to (5.1B.29) by changing the corresponding superscripts + and -, or vice versa.

Substituting (5.1B.10), (5.1B.14), (5.1B.18) to (5.1B.21) and (5.1B.24) to (5.1B.32) into the left-side of the relation (5.1B.17) and by rearranging the terms in this equation we obtain

$$A_{11}(z_m, t_k) + \frac{\Delta t}{2\Delta z} \varphi_{11}(z_m, t_k) = \Delta t E^I(z_m, t_{k-1/2}) - A_{12}(z_m, t_k) - A_2(z_m, t_k) + A(z_m, t_{k-1}) -$$

$$- \frac{\Delta t}{2\Delta z} \left(\varphi_{12}(z_m, t_k) + \varphi_{13}(z_m, t_k) + \varphi_2(z_m, t_k) + \varphi(z_m^+, t_{k-1}) - \varphi(z_m^-, t_{k-1}) \right)$$

(5.1B.33)

where

$$\varphi_{11}(z_m, t_k) = \varphi_{11}(z_m^{++}, t_k) - \varphi_{11}(z_m^{+-}, t_k) - \varphi_{11}(z_m^{-+}, t_k) + \varphi_{11}(z_m^{--}, t_k),$$

(5.1B.34a)

$$\varphi_{12}(z_m, t_k) = \varphi_{12}(z_m^{++}, t_k) - \varphi_{12}(z_m^{+-}, t_k) - \varphi_{12}(z_m^{-+}, t_k) + \varphi_{12}(z_m^{--}, t_k),$$

(5.1B.34b)

$$\varphi_{13}(z_m, t_k) = \varphi_{13}(z_m^{++}, t_k) - \varphi_{13}(z_m^{+-}, t_k) - \varphi_{13}(z_m^{-+}, t_k) + \varphi_{13}(z_m^{--}, t_k),$$

(5.1B.34c)

$$\varphi_2(z_m, t_k) = \varphi_2(z_m^{++}, t_k) - \varphi_2(z_m^{+-}, t_k) - \varphi_2(z_m^{-+}, t_k) + \varphi_2(z_m^{--}, t_k).$$

(5.1B.34d)

The system of N equations (5.1B.33) is possible, with considering (5.1B.18) to (5.1B.21) and (5.1B.23) to (5.1B.32), to rewrite into the following matrix equation

$$\left([A_{11C}(m, n)] + \frac{\Delta t}{2\Delta z} [\varphi_{11C}(m, n)] \right) \{ I(m, t_k) \} = \{ \Delta t E^I(m, t_{k-1/2}) \} - [A_{12C}(m, n)] \{ I(n, t_{k-1}) \} -$$

$$- \{ A_2(m, t_k) \} + \{ A(m, t_{k-1}) \} - \frac{\Delta t}{2\Delta z} \left([\varphi_{12C}(m, n)] \{ I(m, t_{k-1}) \} + \right.$$

$$\left. + [\varphi_{13C}(m, n)] \left\{ \int_0^{t_{k-1}} I(m, t) dt \right\} + \{ \varphi_2(m, t_k) \} + \{ \varphi(m^+, t_{k-1}) \} - \{ \varphi(m^-, t_{k-1}) \} \right)$$

(5.1B.35)

where

$$\varphi_{11C}(m, n) = \varphi_{11C}(m^+, n^+) - \varphi_{11C}(m^+, n^-) - \varphi_{11C}(m^-, n^+) + \varphi_{11C}(m^-, n^-),$$

(5.1B.36a)

$$\varphi_{12C}(m, n) = \varphi_{12C}(m^+, n^+) - \varphi_{12C}(m^+, n^-) - \varphi_{12C}(m^-, n^+) + \varphi_{12C}(m^-, n^-),$$

(5.1B.36b)

$$\varphi_{13C}(m, n) = \varphi_{13C}(m^+, n^+) - \varphi_{13C}(m^+, n^-) - \varphi_{13C}(m^-, n^+) + \varphi_{13C}(m^-, n^-),$$ (5.1B.36c)

$$\varphi_2(m, n) = \varphi_2(m^+, n^+) - \varphi_2(m^+, n^-) - \varphi_2(m^-, n^+) + \varphi_2(m^-, n^-).$$ (5.1B.36d)

In (5.1B.35) we denote the matrix of the size NxN and $Nx1$ by square and brace brackets, respectively. It is apparent that the left side of the equation (5.1B.35) contains only the unknown currents at the time instants $t=t_k$, however, the right-side contains the known currents at time instants $t\pm t_{k-1}$. The algorithm can start with the assumption $\{I(m, t_0)\}=\{0\}$ and computing $\{I(m, t_1)\}$. When the current are computed $\{I(m,t_1)\}$, then it is possible to compute $\{I(m,t_2)\}$ and so on. Further, it should be noted, that it is necessary to solve the inverse matrix. However, this one does not depend on time and it is spare. Thus, the inverse matrix is computed only once.

Using the implicit algorithm is demonstrated in the layer A on the analysis of the wire dipole.

Numerical model of antenna excitation

Let's go back to the excitation of our wire antenna, and discus the appropriate numerical model of the excitation. Let's suppose that the feeding ports of our antenna are located in the position of the current cells z_1, z_2, …, z_N of our discretized wire antenna (fig. **5.1B.1**); denote this place by z_f. The length of the feeding port is equal to the length of the discretization segment. An antenna can generally work in a receiving or transmitting mode.

If the antenna works in the receiving mode, the plane wave electromagnetic wave incidents on antenna's surface, and its transient dependence can be arbitrary (in our case it is defined by a Gaussian pulse modulated by a harmonic signal (5.1A.7)). Depending on the direction of the incident wave with the respect to antenna axis, the transient dependence of the incident wave at the current cells z_1, z_2, …, z_N is delayed. The incident wave induces in the wire antenna a current. The current response at the location of the feeding port z_f can be recorded.

If the antenna works in the transmitting mode, the situation is analogical to the one described in chapter 4.1. In the transmitting mode the voltage source is connected to the feeding port of the antenna at z_f. This source evokes the intensity of the electric field at the feeding port (in our case the time dependence is described by the Gaussian pulse modulated by a harmonic signal). Since the excitation source is connected only at the location of the feeding port z_f, at the other current cells the intensity of the excitation field is equal to zero.

5.1 Time-domain modeling of wire antennas by method of moments

Matlab program

Program in Matlab for the analysis of the wire dipole in the time domain is available here. The implicit algorithm is implemented in the m-file. The source is described in detail, and the important parts of the script are discussed in the layer D. At the beginning of the script it is possible to choose the parameters of the analyzed dipole and the excitation pulse. The results of the analysis are depicted in four independent windows (the excitation pulse, the current response in the center of the dipole, the real and the imaginary part of the input impedance computed from the excitation pulse and the computed current response).

5.1 Time-domain modeling of wire antennas by method of moments

Developing Matlab

The program for the analysis of the wire dipole by the implicit approach is from the point of the implementation demanding. Firstly it is necessary to compute the excitation Gaussian pulse modulated by a harmonic signal (5.1A.7):

```
EG=E0*4/(c*T*sqrt(pi)).*exp(-(4/T)^2*(t-t0-delt/2).*(t-t0-
delt/2)).*cos(2*pi*f0*(t-delt/2));
```

Secondly the values of the coefficients $\kappa(m, n)$ and $\kappa(m^{\pm}, n^{\pm})$ are evaluated according to the relations (5.1B.5) and (5.1B.13). Further, the distances between the discretization segments, according to (5.1B.7) and (5.1B.12), are computed. We have an optional choice for this calculation. We can use either the accurate relations (5.1B.7a) and (5.1B.12a), or the approximate relations with neglecting of the radius of the antenna wire (5.1B.7b) and (5.1B.12b), if the following condition is met: $c\Delta t \gg a$.

```
Rmn=zeros(Nz+1,1);
kappa=zeros(Nz+1,1); % kappa according to (5.1B.5)
for m=0:Nz % (5.1B.13)
    hlp1=log((m+0.5)*delz + sqrt(((m+0.5)*delz)^2+a^2));
    hlp2=log((m-0.5)*delz + sqrt(((m-0.5)*delz)^2+a^2));
    kappa(m+1)=(hlp1-hlp2);
    % distance of discretization elements (5.1B.7) and (5.1B.12), optional
    choice
    % if c*delt>>a is met, it is better to use (5.1B.7b) and (5.1B.12b)
    neglecting of radius of dipole (layer B)
    % R(m+1)=sqrt(a^2+(m*delz)^2); % (5.1B.7a) a (5.1B.12a)
    R(m+1)= m*delz; % (5.1B.7b) a (5.1B.12b)
end
kappaFI=1/4/pi/ep*kappa; % kappa for scalar potential (5.1B.13)
kappa=mi/4/pi*kappa; % kappa for vector potential (5.1B.5)
```

Further, we evaluate the values of the matrix elements in (5.1B.35), i.e. $[A_{11C}(m, n)]$, $[A_{12C}(m,n)]$ according to the relations (5.1B.35) and (5.1B.22) for the vector potential and $[\varphi_{11C}(m, n)]$, $[\varphi_{12C}(m,n)]$ and $[\varphi_{13C}(m, n)]$ according to the relations (5.1B.36), (5.1B.27b) and (5.1B.29) for the scalar potential. Here it is necessary to notice that the elements of these matrixes concern currents for time $t_{Rk} > t_{k-1}$. In case of the matrixes of the vector potential $[A_{11C}(m,n)]$, $[A_{12C}(m, n)]$, the source code can be written:

```
A11C=zeros(N,N); % coefficients of vector potential (5.1B.35) and (5.1B.22)
A12C=zeros(N,N);
for m=1:N,
    for n=1:N,
        mn=abs(m-n); % consider terms at time tk>=tRk>tk-1
        if(R(mn+1)<cdelt) % distance has to be smaller than c*delt
            A11C(m,n)=A11C(m,n)+(1-R(mn+1)/cdelt)*kappa(mn+1);
            A12C(m,n)=A12C(m,n)+R(mn+1)/cdelt*kappa(mn+1);
        end
    end
end
```

In case of the matrixes of the scalar potential $[\varphi_{11C}(m, n)]$, $[\varphi_{12C}(m,n)]$ and $[\varphi_{13C}(m, n)]$ only a part of the source code for the calculation of the contribution of the first term with indexes (m^{+}, n^{+}) in (5.1B.36) will be shown. The calculation of the other terms with the indexes indexi (m^{+}, n^{-}), (m^{-}, n^{+}) and $(m^{-}, n-)$ is analogical:

```
FI11C=zeros(N,N); % coefficients of scalar potential (5.1B.36)
FI12C=zeros(N,N); % for ++ according to (5.1B.27b) and (5.1B.29);
FI13C=zeros(N,N); % for other combinations analogically

for m=1:N,
    for n=1:N,
        mn=abs(m-n); % computing of distance indexes between elements
        mnPP=mn; % P is +; M is -
        mnPM=mn+1;
        mnMP=abs(mn-1);
        mnMM=mn;
        if(m<n)
            hlp=mnPM;
            mnPM=mnMP;
            mnMP=hlp;
        end
        % terms contribute if the following condition is met: tRk>tk-1
        if(R(mnPP+1)<cdelt)
            FI11C(m,n)=FI11C(m,n) -delt/delz/2*(1-R(mnPP+1)
```

```
                 /cdelt)^2*kappaFI(mnPP+1); %(5.1B.29a)
                 FI12C(m,n)=FI12C(m,n) -delt/delz/2*(1-(R(mnPP+1)
                 /cdelt)^2)*kappaFI(mnPP+1); %(5.1B.29b)
                 FI13C(m,n)=FI13C(m,n)-1/delz*kappaFI(mnPP+1); %(5.1B.27b)
         end
         .
         .
         .
      end
 end
```

Since the inverse matrix in (5.1B.35) does not depend on time, we evaluate it before the time loop:

```
invA11CFI11C=inv(A11C+delt/delz/2*FI11C); % inverse matrix (5.1B.35)
```

Finally we implement the time loop where we compute from the known time "samples" of the vector and scalar potential the future samples of these quantities. Note that the values of currents, or the values of the integrals of currents, which do not fit to discrete time instants, are between the time instants linearly approximated. The integral of currents, which is computed according to the trapezoid rule, is stored in the variable *IS*. The meaning of the other symbols should be apparent. Again in case of the calculation of the scalar potential, the source code is shown only for the first term with indexes (m^+, n^+) (see (5.1B.34) and (5.1B.36)), the calculation of the contribution of the other terms with indexes (m^+, n^-), (m^-, n^+) and (m^-, n^-) is analogical.

```
% TIME LOOP
for k=3:Nt,
     A2=zeros(N,1); % (5.1B.19b)
     FI2=zeros(N,1); % (5.1B.36d), (5.1B.25b) and further analogically for
     +-,-+,--
     for m=1:N,
          for n=1:N,
               % computing of distance indexes between elements
               mn=abs(m-n); % vector potential
               mnPP=mn; % scalar potential
               mnPM=mn+1; % P is +; M is -
               mnMP=abs(mn-1);
               mnMM=mn;
               if(m<n)
                    hlp=mnPM;
                    mnPM=mnMP;
                    mnMP=hlp;
               end
               % further the terms are computed if: tRk<=tk-1
               if(R(mn+1)>=cdelt) % vector potetntial (5.1B.19b)
                    tk=k-R(mn+1)/cdelt;
                    if(tk>1)
                         ftk=floor(tk); % linear approximation of current
                         In=(I(n,ftk+1)-I(n,ftk))*(tk-ftk)+I(n,ftk);
                         A2(m,1)=A2(m,1)+In*kappa(mn+1);
                    end
               end
               if(R(mnPP+1)>=cdelt) % scalar potential ++; (5.1B.36d),
               (5.1B.25b)
                    tk=k-R(mnPP+1)/cdelt;
                    if(tk>1)
                         ftk=floor(tk); % linear approximation of integral of
                         current
                         ISn=(IS(n,ftk+1)-IS(n,ftk))*(tk-ftk)+IS(n,ftk);
                         FI2(m,1)=FI2(m,1)+ISn/delz*kappaFI(mnPP+1);
                    end
               end
               .
               .
               .
          end
     end
     E(nap,1)=EG(k); % excitation vector is non-zero at the location of
     feeding (5.1B.7)
     % (5.1B.35)
     I(:,k)=invA11CFI11C*(delt*E-A12C*I(:,k-1)-A2+A -delt/2/delz*
     (FI12C*I(:,k-1)+FI13C*IS(:,k-1)+FI2+FI)); % integration of current with
     respect to time (trapezoid rule) (5.1B.15)
     IS(:,k)=(I(:,k)+I(:,k-1))*delt/2+IS(:,k-1); % preparation for the next
     step
     A=A11C*I(:,k)+A12C*I(:,k-1)+A2; % (5.1B.18) a (5.1B.21)
```

```
    FI=FI11C*I(:,k)+FI12C*I(:,k-1)+FI13C*IS(:,k-1)+FI2; %
    (5.1B.30)-(5.1B.33)
    % considering (5.1B.26)-(5.1B.29)
end
```

5.1 Time-domain modeling of wire antennas by method of moments

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

Time-domain modeling is praticularly suitable for...

Possible answers for question 1:

- ○ … a narrowband analysis.
- ○ … a wideband analysis.
- ○ … an analysis which does not depend on the bandwidth.
- ◉ I won't answer this question

Question 2

By the method of moments and MOT (marching on-in-time) method the starting equation is "transformed" to a system of...

Possible answers for question 2:

- ○ … an algebraic equations.
- ○ … a delayed differential equations.
- ○ … a difference equations.
- ◉ I won't answer this question

Question 3

The principle of MOT (marching on-in-time) method is...

Possible answers for question 3:

- ○ … the unknonw (new) values of the currents at the current time are computed from the known (old) values of the currents in the previous time instants and this repeats until the whole response is computed.
- ○ … the current response is computed by solving the large matrix equation in an instant.
- ○ … the current response is approximated by a set of polynomials and unknown coefficients.
- ◉ I won't answer this question

Question 4

The length of the time step at the explicit approach...

Possible answers for question 4:

- ○ … has to be smaller or eaqual to time, which the electromagnetic wave needs for the traveling the smallest distance between the centers of the discretization elements.
- ○ … has to be larger than the time, which the electromagnetic wave needs for the traveling the smallest distance between the centers of the discretization elements.
- ○ … can be chosen arbitrarily.
- ◉ I won't answer this question

Question 5

The implicit aproach is...

Possible answers for question 5:

- … stable than the explicit one.
- … less time consuming than the explicit one, and the inverse matrix do not have to be computed.
- … more accurate than the explicit one.
- I won't answer this question

see result

Chapter 6: Frequency selective surfaces

Introduction

At the end of 18th century, American physicist D. Rittenhouse found out that some colors of a light spectrum are suppressed when a street lamp is observed through a silk handkerchief. This was the first proof of the fact that non-continuous surfaces can exhibit different transmission properties for different frequencies of incident wave. Hence, the surfaces are called frequency-selective surfaces.

Later, frequency selective surfaces appeared even in the range of radio frequencies. In 1919, G. Marconi and B. Franklin used a reflector built form horizontal conductors, which formed a non-continuous parabolic surface together. That way, a continuous reflective surface was imitated for a given frequency.
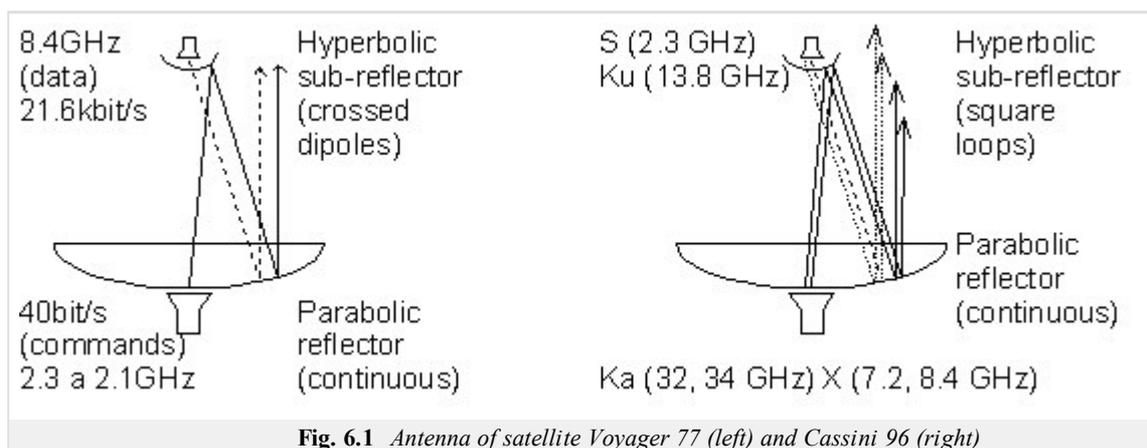
Fig. 6.1 *Antenna of satellite Voyager 77 (left) and Cassini 96 (right)*

Of course, frequency selective surfaces were further investigated and exploited. During the last 40 years especially, an intensive attention was turned to them. The satellite Voyager 77 exploited a frequency selective surface for implementing a double-frequency reflector (the parabolic antenna on the satellite could operate in two frequency bands and could be fed from two spatially separated sources). In the case of the satellite Cassini in 1996, the described principle was extended to four frequency bands. A schematic picture of both the antennas is depicted in fig. **6.1**, their photograph is shown in fig. **6.2**.

Selective surfaces are exploited even in military applications. As a typical example, a radome in the spike of a fighter can be given. A radome has to ensure a total reflection of all frequencies except of the operation frequency band of the radar. Next to the filtering, selective surfaces are also used as polarization filters.

In par. 6.1, numeric modeling of two selected types of frequency selective surfaces is described. Attention is turned to the surface, which consists of equidistantly distributed identical conductive rectangles, and to the surface, which is formed by rectangular slots in a metallic plane.

Fig. 6.2 *Antenna for the satellite Cassini 96 (left). A detailed view on hyperbolic sub-reflector of this antenna (right).*

6.1 Analysis of frequency selective surfaces

Basic theory

In this paragraph, reflection coefficient and transmission coefficient are computed for an infinite periodic frequency selective surface. The attention is turned to the surface consisting of rectangular elements (metallic, slot). First, the surface is assumed to be fabricated from such dielectrics, which properties are identical with surrounding. At the end, a brief note on the analysis of real substrates is provided.

Numeric analysis of frequency selective surfaces

There are two approaches to the analysis of frequency selective surfaces. The first one is based on the method of induced electromotoric forces and enables to analyze surfaces of both the finite extent and the infinite one (in most cases, the method is applied to infinite structures). Analysis by the method of induced electromotoric forces is not described here; readers interested can find more information in [17].

Spectral domain moment method is the second way of analyzing frequency selective surfaces [18]. We concentrate to this method only.
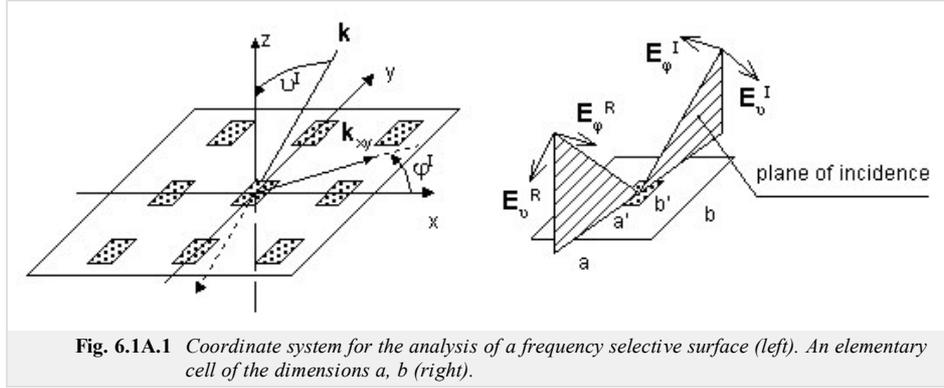
Fig. 6.1A.1 *Coordinate system for the analysis of a frequency selective surface (left). An elementary cell of the dimensions a, b (right).*

Metallic elements

We assume a periodic surface as depicted in fig. **6.1A.1**. The surface is illuminated by a plane wave. Projection of wave vector **k** to the plane of the surface is denoted by $\mathbf{k}_{xy}$. Electric filed intensity of incident wave can be expressed then as follows:

$$\mathbf{E}^I = \mathbf{E}_0 \exp\left[+j(\alpha_0 x + \beta_0 y)\right].$$

(6.1A.1)

Here, α_0, β_0 are negative projections of wave vector of incident wave to the directions of coordinate axes x and y (the exponential term exp(+jkr) is assumed for increasing phase in the direction of propagation). Projections of wave vector in the described coordinate system are given as follows:

$$\alpha_0 = k \sin(\vartheta)\cos(\varphi),$$
$$\beta_0 = k \sin(\vartheta)\sin(\varphi).$$

The symbol $\mathbf{E}_0$ denotes electric-field intensity vector in the origin of the coordinate system and k is free-space wave number.

In general, the incident wave is of both the parallel polarization and the perpendicular one. If field intensity of the incident wave is expressed in spherical coordinate system (fig. **6.1A.1**), then components E_ϑ^I and E_φ^I represent directly the parallel component and the perpendicular one.

If frequency selective surface is going to be analyzed, components of electric field intensity in the plane xy have to be known. We can evaluate them re-computing E_ϑ^I and E_φ^I as follows:

$$\begin{bmatrix} E_x^I \\ E_y^I \end{bmatrix} = \begin{bmatrix} \sin(\vartheta)\cos(\varphi) & -\sin(\varphi) \\ \sin(\vartheta)\sin(\varphi) & \cos(\varphi) \end{bmatrix} \begin{bmatrix} E_\vartheta^I \\ E_\varphi^I \end{bmatrix}.$$

(6.1A.2)

Due to the periodicity of the surface, all the necessary quantities (electric field intensity, current density) in the plane xy can be expressed using Fourier series. Using Fourier expansion, density of electric current **J** [A/m] on metallic element is expressed:

$$\mathbf{J} = \sum_{m=-M}^{M} \sum_{n=-N}^{N} \mathbf{J}(\alpha_m, \beta_n)\exp\left[+j(\alpha_m x + \beta_n y)\right],$$

(6.1A.3)

where

$$\alpha_n = \alpha_0 + \frac{2\pi}{a}m, \quad \beta_n = \beta_0 + \frac{2\pi}{b}n$$

(6.1A.4)

are spatial frequencies. Those spatial frequencies are a spatial analogy to the temporal angular frequency ω, which are used for temporal analysis of signals. The only difference is hidden in the fact that time period T corresponds to spatial periods a and b.

Coefficients of Fourier series can be obtained by integrating current densities over the whole surface of a cell (inverse Fourier transform)

$$\mathbf{J}(\alpha_m, \beta_n) = \frac{1}{ab} \iint\limits_{plocha\,bu\,\check{n}\,ky} \mathbf{J}(x, y)\exp\left[-j(\alpha_m x + \beta_n y)\right].$$

(6.1A.5)

In analogy to current density, Fourier series can express arbitrary periodic quantity. We do that in the next part, where a plane wave excitation of a selective surface is expressed in the spectral domain (i.e., in the domain of spatial frequencies). Determination of reflective properties of the selective surface is the aim of the computation.

Problem formulation

Assume a plane wave falling to a frequency selective surface under angles (ϑ, φ). This wave excites currents in metallic elements, which form a scattered field $\mathbf{E}^S$.

Whereas the magnitude (not the phase) of the intensity of incident wave $\mathbf{E}^I$ is constant over the whole elementary cell, the magnitude of the intensity of the scattered wave $\mathbf{E}^S$ is different in the different points of the cell. Relation between $\mathbf{E}^S$ and $\mathbf{E}^I$ changes according to the surface impedance in a given point. For the surface of conductive elements, we get

$$\mathbf{E}^I + \mathbf{E}^S = \frac{\mathbf{J}}{\gamma}. \qquad (6.1A.6)$$

Here, γ [S.m^{-1}] is electric conductivity of elements. Assuming the electric conductivity of metallic elements to be perfect, the right-hand side of (6.1A.6) approaches zero, and we get

$$\mathbf{E}^I + \mathbf{E}^S = 0. \qquad (6.1A.7)$$

Intensity of scattered electric field $\mathbf{E}^S$ in a point determined by the position vector $\mathbf{r}$ can be evaluated from vector potential $\mathbf{A}$

$$\mathbf{E}^S(\mathbf{r}) = -j\omega\mu\left\{\mathbf{A} + \frac{1}{k^2}\nabla\left[\nabla \cdot A(\mathbf{r})\right]\right\}. \qquad (6.1A.8)$$

As a source of vector potential $\mathbf{A}$, currents $\mathbf{J}$ induced by incident wave to metallic elements of selective surface can be considered. Contribution of current in $\mathbf{r}'$ to the potential in $\mathbf{r}$ is described by Green function $\mathbf{G}$. For vector potential, we get

$$\mathbf{A}(\mathbf{r}) = \iint\limits_{zdroje} \mathbf{G}(\mathbf{r}|\mathbf{r}')\mathbf{J}(\mathbf{r}')dS'. \qquad (6.1A.9)$$

In free space, Green function $\mathbf{G}$ is of the form

$$\mathbf{G} = \frac{\exp(-jk|r-r'|)}{4\pi|r-r'|}. \qquad (6.1A.10)$$

Combining relations (6.1A.9), (6.1A.8) and (6.1A.7), integral equation for space-domain electric field is obtained:

$$-j\omega\mu\left\{\iint\limits_{zdroje}\mathbf{G}(\mathbf{r}|\mathbf{r}') \cdot \mathbf{J}(\mathbf{r}')dS' + \frac{1}{k^2}\nabla\left[\nabla \cdot \iint\limits_{zdroje}\mathbf{G}(\mathbf{r}|\mathbf{r}') \cdot \mathbf{J}(\mathbf{r}')dS'\right]\right\} = -\mathbf{E}^I. \qquad (6.1A.11)$$

Eqn. (6.1A.11) is solved in the domain of spatial frequencies. Eqn. (6.1A.11) is mapped to the spectral domain within two steps. First, eqn. (6.1A.8) is mapped by its rewriting to the component form and by expressing vector potential by Fourier series. That way, we get [19]

$$\begin{bmatrix} E_x^S(\alpha_m, \beta_n) \\ E_z^S(\alpha_m, \beta_n) \end{bmatrix} = -j\omega\mu \begin{bmatrix} 1 - \dfrac{\alpha_m^2}{k^2} & \dfrac{-\alpha_m\beta_n}{k^2} \\ \dfrac{-\alpha_m\beta_n}{k^2} & 1 - \dfrac{\beta_n^2}{k^2} \end{bmatrix} \begin{bmatrix} A_x(\alpha_m, \beta_n) \\ A_y(\alpha_m, \beta_n) \end{bmatrix}. \qquad (6.1A.12)$$

In the second step, the relation is mapped to the domain of spatial frequencies (6.1A.9) [19]

$$\mathbf{A}(\alpha_m, \beta_n) = \mathbf{G}(\alpha_m, \beta_n) \cdot \mathbf{J}(\alpha_m, \beta_n). \qquad (6.1A.13)$$

Evaluating Fourier transform of the function (6.1A.10), we get [19] Green function in spectral domain

$$G(\alpha_m, \beta_n) = \frac{-j}{2\sqrt{k^2 - \alpha_m^2 - \beta_n^2}}, \qquad (6.1A.14)$$

where the square-root of negative imaginary part is assumed only.

Combining (6.1A.13), (6.1A.14) and (6.1A.12), the final relation for electric field in spectral domain is obtained

$$-\frac{1}{2\omega\varepsilon}\sum_{m,n} \begin{bmatrix} \dfrac{k^2 - \alpha_m^2}{\sqrt{k^2 - \alpha_m^2 - \beta_n^2}} & \dfrac{-\alpha_m\beta_n}{\sqrt{k^2 - \alpha_m^2 - \beta_n^2}} \\ \dfrac{-\alpha_m\beta_n}{\sqrt{k^2 - \alpha_m^2 - \beta_n^2}} & \dfrac{k^2 - \beta_n^2}{\sqrt{k^2 - \alpha_m^2 - \beta_n^2}} \end{bmatrix} \begin{bmatrix} J_x(\alpha_m, \beta_n) \\ J_y(\alpha_m, \beta_n) \end{bmatrix} \exp[j(\alpha_m x + \beta_n y)] = -\begin{bmatrix} E_x^I(x, y) \\ E_y^I(x, y) \end{bmatrix}. \qquad (6.1A.15)$$

Problem solution

Solving the problem, an unknown current distribution in (6.1A.15) is approximated by the linear combination of properly chosen basis functions and unknown approximation coefficients. Such formal approximation is substituted into the solved eqn. (6.1A.15). Since the approximation of the solution does not meet the initial relation exactly, we respect this fact by introducing a residual function (residuum). Lower the residuum is, more accurate the approximation is. The residuum is minimized by Galerkin method (residuum is sequentially multiplied by as many basis functions as many unknown approximation coefficients is computed; that way, N linear algebraic equations for N unknown approximation coefficients is obtained).

There are two approaches for choosing basis functions. The first one exploits basis functions, which are non-zero over the whole analyzed area. Functions are elected to physically represent standing waves of a current on an element.

The second approach divides the analyzed region to sub-regions, where current is approximated in terms of basis functions, which functional value is non-zero over a
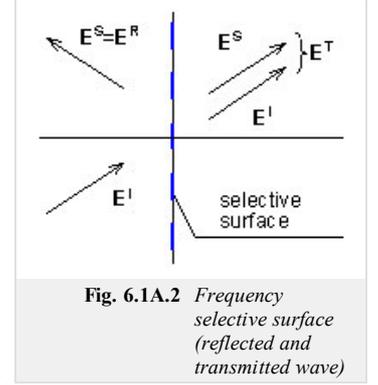
given sub-region only. This approach is advantageous in simple analysis of selective surfaces consisting of arbitrarily shaped elements.

Here, we concentrate on the first group of basis functions only. As basis functions, harmonic functions are used (first approach, the layer B), and a combination of harmonic functions and *Chebyshev polynomials* is exploited (second approach, the layer B).

A more detailed description of the above-given basis functions for solving (6.1A.15) is introduced in the layer B. Here, obtained results are presented only.

Harmonic basis functions

Assume a frequency selective surface consisting of rectangular perfectly conductive elements of the dimensions a' and b'. If current distribution over an element is approximated in terms of harmonic basis functions, we obtain for the mode (1, 1) for the parallel polarization such current distribution, which is depicted in fig. **6.1A.3a**, and for the perpendicular polarization, the distribution from fig. **6.1A.3b** is obtained

a)   b)
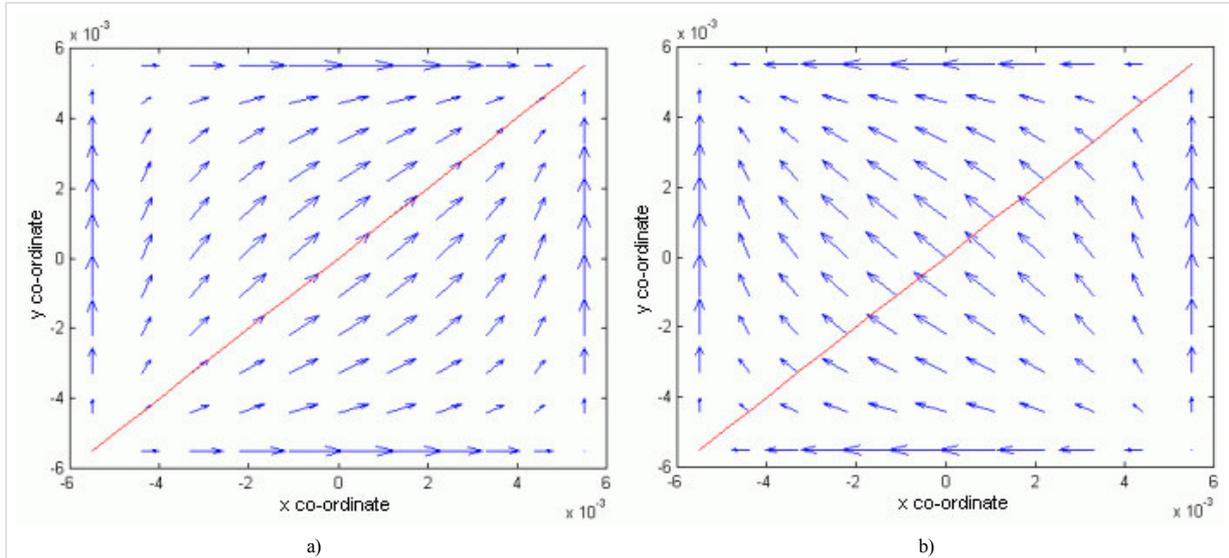
Fig. 6.1A.3 *Directional field of current density of the mode (1,1)*
a) for parallel polarization
b) for perpendicular one.
The cell is of dimensions a = b = 12mm, the metallic element is of dimensions a' = b' = 12mm; f=10GHz, ϑ = 1°, φ = 45°, d = 1.5mm, ε$_r$ = 3.7. Magnitude of marginal currents was increased with respect to reality. Plane of incidence is depicted in red.

Combination of harmonic functions and Chebyshev polynomials

In contrast to purely harmonic basis functions, the function cosine is replaced by *Chebyshev polynomial*, which is of infinite magnitude at the edges of the element. That way, strong currents flowing along edges are correctly modeled.

An equivalent circuit can describe behavior of frequency selective surfaces, which consist of metallic elements, where inductances of elements together with capacitances among ends of elements create a serial resonant circuit (fig. **6.1A.4**).

Reflection coefficient *r* (see fig. **6.1A.4**) is evaluated according to

$$P^T = P^I\left[1 - |\rho|^2\right]. \qquad (6.1A.16)$$

Here, P^I is power of incident wave and P^T denotes power of wave, which was transmitted by the surface to the right-hand half-space. Resistance 377 Ω models free space in front of the surface (left) and behind the surface (right).
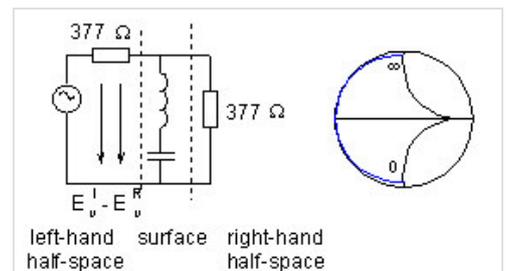
Fig. 6.1A.4 *Equivalent circuit of selective surface in frequency domain; surface consist of metallic elements (left); ϑ = φ = 0. Directivity factor in Smith chart (right).*
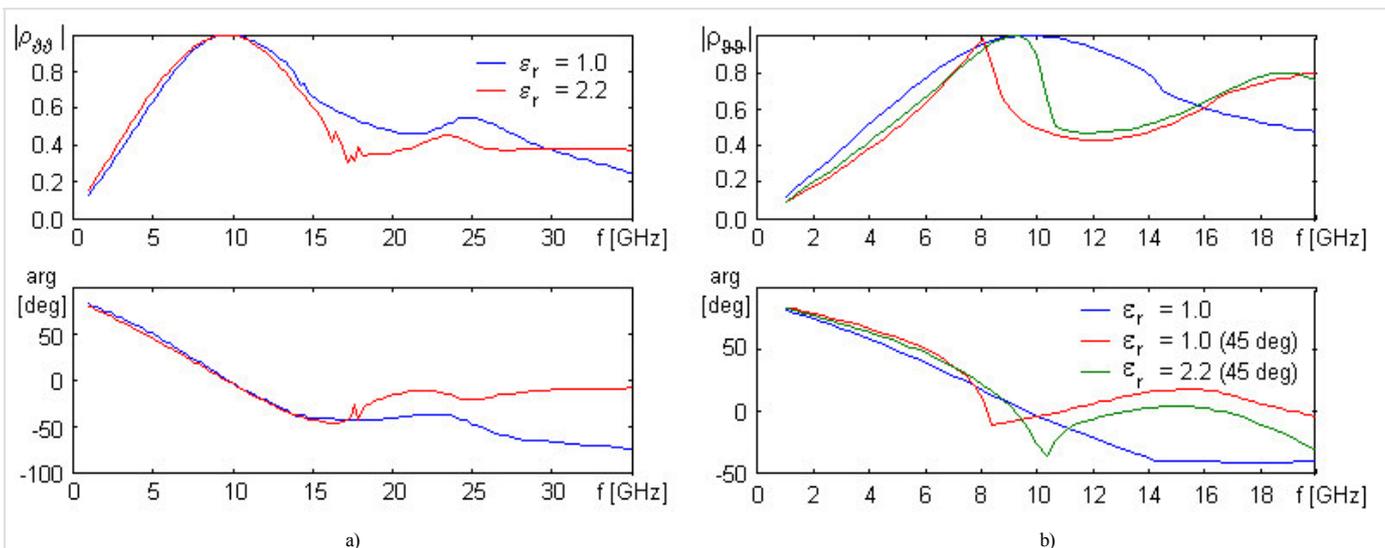
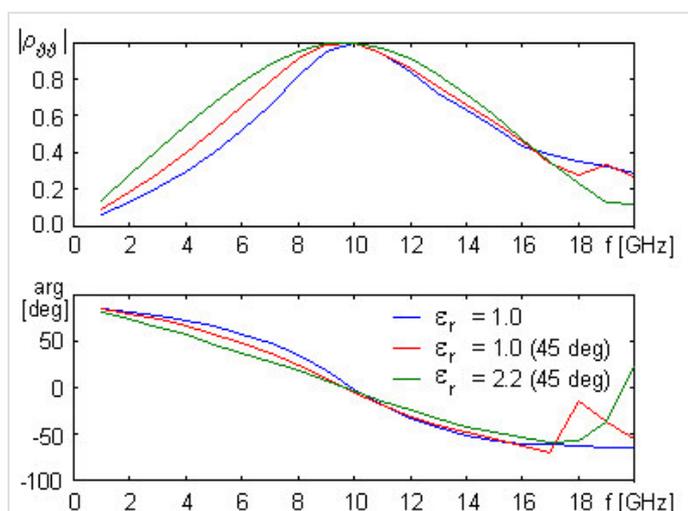**Fig. 6.1A.5** *Surface of rectangular elements, tuned for 10 GHz, parallel polarization*
a) normal incidence θ = 0°, φ = 90°
b) unnormal incidence θ = 45°, φ = 90°

Frequency course of reflection coefficient of the surface for parallel polarization is depicted in fig. **6.1A.5**. Considering the surface without dielectrics, (b = 21 mm, a = 7.5 mm, b' = 19.75 mm, a' = 1.5 mm), there is a break of the steepness of the reflection coefficient course around 14.3 GHz; this phenomenon is caused by the excitation of the first parasitic mode (*grating lobe*).

Adding dielectrics and modifying dimensions (b = 17 mm, a = 7.5 mm, b' = 15.75 mm, a' = 1.5 mm, height of the substrate d = 1.57 mm) causes a small decrease of the selectivity on one hand, but the stability of tuning for un-normal incidence is increased on the other hand (see fig. **6.1A.5b**). Further increase of the stability of reflection coefficient with respect to the angle of incidence can be reached by adding an upper dielectric layer. Thanks to the dielectrics, cell dimensions are smaller and parasitic modes are excited at higher frequencies (17.6 GHz). In fig. **6.1A.5b**, no significant effect of this mode can be observed.

Both the normal incidence and the un-normal one excite parasitic resonance at higher frequencies. At frequencies, which are higher than the frequency of the first parasitic mode, the course of reflection coefficient does not express the total intensity of the reflected wave, but only the intensity of the basic mode (0,0).

Considering perpendicular polarization of incident wave (in contrast to parallel one), reflection coefficient is nearly independent on the angle of incidence (see fig. **6.1A.6**). If a surface without dielectrics is considered, dimensions b = 21 mm, a = 7.5 mm, b' = 19.75 mm, a' = 1.5 mm, are assumed., If dielectrics is considered, dimensions b = 17 mm, a = 7.5 mm, b' = 15.75 mm, a' = 1.5 mm and the height of the substrate d = 1.57 mm are assumed.

Fig. 6.1A.6 *Surface of rectangular elements, tuned for 10 GHz, unnormal incidence, θ = 45°, φ = 0°, perpendicular polarization.*

Slot elements

In contrast to the surfaces of metallic elements, frequency selective surfaces of slots exhibit an *opposite* dependency of reflection coefficient, i.e. they behave as a band-stop filter. If no dielectrics and a single metallic layer are assumed, then slot surfaces are true complements of metallic-element surfaces.

Fig. 6.1A.7 *Complementary behavior of slot surfaces and metallic-element ones. Reflection coefficient of metallic-element surface (1) equals to transmission coefficient of slot surface (2).*
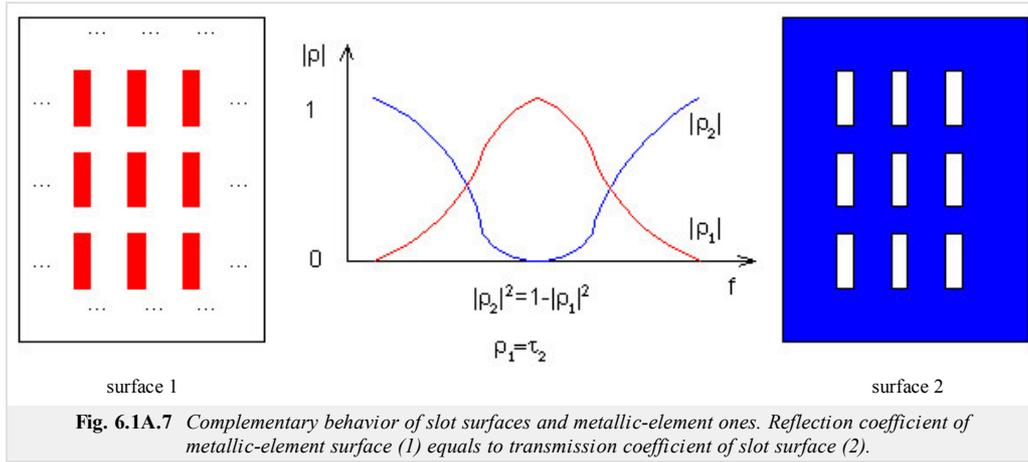
Now, we turn our attention to the derivation of equations for magnetic intensity over the aperture of slot surface in spectral domain. Considering the case without dielectrics and with a single metallic layer, duality principle can be applied: in (6.1A.15), we exchange **E** with **H**, ε with μ, **J** with $\mathbf{J_M}$

$$-\frac{1}{2\omega\mu}\sum_{m,n}\begin{bmatrix}\dfrac{k^2-\alpha_m{}^2}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} & \dfrac{-\alpha_m\beta_n}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}}\\[3mm]\dfrac{-\alpha_m\beta_n}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} & \dfrac{k^2-\beta_n{}^2}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}}\end{bmatrix}\begin{bmatrix}2J_x^{M+}(\alpha_m,\beta_n)\\[2mm]2J_y^{M+}(\alpha_m,\beta_n)\end{bmatrix}\exp\left[j(\alpha_m x+\beta_n y)\right]=-\begin{bmatrix}2H_x^I(x,y)\\[2mm]2H_y^I(x,y)\end{bmatrix}. \qquad (6.1A.17)$$

Comparing (6.1A.17) and (6.1A.15), a certain inconsistency can be revealed: coefficient 2 appeared at magnetic current density and exciting magnetic intensity. Next to the duality principle, the following fact had to be considered:

Field radiated by electric currents of the metallic-element surface is not scattered by those elements, whereas magnetic currents in slots produce a field, which is scattered by the conductive plane **[19]**. In order to include the influence of the surrounding of the conductive plane, following steps have to be done:

1. Original electric currents are replaced by magnetic ones (flowing on the surface of the slot). Magnetic current density can be computed applying $\mathbf{J}_M{}^+ = \mathbf{z}\times\mathbf{E}_{ap}$, where $\mathbf{E}_{ap}$ is electric field intensity in the aperture (similarly to electric current, even the magnetic conductor of a real thickness is flown by the current from one side only - from left one).

2. The slot is assumed to be filled in by perfect electric conductor (that way, an infinite continuous ground plane is obtained). This assumption does not change the situation any way because at the boundary of metal and air, the tangential components of electric field intensity are related by $\mathbf{E}_t^{kov}-\mathbf{E}_t^{vzduch}=\mathbf{J}_M{}^+\times(-z)$, and field intensity in metal $\mathbf{E}_t^{kov}=0$ is zero.

3. Mirror principle is applied, which converts the problem to free-space one (fig. **6.1A.8** right).
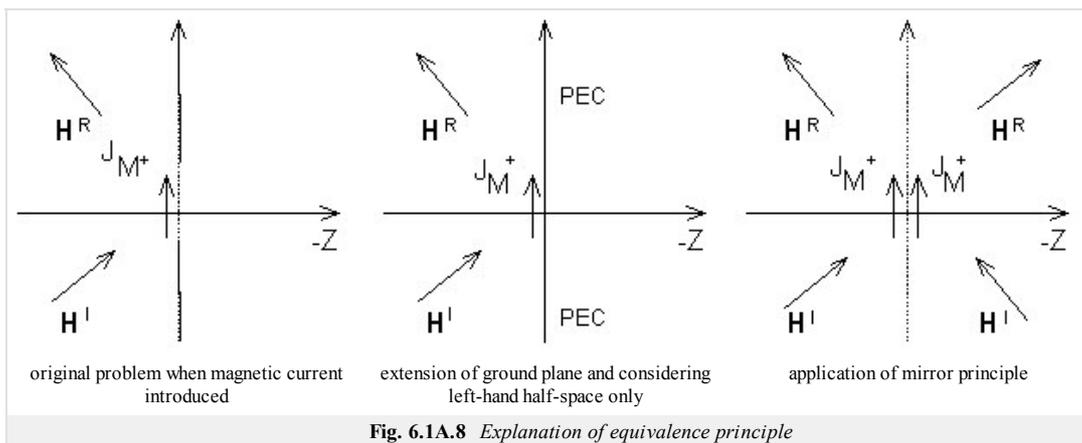

original problem when magnetic current introduced

extension of ground plane and considering left-hand half-space only

application of mirror principle

Fig. 6.1A.8 *Explanation of equivalence principle*

In analogy to metallic elements, dual formulation of transmission coefficient is possible τ.

In fig. **6.1A.9**, an example for normal incidence of a wave to a frequency selective surface with slots in vacuum ($b = 21$ mm, $a = 7.5$ mm, $b' = 19.75$ mm, $a' = 1.5$ mm) and on dielectric substrate ($b = 14$ mm, $a = 7.5$ mm, $b' = 13.25$ mm, $a' = 1.5$ mm and the height of the substrate $d = 1.57$ mm) are given. Obviously, selectivity of transmission coefficient is better for dielectrics (in case of metallic-element surface, the situation was opposite). Next, influence of dielectrics is different compared to metallic-element surface, and therefore, slots are of different dimensions compared to metallic-element surface with dielectrics. Dielectrics even cause non-zero value of reflection coefficient on resonant frequency. Even here, parasitic modes appear.

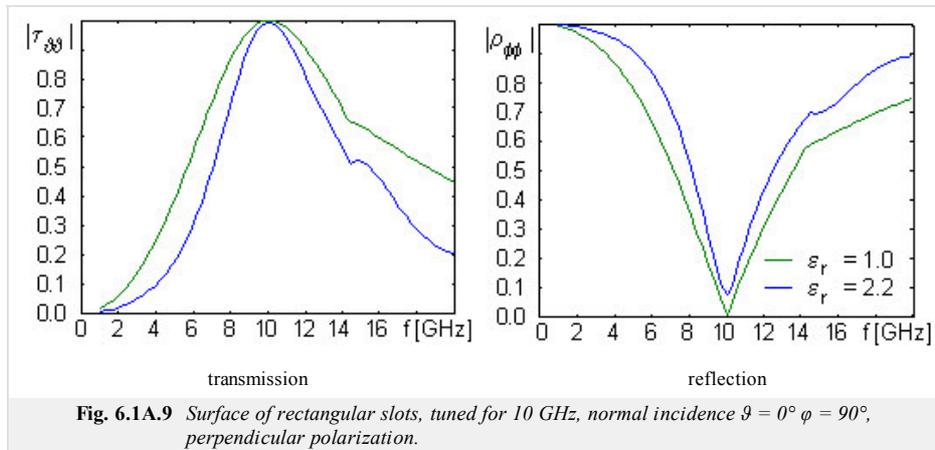**Conclusion**

The method of the analysis of frequency selective surfaces, given in this chapter, does not include the influence of dielectrics, but given examples do that [17]. Another way of the description of the influence of dielectrics can be found in [19].

Real applications use several sequentially lined selective surfaces. That way, selectivity can be significantly improved. Design of such surfaces is rather complicated.

Frequency selective surfaces can be modeled in commercial programs based on finite elements (HFSS) or on moment method (IE3D). In both cases, only a single element is analyzed thanks to the periodicity.

Designing frequency selective surfaces, physical basis of their behavior and limitations have to be known.

Fig. 6.1A.9 *Surface of rectangular slots, tuned for 10 GHz, normal incidence $\vartheta = 0° \varphi = 90°$, perpendicular polarization.*

6.1 Analysis of frequency selective surfaces

Advanced theory

In this part of the textbook, we take a detailed look on basis functions, which can be used when numerically analyzing frequency selective surfaces (current distribution over an element is sought).

In the layer A, we derived a relation for electric field of a frequency selective surface, which consists of metallic elements, in the domain of spatial spectra

$$-\frac{1}{2\omega\varepsilon}\sum_{m,n}\begin{bmatrix} \dfrac{k^2-\alpha_m{}^2}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} & \dfrac{-\alpha_m\beta_n}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} \\[3mm] \dfrac{-\alpha_m\beta_n}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} & \dfrac{k^2-\beta_n{}^2}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} \end{bmatrix}\begin{bmatrix} J_x(\alpha_m,\ \beta_n) \\[2mm] J_y(\alpha_m,\ \beta_n) \end{bmatrix}\exp\big[j(\alpha_m x+\beta_n y)\big]=-\begin{bmatrix} E_x^I(x,\ y) \\[2mm] E_y^I(x,\ y) \end{bmatrix}. \qquad (\,6.1B.1\,)$$

In the above-given relation, ω denotes angular frequency of a wave, ε is permittivity of the surrounding of selective surface, α_m and β_n are spatial frequencies, k denotes free-space wave number, J_x and J_y are components of current density vector over the element, and finally, E_x^I and E_y^I denote components of electric field intensity of incident wave.

Solving the problem, we approximate the unknown current distribution in (6.1B.1) in terms of properly chosen basis functions and unknown approximation coefficients. Such formal approximation is substituted to the solved equation (6.1B.1). Since the approximation does not meet the solved equation exactly, we respect this fact by introducing a residual function (residuum). Smaller values the residual function reaches, more accurate the approximation is. The residuum is minimized by Galerkin method (residuum is sequentially multiplied by as many basis functions as many unknown approximation coefficients have to be computed; that way the set of N linear algebraic equations for N unknown approximation coefficients is obtained).

There are two approaches for choosing basis functions. The first one exploits basis functions, which are non-zero over the whole analyzed area. Functions are elected to physically represent standing waves of a current on an element.

The second approach divides the analyzed region to sub-regions, where current is approximated in terms of basis functions, which functional value is non-zero over a given sub-region only. This approach is advantageous in simple analysis of selective surfaces consisting of arbitrarily shaped elements.

Harmonic basis functions

Assume a frequency selective surface consisting of rectangular perfectly conductive elements of the dimensions a' and b'. Components of current density J are approximated as follows [17]

$$\mathbf{J}(x,\ y)=\sum_{p=0}^{P}\sum_{q=0}^{Q}A_{pq}\Psi_{pq}^{TM}+F_{pq}\Psi_{pq}^{TE}, \qquad (\,6.1B.2\,)$$

where

$$\Psi_{pq}^{TE}(x,\ y)=\left[\frac{p\pi}{a'}\sin\left(\frac{p\pi x}{a'}\right)\cos\left(\frac{q\pi y}{b'}\right)\mathbf{u}_x+\frac{q\pi}{b'}\cos\left(\frac{p\pi x}{a'}\right)\sin\left(\frac{q\pi y}{b'}\right)\mathbf{u}_y\right]\exp\big[j(\alpha_0 x+\beta_0 y)\big],$$
$$\Psi_{pq}^{TM}(x,\ y)=\left[\frac{p\pi}{b'}\sin\left(\frac{p\pi x}{a'}\right)\cos\left(\frac{q\pi y}{b'}\right)\mathbf{u}_x-\frac{q\pi}{a'}\cos\left(\frac{p\pi x}{a'}\right)\sin\left(\frac{q\pi y}{b'}\right)\mathbf{u}_y\right]\exp\big[j(\alpha_0 x+\beta_0 y)\big]. \qquad (\,6.1B.3\,)$$

Here, $\Psi_{pq}{}^{TE}$ and $\Psi_{pq}{}^{TM}$ are two different functions for parallel polarization and perpendicular one. These functions differ in two aspects:

- Indeces p, q for the function $\Psi_{pq}{}^{TE}$ can reach zero, and for $\Psi_{pq}{}^{TM}$ are one and higher.
- Component J_y for perpendicular polarization has to be of the opposite sign than J_y for parallel polarization.

In order to get better notion about the difference between currents, which are excited by electric intensity of incident wave, which is parallel (perpendicular) to the plane of incidence, we depict directional fields of current density of mode (1,1) for parallel polarization and perpendicular one in fig. **6.1B.1**.

a) b)
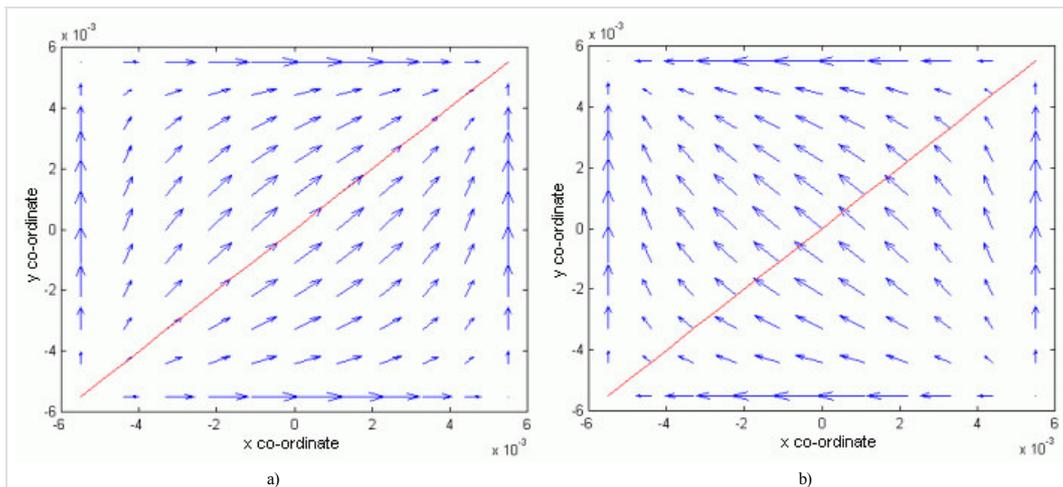
Fig. 6.1B.1 *Directional field of current density of the mode (1,1)*
a) for parallel polarization
b) for perpendicular one.
The cell is of dimensions a = b = 12mm, the metallic element is of dimensions a' = b' = 12mm; f=10GHz, ϑ = 1°, φ = 45°, d = 1.5mm, ε_r = 3.7. Magnitude of marginal currents was increased with respect to reality. Plane of incidence is depicted in red.

Combination of harmonic functions and Chebyshev polynomials

$$\mathbf{J}^{TE}=\sin\left[\frac{p\pi}{a'}\left(x+\frac{a'}{2}\right)\right]\frac{T_q(2y/b')}{\sqrt{1-(2y/b')^2}}\,\mathbf{x}+\frac{T_p(2x/a')}{\sqrt{1-(2x/a')^2}}\sin\left[\frac{q\pi}{b'}\left(y+\frac{b'}{2}\right)\right]\mathbf{y} \qquad (\,6.1B.4a\,)$$

$$\mathbf{J}^{TM} = \sin\left[\frac{p\pi}{a'}\left(x + \frac{a'}{2}\right)\right]\frac{T_q(2y/b')}{\sqrt{1-(2y/b')^2}}\mathbf{x} - \frac{T_p(2x/a')}{\sqrt{1-(2x/a')^2}}\sin\left[\frac{q\pi}{b'}\left(y + \frac{b'}{2}\right)\right]\mathbf{y}$$ (6.1B.4b)

In contrast to harmonic basis functions, the function cosine is replaced by *Chebyshev polynomial* in order to reach infinite values at the edges of elements. That way, currents flowing along edges are correctly modeled.

Now, eqn. (6.1B.1) is solved by Galerkin method: eqn. (6.1B.1) is sequentially multiplied by basis functions and the product is integrated over the surface of the element. For purely harmonic basis functions, we get:

$$\int_a^b w_m(z)R(z)dz = 0 \quad m = 0,1,\ldots N,$$

$$\sum_{p,q} A_{p,q}\left\{\sum_{m,n} \Psi_{rs}^{TE*}(a_m,\beta_n)K(a_m,\beta_n)\Psi_{pq}^{TE*}(a_m,\beta_n)\right\} + F_{pq}\left\{\sum_{m,n} \Psi_{rs}^{TE*}(a_m,\beta_n)K(a_m,\beta_n)\Psi_{pq}^{TE*}(a_m,\beta_n)\right\} = \frac{-1}{ab}\iint\limits_{p\check{r}es\,element} E_{\tan}^I(x,y)\Psi_{pq}^{TE*}(x,y)dxdy,$$ (6.1B.5a)

$$\sum_{p,q} A_{p,q}\left\{\sum_{m,n} \Psi_{rs}^{TM*}(a_m,\beta_n)K(a_m,\beta_n)\Psi_{pq}^{TM*}(a_m,\beta_n)\right\} + F_{pq}\left\{\sum_{m,n} \Psi_{rs}^{TM*}(a_m,\beta_n)K(a_m,\beta_n)\Psi_{pq}^{TM*}(a_m,\beta_n)\right\} = \frac{-1}{ab}\iint\limits_{p\check{r}es\,element} E_{\tan}^I(x,y)\Psi_{pq}^{TM*}(x,y)dxdy.$$ (6.1B.5b)

If the above-given set is solved out, electric intensity $\mathbf{E}^S$ in spectral domain is computed. Intensity $\mathbf{E}^S$ appears at the left-hand side of eqn. (6.1B.1)

$$\mathbf{E}^S = -\frac{1}{2\omega\varepsilon}\sum_{m,n}\begin{bmatrix}\dfrac{k^2-\alpha_m{}^2}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} & \dfrac{-\alpha_m\beta_n}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} \\ \dfrac{-\alpha_m\beta_n}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}} & \dfrac{k^2-\beta_n{}^2}{\sqrt{k^2-\alpha_m{}^2-\beta_n{}^2}}\end{bmatrix}\begin{bmatrix}J_x(\alpha_m,\beta_n) \\ J_y(\alpha_m,\beta_n)\end{bmatrix}\exp[j(\alpha_m x + \beta_n y)] = -\begin{bmatrix}E_x^I(x,y) \\ E_y^I(x,y)\end{bmatrix}.$$ (6.1B.6)

Computing electric intensity of scattered wave $\mathbf{E}^S = \mathbf{E}^R$, only the basic harmonic component on the frequency α_0, β_0 is assumed. If the period a is small (smaller than one half of the wavelength) and frequency approaches the first resonance of the surface, then a uniform field of the scattered wave exists in the far field region only. If the condition is not met, then parasitic modes can be excited (higher spatial frequencies are not filtered by the near-field zone, and corresponding waves propagate further).

Let us explain the physical notion of parasitic modes. Explanation is illustrated by fig. **6.1B.2**. Here, a two-element antenna array is depicted, which is illuminated by a plane wave arriving from the direction ϑ. After the incidence, the wave is reflected according to rules of geometric optics. A part of energy can propagate even in the direction of the parasitic (grating) lobe.

Fig. 6.1B.2 *Explanation of forming parasitic lobes (two-element antenna)*
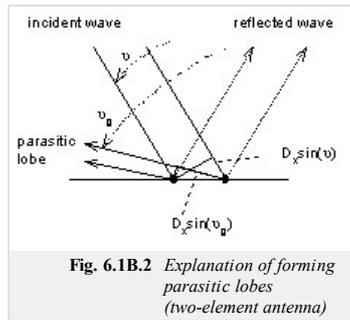
Formation of a parasitic lobe is conditioned by reaching a phase shift, which equals to the integral multiple of 2π by the beam no. 1 on a bold trajectory. From the mathematical point of view:

$$\beta D_x \sin(\vartheta_g) + \beta D_x \sin(\vartheta) = n2\pi.$$ (6.1B.7)

The direction, where the parasitic lobe appears, depends on the frequency of the incident wave and on the distance of antenna elements only. The lowest frequency of exciting the parasitic mode for a given angle of incidence, is for $\vartheta_g = 90°$:

$$f_{g0} = \frac{nc}{D_x(\sin\vartheta+1)}.$$ (6.1B.8)

if the distance of antenna elements equals to $a = 21$ mm and if perpendicular incidence of wave are assumed, then the first parasitic lobe is excited on the frequency

$$f_g = \frac{1\cdot 3\cdot 10^8}{0{,}021\cdot(\sin 0+1)} = 14{,}3\,GHz.$$ (6.1B.9)

6.1 Analysis of frequency selective surfaces

Matlab program

Spectral domain moment method is implemented in the computer program for analysis frequency selective surfaces (FSS) with the rectangular patches. This program is composed from the several separated m-files. These m-files are called after setting input parameters and after request for the computation. The N-dimensionals Matlab arrays and the nested loops are used. Program is controlled from user friendly GUI interface.

Program Description

The computer program for analysis frequency selective surfaces with rectangular patches can computes the magnitude and the angle of the reflection coefficient, transmission coefficient in decibels and the current distribution on the conductive patches. The spectral domain method for the computing these parameters is used. For the approximation of the current distribution are available two types of the basis functions in Popupmenu at the top of the window (fig. **6.1C.1**).

The main window includes Edits for entering the dimensions of the cells and for the dimensions of the conductive patches. Meanings of the parameters are illustrated on the picture, which represents the cell and the patch of FSS. At the right side of the window are Edits for the setting the step and the interval of the frequencies, which will be used for the analysis. The program enables setting the perpendicular or the parallel polarization and the angles of the incidence (theta and phi are common angles in the spherical coordinates). The Edits named as M, N denotes the numbers of the harmonics for spatial frequencies in the directions x and y, respectively. Another Edits named as P, Q are for entering the level of the basis functions. At the bottom

Fig. 6.1C.1 *GUI of the program in Matlab*

of the window are Edits for the setup the depiction of the current distribution. The parameter Nx and Ny sets the density of the depiction and the parameter named as Step_I is step of the frequencies for which is the current distribution depicted. When we have set all these parameters we can choose the required parameter by the Checkboxes and click on the button $\boxed{\text{Analyze}}$. The program is closed by the click on the button $\boxed{\text{Close}}$.
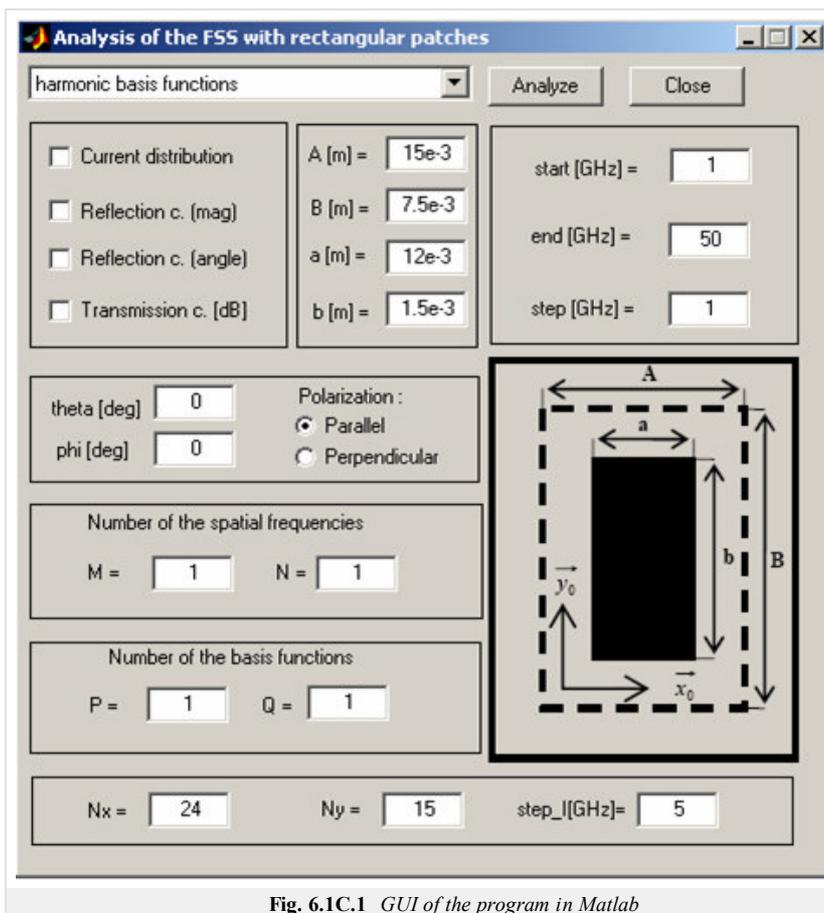
Results of the Analysis

In this section the frequency selective surface with rectangular patches is analyzed with the program in the Matlab and with the ANSOFT Designer. The results of analyses and the computing times are compared. The conductive patch is of the dimensions a = 12 mm, b = 1.5mm, the cell is of the dimensions A = 15 mm, B = 7.5 mm. Magnitude and angle of the reflection coefficient computed in our program are depicted in fig. **6.1C.2**. The FSS was analyzed for parallel polarization with angles of incidence $\varphi = 0°$, $\vartheta = 0°$ in the interval (5 ÷ 50) GHz with the step 1GHz. We considered M = N = P = Q = 1. The results of the analysis from program in Matlab and from ANSOFT Designer are almost identical. When we increas the number of spatial frequencies to M = N = 5 and level of basis functions to P = Q = 4 we obtain the results in the fig. **6.1C.4**. The results in the fig. **6.1C.4** and in the fig. **6.1C.2** are again almost identical, so increasing the number of the spatial frequencies and level of the basis functions was not necessary. We are trying to use lower numbers of the spatial frequencies and lower levels of the basis functions to save the computing time. The computing times of analyses are presented in the Tab.1. The computing time of analysis in the ANSOFT Designer was 149 sec. The similar results was computed in Matlab in 1.6 sec (see fig. **6.1C.2** and fig. **6.1C.3**). We have to consider, that ANSOFT Designer computing all parameters of the FSS for both possible polarizations in the same time. In the figure are depicted computing times depending on the parameter n = (M = N = P = Q) for both types of the basis functions. The computing times of the basis functions with Chebyshev polynomials are lower then the computing times of the harmonic basis functions.

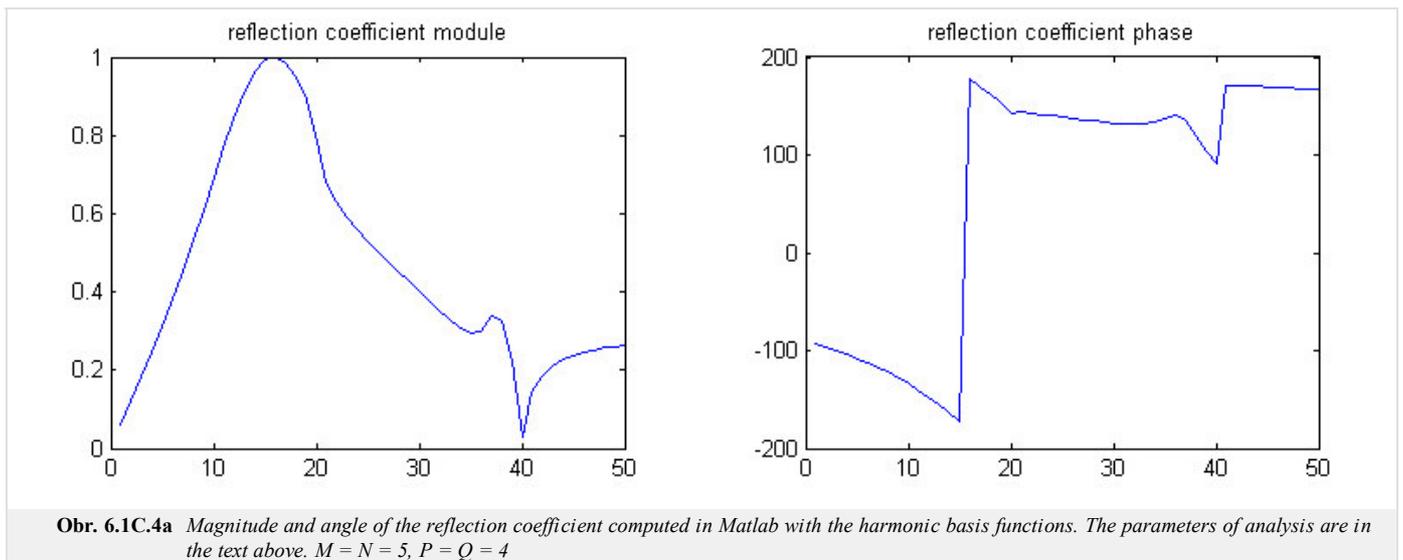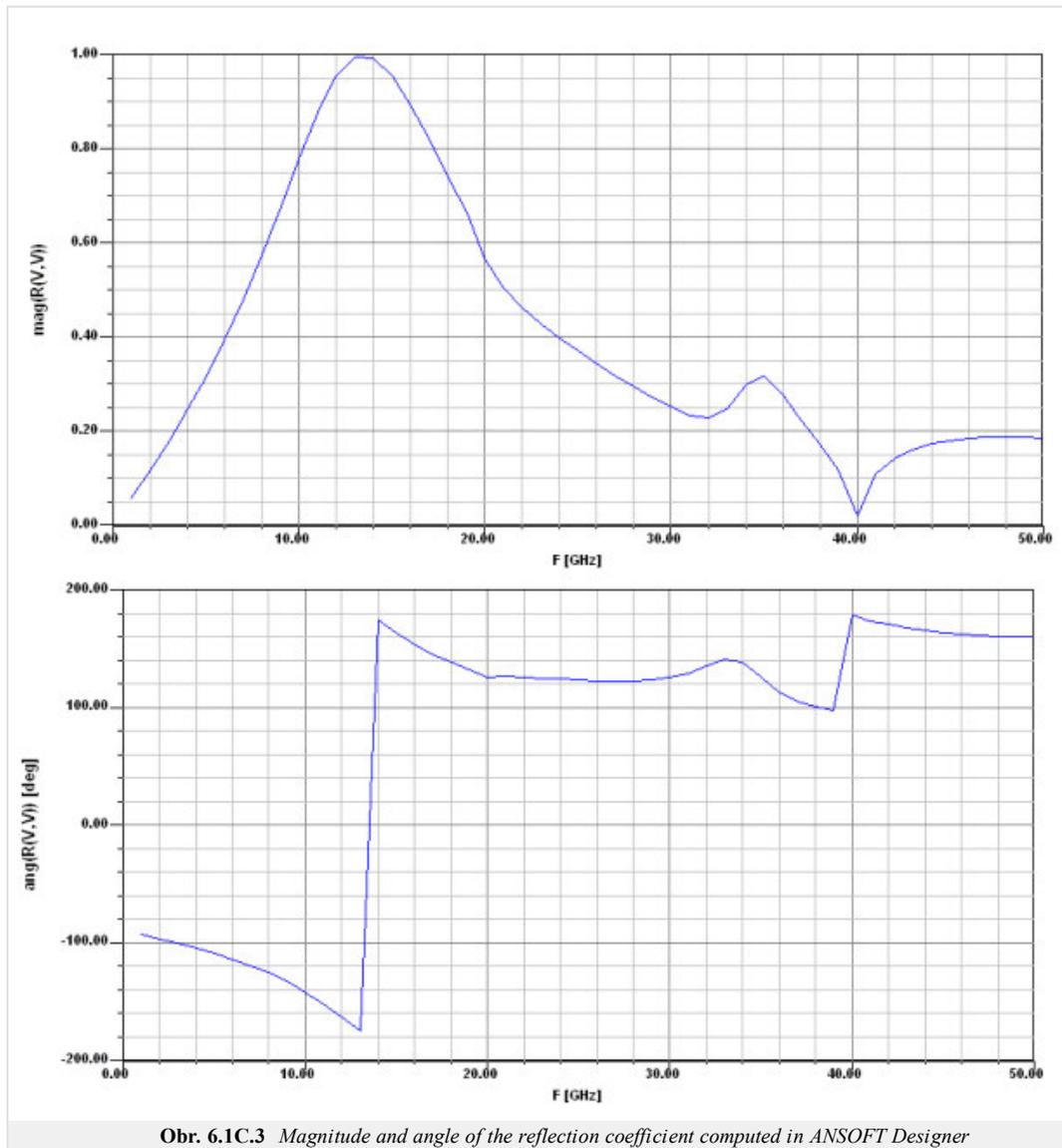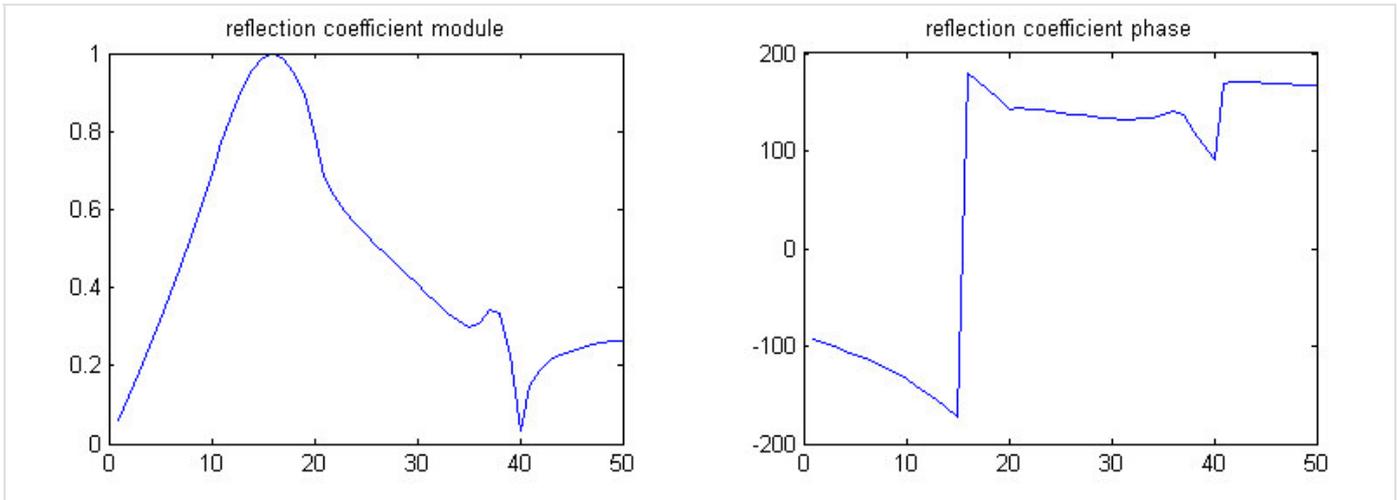| Tab. 6.1C.1 *Comparsion of the computing times for harmonic basis functions and for the basis functions with Chebyshev polynomials. The parameter n is equals M = N = P = Q. The computing time was measured for analysis in the interval (1÷50) GHz with step 1 GHz.*					
n [-]	**1**	**2**	**3**	**4**	**5**
t [s] (basis functions with Chebyshev polynomials)	1,6	10,4	60,8	251,5	807,5
t [s] (harmonic basis functions)	1,6	10,5	62,1	268,9	908,4
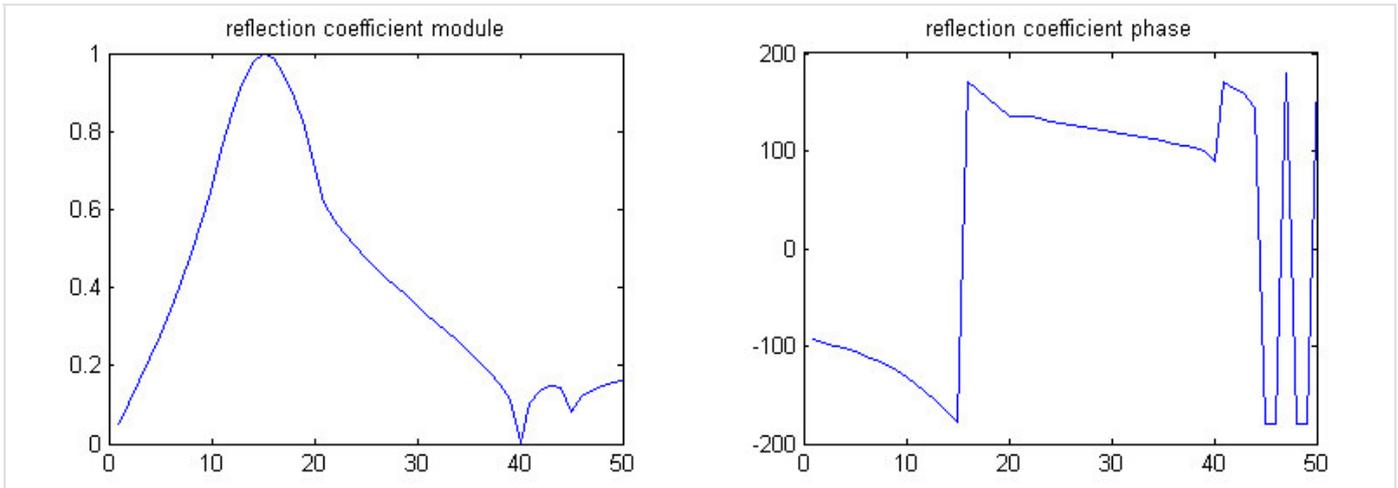
Obr. 6.1C.2a *Magnitude and angle of the reflection coefficient computed in Matlab with the harmonic basis functions. The parameters of analysis are in the text above. M = N = P = Q = 1*
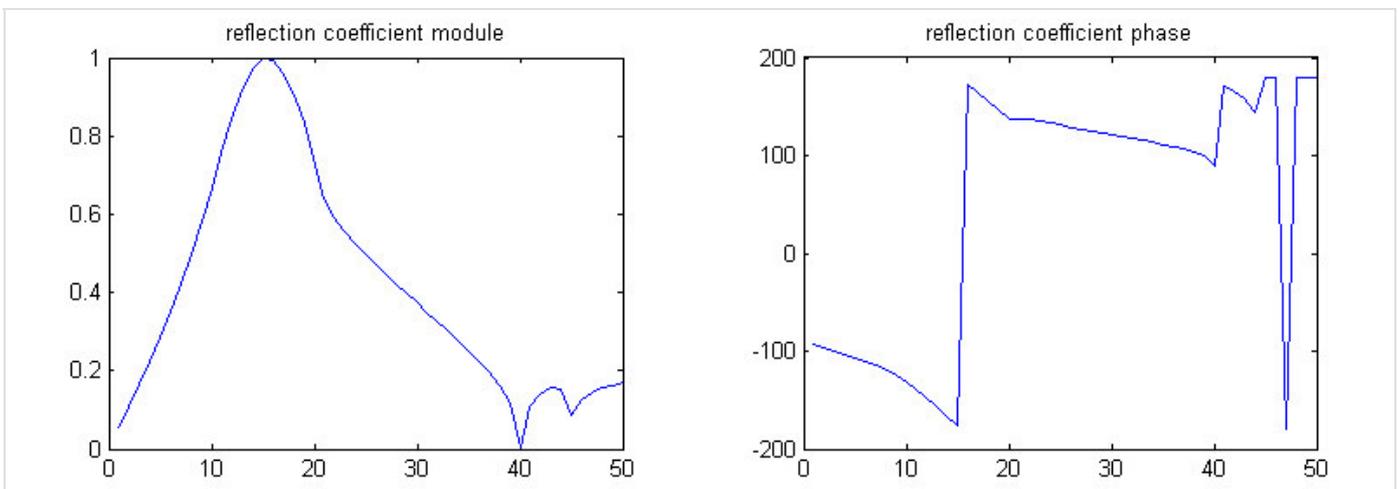

Obr. 6.1C.2b *Magnitude and angle of the reflection coefficient computed in Matlab with the basis functions with Chebyshev polynomials. The parameters of analysis are in the text above. M = N = P = Q = 1*

Obr. 6.1C.3 *Magnitude and angle of the reflection coefficient computed in ANSOFT Designer*

Obr. 6.1C.4a *Magnitude and angle of the reflection coefficient computed in Matlab with the harmonic basis functions. The parameters of analysis are in the text above. M = N = 5, P = Q = 4*

Obr. 6.1C.4b *Magnitude and angle of the reflection coefficient computed in Matlab with the basis functions with Chebyshev polynomials. The parameters of analysis are in the text above. M = N = 5, P = Q = 4*
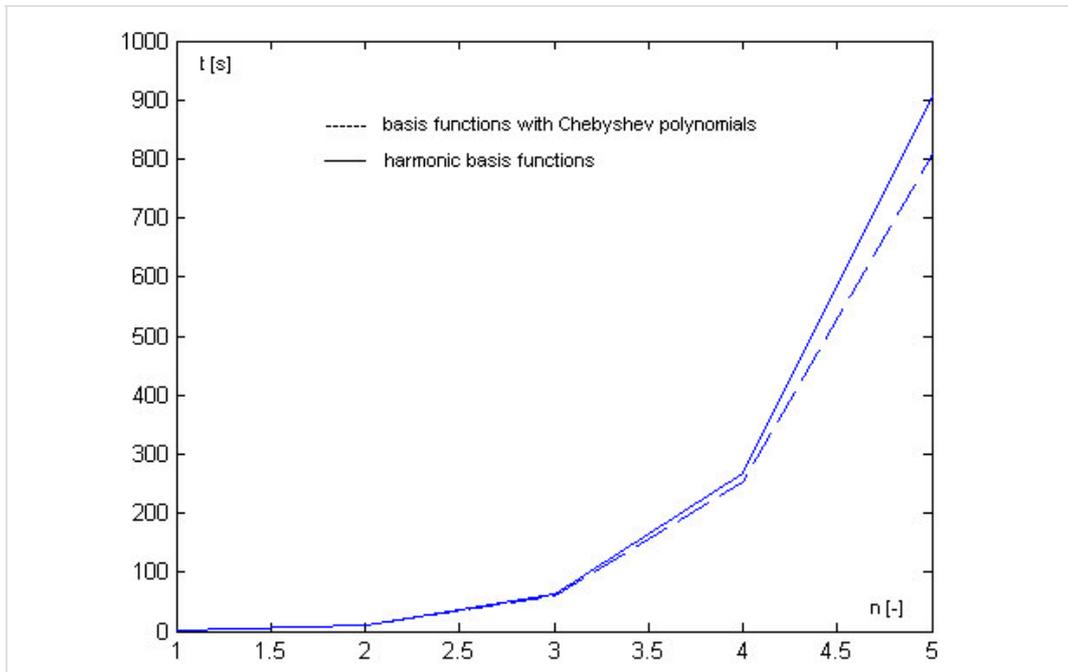
Fig. 6.1C.5 *Comparsion of the computing times of the analyses with the harmonic basis functions with Chebyshev polynomials.*

6.1 Analysis of frequency selective surfaces

Java applet

6.1 Analysis of frequency selective surfaces

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

Frequency selective surfaces …

Possible answers for question 1:
- ○ … are used to amplify electromagnetic waves.
- ○ … can behave as a perfect mirror at one frequency, and as a free space at another frequency.
- ○ … are sensitive to frequencies and insensitive to polarizations of waves.
- ◉ I won't answer this question

Question 2

For the analysis of FSS, we usually use …

Possible answers for question 2:
- ○ … the frequency domain method of moments.
- ○ … the time-domain moment method.
- ○ … the spectral domain method of moments.
- ◉ I won't answer this question

Question 3

The suppression of wave transitions at given frequencies is caused …

Possible answers for question 3:
- ○ … by destructive interferences of waves radiated by currents induced on FSS elements.
- ○ … increasing losses in metallic and dielectric parts of the FSS.
- ○ … by the critical frequency of waveguides formed by FSS elements.
- ◉ I won't answer this question

Question 4

Current distribution on the FSS elements can be approximated in terms of …

Possible answers for question 4:
- ○ … harmonic and Chebychev basis functions.
- ○ … piecewise constant basis functions.
- ○ … Dirac pulses.
- ◉ I won't answer this question

Question 5

Analysis of rectangular slots in a metallic plane and rectangular patches on a dielectric plane …

Possible answers for question 5:

- ○ … are the same thanks to the duality principle.

- ○ … have to be performed differently due to the different boundary conditions.

- ○ … are similar, but the magnetic currents require small changes of the algorithm.

- ● I won't answer this question

<div align="center">

see result

</div>

Chapter 7: Radio optics

Introduction

The term electric signal usually denotes a time-variant voltage $U(t)$, current $I(t)$ or electric field intensity $E(t)$. Time dependency of those quantities contains an information.

Every electric signal can be composed of a finite or infinite number of harmonic voltages (currents, intensities) of various frequencies, amplitudes and initial phases. Amplitudes and phases of those spectral components are given by the spectral function $S(\omega)$. The signal can be therefore described by the relation

$$U(t) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} S(\omega) e^{j\omega t} d\omega. \tag{7.1}$$

In that relation, ω is angular frequency and t is time.

The signal is said to be given by the addition of its spectral components or by inverse Fourier transform of its spectral function (according to eqn. (7.1)). Since a simple harmonic oscillation is the *basic building stone* of a signal, we consider this oscillation as elementary signal even it does not bear any information. The elementary signal is determined by its amplitude, phase and (temporal) frequency f, or time period $T = 1/f$.

Electromagnetic wave can contain information in time dependency of electric field intensity $E(t)$. Since the wave exists in the space, information can be hidden even in the *spatial* distribution of intensity. Then, we speak about optical signal. The spatial signal is richer than the electric one because it is a function of two variables (x, y). It is associated with an electromagnetic wave (with amplitude distribution of field intensity in the plane, which is perpendicular to the direction of propagation).

In analogy to the elementary electric signal, elementary spatial signal can be introduced. If coordinate system x, y is given in a plane, which is perpendicular to the direction of propagation, then a pure harmonic variation of amplitude of field intensity in an axis direction is elementary optical signal. In the direction of the second axis, amplitude of field intensity is constant (independent to y). If such wave illuminates a projection screen, a set of dark and light strips of unclear transitions can be seen (the darkest strips correspond to the minimum of the harmonic function, the lightest strips to the maximum). Distance between two neighboring strips is called spatial period T_x. Reverse value of the spatial period is called spatyial frequency f_x. From the mathematical point of view, elementary optical signal for *x*-direction can be expressed as follows:

$$E(x) = E_{mx}\left[1 + \sin(2\pi f_x x)\right] = E_{mx}\left[1 + \sin(\omega_x x)\right]. \tag{7.2}$$

For *y*-direction:

$$E(y) = E_{my}\left[1 + \sin(2\pi f_y x)\right] = E_{my}\left[1 + \sin(\omega_y x)\right]. \tag{7.3}$$

Composing these two elementary signals together, two-dimensional spatial signal is obtained:

$$E(x,\,y) = E_m\left[1 + \sin(\omega_x x + \omega_y y)\right]. \tag{7.4}$$

Optical signals can be written in symbolic way. Then, we get:

$$E(x,\,y) = E_m\left[1 + \exp(j\omega_x x) + \exp(j\omega_y y)\right]. \tag{7.5}$$

In analogy to the time domain, every spatial signal $E(x, y)$ can be composed of elementary spatial signals of different spatial frequencies, of different amplitudes and different initial phases. Fourier transform corresponding to (7.1) is, of course, two-dimensional.

Elementary spatial signal is non-uniform plane wave. Handling those waves is rather complicated, and therefore, non-uniform waves are decomposed to three uniform plane waves. In fig. **7.1**, blue arrows denote directions of propagation of three paraxial waves, which amplitudes are $E/2$, E and $E/2$, and which propagate in three symmetric directions with respect to the axis z. Symbols **k** denote wave vectors of respective waves. Now, the resultant field distribution in the plane xy in the parallel direction to the axis y is computed. For that purpose, projections of wave vectors to the direction y have to be determined:
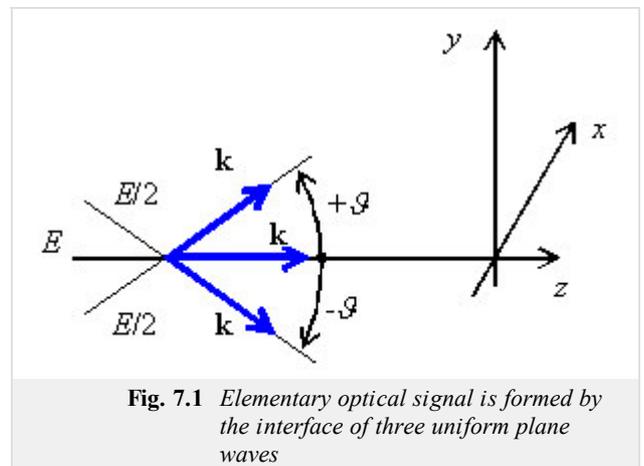
Fig. 7.1 *Elementary optical signal is formed by the interface of three uniform plane waves*

$$\tag{7.6}$$

The computed intensity is given by the addition of three plane waves, which propagate in the direction y , and which are described by wave numbers k_{1y}, k_{0y}, k_{2y}:

$$E\left(y\right) = \frac{E_m}{2}\exp\left(jk_{1y}y\right) + E_m + \frac{E_m}{2}\exp\left(jk_{2y}y\right) =$$

$$= E_m\left[\frac{\exp\left(+jk\sin\vartheta y\right)}{2} + 1 + \frac{\exp\left(-jk\sin\vartheta y\right)}{2}\right] = \qquad (\,7.7\,)$$

$$= E_m\left[1 + \cos\left(k\sin\vartheta y\right)\right]$$

That way, we computed the cosine distribution of the amplitude Ey, and consequently, the elementary optical signal of the spatial frequency $\omega_y = k.\sin\vartheta$. Now, a general spatial signal can be decomposed by Fourier transform to elementary optical signals, and each elementary signal can be further decomposed to the triplet of uniform waves.

The above-described principles form the basics of radio optics - on radio frequencies, the spatial distribution of signals is handled the same way as in optics. Considering the electromagnetic nature of the spatial signal, there is no reason to deny this approach.

In the par. 7.1, we describe Gaussian beam - a narrow beam of coherent radiation. As an example of Gaussian beam, a laser beam of the basic mode TEM$_{00}$ can be considered, which leaves the source with planar equiphase surface perpendicular to the direction of propagation. The beam is not intentionally modulated, but the transversal field distribution is not constant. The maximum intensity appears as the axis of the beam and the intensity decreases towards the margin of the beam following the Gaussian function. Gaussian beam is let to pass a system of spatial layers and lenses, and variation of its parameters is observed. The task is typical for optical applications, but we can meet it even on *low* (radio) frequencies.

Par. 7.2 deals with the passage of Gaussian beam through various spatial elements (lenses, layers of media, etc.), and with the description of those phenomena by matrix notation. Gaussian beam is described by the vector of parameters, optical elements are described by matrices. Multiplying a vector of Gaussian beam parameters at the input of an optical element by the matrix of an optical element, a vector of Gaussian beam parameters at the output of an optical element is obtained.

7.1 Gaussian beam

Basic theory

Investigating parameter variation of coherent optic beam, which passes an optical system, the simplest laser beam TEM$_{00}$ is assumed. When leaving the source, the beam is *parallel* to the optical axis (its planar equiphase surface is perpendicular to the direction of propagation). Even if the beam is not modulated, transversal distribution of field intensity is not constant.

Waves, which equiphase surface normal declines for a very small angle form the optical axis (axis z), are called paraxial waves. Those waves have to meet the equation

$$\nabla_{\mathrm{T}}^2 A - j2k\frac{\partial A}{\partial z} = 0, \qquad (\text{7.1A.1})$$

where

$$\nabla_{\mathrm{T}}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

is the transversal part of Laplace operator, z is a coordinate of the longitudinal axis, k denotes wave number, A is field intensity.

Gaussian beam is one of possible solutions of eqn. (7.1A.1). Power of Gaussian beam is concentrated into a narrow cone. Field distribution in an arbitrary transversal plane is given by a circularly symmetric Gaussian function; the axis of symmetry is identical with the optical axis. The width of Gaussian beam is minimal is a so-called neck (here, the equiphase surface is planar). The width gradually increases on both the sides of the neck (equiphase surfaces are gradually curved).

As already noted, transversal distribution of field intensity of the basic mode TEM$_{00}$ is described by Gaussian function (the maximum intensity on the axis gradually decreases towards margin):

$$E(x,\, y) = E_{max}\exp\!\left(-\frac{\rho^2}{a_0^2}\right) = E_{max}\exp\!\left(-\frac{x^2+y^2}{a_0^2}\right), \qquad (\text{7.1A.2})$$

where r denotes the radial distance of the point $(x,\, y)$ from the beam axis, and a_0 is so-called beam radius (a radial distance from the beam axis, where field intensity decreases to the value E_{max}/e; $e = 2.718...$.

Complex amplitude

Assume a paraxial plane wave $\exp(-jkz)$, where $k = 2\pi/\lambda$ is wave number, λ denotes wavelength and z is optic axis coordinate. The wave is modulated by the envelope $A(r, z)$, which varies relatively slowly in the direction of the optical axis z. Then, the complex amplitude meets

$$U(r) = A(r,\, z)\exp(-jkz). \qquad (\text{7.1A.3})$$

The envelope is assumed to stay approximately constant when the distance changes for $\Delta z = \lambda$. We speak about a local plane wave, which equiphase surface normals form paraxial beams. Rearranging equations, the definition of the complex envelope of Gaussian beam is obtained:

$$A(r,\, z) = \frac{A_1}{q(z)}\exp\!\left[-jk\frac{\rho(r)^2}{2q(z)}\right], \quad q(z) = z + jz_0, \quad \rho^2 = x^2 + y^2. \qquad (\text{7.1A.4})$$

Here, z_0 is Rayleigh distance.

In order to separate the amplitude and the phase of a complex envelope, function $1/q(z)$ is rewritten to the real part $R(z)$ and to the imaginary one, which is represented by the function $W(z)$. Hence,

$$\frac{1}{q(z)} = \frac{1}{R(z)} - j\frac{1}{\pi W^2(z)}. \qquad (\text{7.1A.5})$$

The function $R(z)$ describes the half-width of Gaussian beam, and $W(z)$ is curvature radius of the beam equiphase surface.

For beam parameters, we can further define:

$$W(z) = W_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2}, \qquad (\text{7.1A.6})$$

$$R(z) = z\left[1 + \left(\frac{z}{z_0}\right)^2\right], \qquad (\text{7.1A.7})$$

$$W_0 = \sqrt{\frac{\lambda z_0}{\pi}}. \qquad (\text{7.1A.8})$$

Properties of Gaussian beam

Gaussian beam is uniquely determined by the following parameters:

- Intensity of radiation
- Power of the beam
- Radius of the beam

The above-given parameters are going to be explained more in detail now.

Intensity of radiation

Intensity of radiation is a function of the axial distance z and the radial one $\rho = (x^2 + y^2)^{1/2}$,

$$I(\rho, z) = I_0 \left[\frac{W_0}{W(z)}\right]^2 \exp\left[-\frac{2\rho^2}{W^2(z)}\right].$$ (7.1A.9)

At the beam axis ($\rho = 0$), intensity is of the maximal value I_0 for $z = 0$. Increasing z, intensity gradually decreases, and for $z = \pm z_0$, the intensity is of the half of the maximum value I_0.

Power of the beam

The total power, which is transmitted by the beam, is given by the integral of the product of the beam intensity and its transversal surface

$$P = \int_0^\infty I(\rho, z) 2\pi\rho d\rho.$$ (7.1A.10)

At the axis, we get

$$P = \frac{1}{2} I(0,z)\pi [W(z)]^2.$$ (7.1A.11)

Inside a circle of the radius $\rho_0 = W(z)$, approximately 86 % of the total power is transmitted. Through a circle of the radius 1.5 $W(z)$, approximately 99 % of power passes

Since Gaussian beams are often characterized by the transmitted power P, the intensity I is useful to be expressed as a function of P

$$I(\rho, z) = \frac{2P}{\pi W^2(z)} \exp\left[-\frac{2\rho^2}{W^2(z)}\right].$$ (7.1A.12)

Radius of the beam

In every transversal plane, the intensity is maximal at the optical axis (z). Since most power propagates in the area of the radius $W(z)$, we understand $W(z)$ as the radius of the beam. Dependency of the radius on the coordinate z is given by the relation

$$W(z) = W_0 \sqrt{1 + \left(\frac{z}{z_0}\right)^2}.$$ (7.1A.13)

In the plane z = 0, the radius is of the minimal walue W_0, which is called the position of the maximum necking of the beam.

Further information can be found in **[16]**.

7.1 Gaussian beam

Matlab program

The program serves to computing and depicting parameters of Gaussian beam when passing an optical element.

First, the path of Matlab has to be directed to the folder **Gauss**. The program is run via the m-file `gauss.m`. Passing the introductory window, the following input parameters are required to be typed: wavelength, radius of the beam, and radius of the equiphase surface. Then, the type of the optical element has to be chosen from a combo. Following types are at our disposal:

- Free space
- Planar mirror
- Spherical mirror
- Planar boundary
- Spherical boundary

For each of those elements, supplementary parameters have to be determined. Among those parameters, the distance behind the optical element, where the radius of the equiphase surface and the radius of the beam are going to be observed, has to be given. Then, the graphical representation of the beam is displayed, and in the table, the beam radius R_2 and the equiphase surface radius W_2 depending on the distance z from the optical element are listed. Moreover, the button for finishing the program (End) and for entering new parameters are shown.

7.1 Gaussian beam

Developing Matlab

In the program, ABCD law for passing Gaussian beam trough an optical system, is exploited. Entering wavelength, radius of the beam W_1, and radius of the equiphase surface R_1, complex curvature at the beginning of the system q_1 is evaluated, which consists of a real component a_1 and of an imaginary one b_1:

```
a1 = k^2 * R1 * W1^4 / (k^2 * W1^4 + 4*R1^2);
b1 = -2*k * R1^2 * W1^2 / (k^2 * W1^2 + 4*R1^2);
```

This complex parameter is used for evaluating beam curvature behind the optical element

```
q2 = (A*q1 + B) / (C*q1 + D);
```

Here, A, B, C, D are elements of the matrix of the optical element.

In order to determine the beam radius W_2 and the equiphase surface radius R_2 behind the optical element, following relations are used:

```
W2 = sqrt( (2*(A*a1+B)^2 + (A*b1)^2) / (k*b1*((A*a1+B)*C - A*(C*a1+D))));
R2 = ((A*a1+B)^2+(A*b1)^2)/((A*a1+B)*(C*a1+D)-A*(C*a1+D));
```

Next, variation of parameters W and R in a given distance behind the optical elements is computed using the following relations:

```
W2z(zz) = sqrt((W2)^2*(1+zn/R2)^2+(2*zn/(k*W2))^2); R2z(zz) = ((R2+zn)^2*
(k*W2^2)^2+4*zn^2*R2^2)/((R2+zn)*(k*W2^2)^2+4*zn*R2^2);
```

Evaluated parameters W and R are shown for single equiphase surfaces depending on the distance from the optical element.

7.1 Gaussian beam

Java applet

7.1 Gaussian beam

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

Paraxial waves are characteristic by …

Possible answers for question 1:

○ … the propagation exactly in parallel to the optical axis.

○ … a small angle between the optical axis and the normal to the wave form.

○ … are sensitive to frequencies and insensitive to polarizations of waves.

◉ I won't answer this question

Question 2

The Gaussian beam is called Gaussian …

Possible answers for question 2:

○ … according to its inventor Gauss.

○ … because the intensity follow the Gaussian function in the direction of propagation.

○ … because the transversal intensity distribution of the basic mode is described by the Gaussian function.

◉ I won't answer this question

Question 3

In order to separate the real and the imaginary part of the complex amplitude …

Possible answers for question 3:

○ … the real part of $1/q(z)$ is expressed as a function of the wave form radius, and the imaginary one as a function of the beam width.

○ … the real part of $1/q(z)$ is expressed as a function of the beam width, and the imaginary one as a function of the wave form radius.

○ … the method of the separation of variables has to be applied.

◉ I won't answer this question

Question 4

The Gaussian beam is uniquely defined by …

Possible answers for question 4:

○ … the intensity, the power and the radius of the beam.

○ … the intensity, the polarization and the radius of the beam.

○ … the power and the polarization of the beam.

◉ I won't answer this question

Question 5

In the neck, the Gaussian beam …

Possible answers for question 5:

○ … exhibits the minimum width and the equiphase surface is not uniquely defined.

○ … is of the minimum power.

○ … exhibits the minimum width and the planar equiphase surface.

● I won't answer this question

see result

7.2 Passage of Gaussian beam through optic elements

Basic theory

In this paragraph, influence of various optical elements to the parameters of passing Gaussian beam is studied. If Gaussian beam propagates through a circularly symmetric system, which consists of a sequence of optical elements, then the nature of a beam stays Gaussian, and only its parameter change. This change of parameters can be simply computed by matrix optics.

Parameters of Gaussian beam can be observed depending on the position of the observation point (a coordinate on the optical axis) and on the angle between the position vector of the observation point and the optical axis. In paraxial approximation, the position and the angle are mutually associated by two algebraic equations. The optical system is therefore described by the matrix of size 2×2. This matrix is called the matrix of beam. In general:

$$
\begin{aligned}
y_2 &= A y_1 + B \Theta_1 \\
\Theta_2 &= C y_1 + D \Theta_1
\end{aligned}
\qquad
\begin{bmatrix} y_2 \\ \Theta_2 \end{bmatrix} =
\begin{bmatrix} A & B \\ C & D \end{bmatrix}
\begin{bmatrix} y_1 \\ \Theta_1 \end{bmatrix}.
\tag{7.2A.1}
$$

Here, y_1 is position of the input of an optical element, y_2 is position of the output, θ_1 is an angle at the input of the optical element, and θ_2 is an angle at the output (with respect to the optical axis).

ABCD law

Let us denote parameters of Gaussian beam at the input plane of the optical element as q_1, and parameters of Gaussian beam at the output plane of the optical element as q_2. The optical element is described by the matrix $[A, B; C, D]$. We can show that all the introduced quantities are related by the equation

$$
q_2 = \frac{A q_1 + B}{C q_1 + D}.
\tag{7.2A.2}
$$

Since parameters q determine the half-width of the Gaussian beam W and its curvature radius R, eqn. (7.2A.2), which is called ABCD law, describes the transform of the Gaussian beam by an arbitrary paraxial optical system.

Transmission matrices of simple optical elements

In this paragraph, matrices of the most frequently used optical elements are given.

Propagation in vacuum

A distance d in free space can be considered as the simplest optical element. Since wave propagates along beams, coordinates of the beam, which passes through the distance d, change according to the equation $y_2 = y_1 + \theta_1 d$ and $\theta_2 = \theta_1$. The transmission matrix **M** is therefore
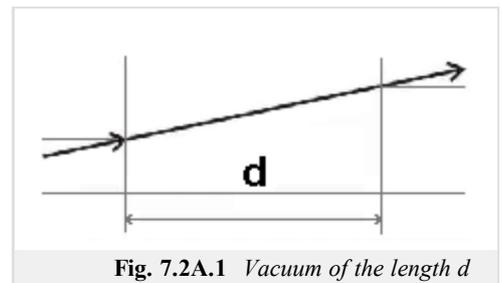

Fig. 7.2A.1 *Vacuum of the length d*

$$
M = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}.
\tag{7.2A.3}
$$

Refraction on planar boundary

On planar boundary of two media of refraction indexes n_1 and n_2, angles of the beam change according to the Snell law
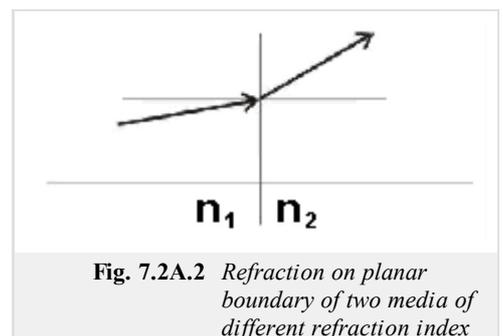

Fig. 7.2A.2 *Refraction on planar boundary of two media of different refraction index*

$$n_1 \sin(\Theta_1) = n_2 \sin(\Theta_2). \qquad (\,7.2A.4\,)$$

In paraxial approximation, we have $n_1\theta_1 \approx n_2\theta_2$, and therefore, position of the beam stays unchanged, i.e. $y_2 = y_1$. The transmission matrix **M** is therefore

$$M = \begin{bmatrix} 1 & 0 \\ 0 & n_1/n_2 \end{bmatrix}. \qquad (\,7.2A.5\,)$$

Refraction on spherical boundary

Relation between angles θ_1 and θ_2 for paraxial beams refracting on a spherical boundary of two media is given by


convex boundary: R>0
concave boundary: R<0

Fig. 7.2A.3 *Refraction on spherical boundary of two media of different refraction index*
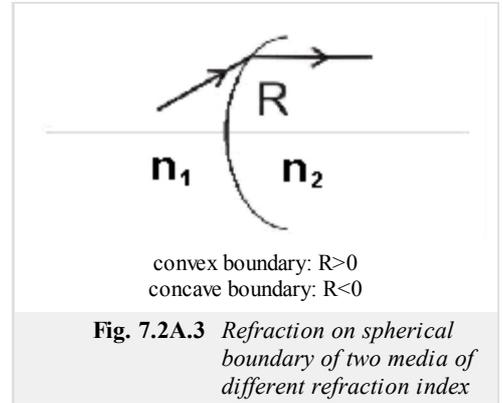
$$\Theta_2 \approx \frac{n_1}{n_2}\Theta_1 - \frac{n_2-n_1}{n_2 R}\, y. \qquad (\,7.2A.6\,)$$

The distance of the beam from the axis stays unchanged, i.e. $y_2 \approx y_1$. The transmission matrix **M** is therefore:

$$M = \begin{bmatrix} 1 & 0 \\ -\dfrac{(n_2-n_1)}{n_2 R} & \dfrac{n_1}{n_2} \end{bmatrix}. \qquad (\,7.2A.7\,)$$

Passage through a thin lens

Relation between angles θ_1 and θ_2 for paraxial beams, which pass through a thin lens of the focus distance *f*, is:


convex lens: f>0
concave lens: f<0

Fig. 7.2A.4 *Passage through a lens of focus distance f*
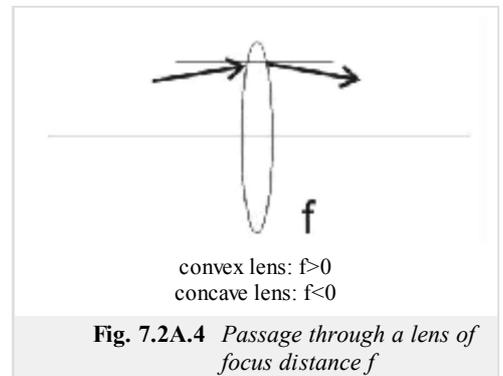
$$\Theta_2 = \Theta_1 - \frac{y}{f}. \qquad (\,7.2A.8\,)$$

The distance form the axis stays unchanged. The transmission matrix **M** is therefore:

$$M = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix}. \qquad (\,7.2A.9\,)$$

Reflection from planar mirror

Reflecting beam from a planar mirror, the position of a beam stays unchanged ($y_2 = y_1$). For angles, we get $\theta_2 = \theta_1$. Hence, the transmission matrix is unitary
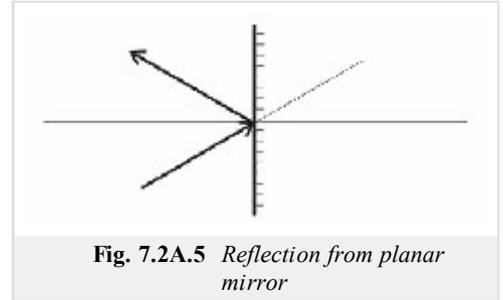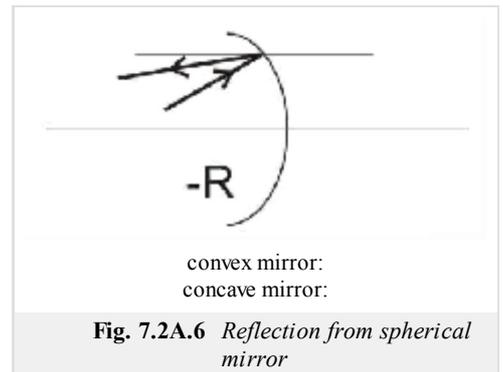

Fig. 7.2A.5 *Reflection from planar mirror*

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$ (7.2A.10)

Reflection from spherical mirror

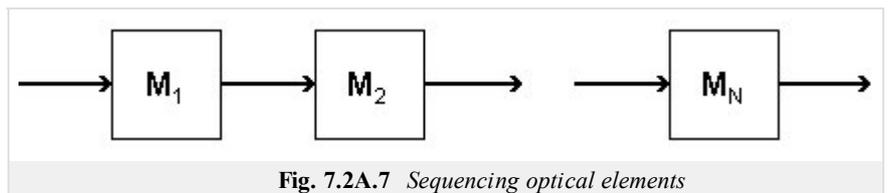Exploiting (7.2A.9), we obtain:

convex mirror:
concave mirror:

Fig. 7.2A.6 *Reflection from spherical mirror*

$$\left(-\Theta_2\right) + \Theta_1 \approx \frac{2y}{-R},$$ (7.2A.11)

$$M = \begin{bmatrix} 1 & 0 \\ 2/R & 1 \end{bmatrix}.$$ (7.2A.12)

Transmission matrix of the sequence of optical elements

Sequence of optical elements, which are described by matrices $\mathbf{M}_1$, $\mathbf{M}_2$, ..., $\mathbf{M}_N$, is equivalent to a single optical element of the transmission matrix:

Fig. 7.2A.7 *Sequencing optical elements*

$$\mathbf{M} = \mathbf{M}_1 \cdot \mathbf{M}_2 \cdot ... \cdot \mathbf{M}_N$$ (7.2A.13)

Note the order of the matrix multiplication. The matrix of the element, which is entered by the beam first, appears on the right, and therefore, it multiplies the column vector of the incident beam first.

More information can be found in **[16]**.

7.2 Passage of Gaussian beam through optic elements

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**5** of **5**:

Question 1

The beam matrix of an optical element …

Possible answers for question 1:

- ○ … associates the equiphase surface radius and the beam width at the input to the waveform radius and the beam width at the output.
- ○ … associates the input position and the input propagation angle to the output position and the output propagation angle.
- ○ … enables to compute the output power from the input one.
- ◉ I won't answer this question

Question 2

The beam matrix of the planar mirror …

Possible answers for question 2:

- ○ … is non-diagonal.
- ○ … is complex.
- ○ … is unitary.
- ◉ I won't answer this question

Question 3

The beam matrix of the free space …

Possible answers for question 3:

- ○ … contains units in the diagonal and the distance above the diagonal.
- ○ … is unitary.
- ○ … is not diagonal.
- ◉ I won't answer this question

Question 4

In case the optical system consists of several elements, the resultant beam matrix …

Possible answers for question 4:

- ○ … is a product of beam matrices of elements.
- ○ … is a sum of beam matrices of elements.
- ○ … cannot be simply expressed.
- ◉ I won't answer this question

Question 5

The focal distance of the convex lens …

see result

Chapter 8: Analysis and simulation of semiconductor structures

Introduction

This chapter deals with semiconductor structures whose the mathematical description of the electromagnetic field is so different that you can not use any of the methods or computing environment presented in this textbook. We use multi-physical computing environment COMSOL Multiphysic, which is based on finite element method.

In the first chapter we present the macroscopic description of semiconductor structures called drift and diffusion model.

8.1 Drift diffusion model

Basic theory

1   Basic Semiconductor Equations

The device modeling represents the problem of charge transport in macroscopic scale in comparison with the semiconductor lattice dimensions. In this approach we assume that the movement of electron between two collisions is described by Newton law and only the interactions with lattice are described by quantum mechanics. So, the carriers can be described statistically thought the distribution function that is given by solving of Boltzman transport equation. However, this is a difficult task to accomplish, thus through several idealistic simplification of Boltzman equation we obtain the practical system of equations (8.1A.1) called the drift-diffusion model. These equations are also called the basic semiconductor equations. They can be easily deduced from Maxwells equations.

$$\nabla \cdot \left(\nabla \Psi \right) = \frac{q}{\varepsilon}(n - p - C),$$ (8.1A.1a)

$$\nabla \cdot (\mathbf{J}_n) - q \frac{\partial n}{\partial t} = qR,$$ (8.1A.1b)

$$\nabla \cdot (\mathbf{J}_p) + q \frac{\partial p}{\partial t} = -qR,$$ (8.1A.1c)

$$\mathbf{J}_n = qn\mu_n \mathbf{E} + qD_n \nabla n,$$ (8.1A.1d)

$$\mathbf{J}_p = qp\mu_p \mathbf{E} - qD_p \nabla p.$$ (8.1A.1e)

Equation (8.1A.1a) is a Poisson equation for electric potential Ψ, where the charge density is defined by elementary charge q and the concentration of negatively charged electrons n, the concentration of positively charged holes p and fixed charge concentration C. ε in this equation denotes permittivity of used semiconductor. Values of relative permittivity for the most common materials are listed in tab. **8.1A.1**.

Continuity equation (8.1A.1b) and (8.1A.1c) express that the sources and sinks of current density are fully compensated by the time change of free charge and joint by function R, which expresses the rate of generation and recombination of electrons and holes. $\mathbf{J}_n$ denotes current density caused by electrons, $\mathbf{J}_p$ current density caused holes and t is the time. Expression of the rate R requires a good knowledge of the physical mechanisms of generation and recombination in semiconductor materials and is one of the key parameters for obtaining the relevant results of simulations of semiconductor structures.

Tab. 8.1A.1 *Typical values of relative permittivity [44]*.

Materiál	ε_r []
Si	11,7
SiO$_2$	3,9
Si$_3$N$_4$	7,2
GaAs	12,5
Ge	16,1

In both current densities equations (8.1A.1d) and (8.1A.1e) the first term on the right side is a component of current density caused by the Lorenz force, taking into account only the influence of the electric field $\mathbf{E}$ (the effect of magnetic induction is neglected), and establishes the effective mobility of electrons μ_n and holes μ_p Their value is generally defined by empirical relationships. Mobility of charge carriers is a quantity which has a large influence on the properties of semiconductor components and therefore its correct modeling is very important. The second term on the right side of equations (8.1A.1d) and (8.1A.1e) represents the effect of carrier diffusion in the direction of the gradient of concentration and introduces Einstein diffusion constants

$$D_n = \mu_n \frac{kT}{q},$$ (8.1A.2a)

$$D_p = \mu_p \frac{kT}{q},$$ (8.1A.2b)

where k is the Boltzmann constant and T is the temperature.

2   Physical Parameters

The equations (8.1A.1) constitute the basic equation system for analyzing and simulation of most semiconductor devices. However, there may be cases when become important physical phenomena that are not sufficiently covered by these equations. That is the case of tunneling currents through very thin layers, which have a significant effect on the properties of the devices. Usually are used the empirical corrections than more complex system of equations to describe such phenomena. Physical parameters of these equations define the geometry of the device, kind of semiconductor material and used manufacturing technology. Their modeling will be discussed in the following paragraphs.

Doping Profile

Distribution of impurities in the semiconductor volume, resulting in N type or P type semiconductor, determines the geometry and function of a semiconductor device. It is therefore essential input information to achieve a accurate results of simulations, so the manufacturing processes such as ion implantation, diffusion, thermal oxidation, epitaxial accretion and others should be properly modeled. The scope of following

paragraphs will be restricted only to processes of ion implantation and diffusion of impurities in volume of semiconductor substrate. The resulting profiles of these processes are often modeled by Gaussian function. An example of doping profile of P-N diode is shown in fig. **8.1A.1**.
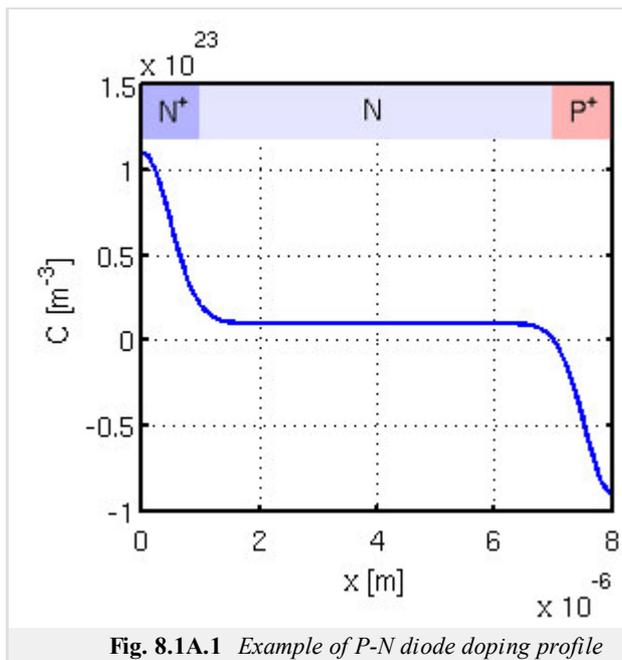

Fig. 8.1A.1 *Example of P-N diode doping profile*

Carriers Mobility

The movement of carriers in semiconductors is disturbed by collisions with the crystal lattice, impurities and defects, therefore, track of the carriers movement between two points is not a straight line. In order of easily imaginable description this phenomenon was introduced the mobility, which is modeled with respect to the mechanisms that affect the carriers collision. These mechanisms may be caused by

- atomic lattice material
- ionized impurities
- carrier-carrier collisions
- neutral impurities
- intensity of the electric field

Often can be used the value $\mu_{n,p}^{300}$ specified for the given material at a temperature of 300K presented in tab. **8.1A.2**.

materiál	μ_n^{300} [m²V⁻¹s⁻¹]	μ_p^{300} [m²V⁻¹s⁻¹]
Si	0,14	0,04
GaAs	0,8	0,04
Ge	0,38	0,18

Tab. 8.1A.2 *Typical values of carriers mobility at temperature 300K [44].*

Generation and Recombination of Carriers

In the previous paragraphs was mentioned function R, which describes the rate of generation and recombination of electrons and holes. The dominant mechanism is Shockley-Read-Hall recombination and generation, which is modeled as a trap between valence and conductive band of semiconductor. The resulting function has the following form

$$R^{SRH} = \frac{np - n_i^2}{\tau_p(n + n_i) + \tau_n(p + p_i)}, \tag{8.1A.3}$$

where, τ_n and τ_p are carriers lifetime and n_i is the intrinsic concentration.

3   Analysis of Basic Equations and Boundary Conditions

In the first section was given a set of basic equations (8.1A.1). It is important to note that in the current densities in the relations (8.1A.1d) and (8.1A.1e) have been omitted the currents caused by a narrowing of the forbidden zone and temperature gradient, because their effect is

considered to be negligible. However, the (8.1A.13) may become incorrect if any of these phenomenon become significant. Using (8.1A.1a) and by substituting the current density equations (8.1A.1d) and (8.1A.1e) into the continuity equation (8.1A.1b) and (8.1A.1c) we obtain a system of three partial differential equations (8.1A.4) with variables Ψ, n a p.

$$\nabla \cdot \left(\nabla \psi \right) - \frac{q}{\varepsilon}(n - p - C) = 0, \qquad (8.1A.4a)$$

$$\nabla \cdot (D_n \nabla n - n\mu_n \nabla \psi) - R = \frac{\partial n}{\partial t}, \qquad (8.1A.4b)$$

$$\nabla \cdot (D_p \nabla p + p\mu_p \nabla \psi) - R = \frac{\partial p}{\partial t}. \qquad (8.1A.4c)$$

For mathematical analysis we need to know the initial estimation of Ψ, n and p and boundary conditions in analyzed domain. The bounded domain D is generally three-dimensional, like practically all semiconductor structures. However, in many cases, this domain can be considered as two-dimensional, or even one-dimensional, which make the analyzed the problem significantly easier.

Lets ∂D denotes partial boundary of domain D. It can be divided into two parts

$$\partial D = \partial D_p \cup \partial D_a \qquad (8.1A.5)$$

where ∂D_p denotes those parts of the boundaries that are real, physical boundaries, such as contacts and interfaces. ∂D_a indicates artificial interfaces that are introducing, for example, to exclusion of sub-structure on the large substrate or to the separation of neighboring devices on a common substrate.

The illustrative description of idealized 2D geometry of MOS transistor is depicted in fig. **8.1A.2**. The entire domain is represented by polygon A-B-C-D-E-F-G-H-A. Equation (8.1A.4) is valid only in the subdomain A-B-E-F-G-H-A. Insulator B-C-D-E-B can be characterized either by the Laplace equation for electrostatic potential (8.1A.6) or by assumption of absence of charge carriers (8.1A.7).

$$\nabla \cdot \nabla \left(\Psi \right) = 0 \qquad (8.1A.6)$$

$$n = p = C = 0 \qquad (8.1A.7)$$

However, the use of these equations makes it impossible to determine the current passing through the base and the influence of the charge on insulating oxide layer. The boundaries A-B, E-F, C-D and B-E can be considered as physical boundaries representing three contacts and interface between the semiconductor and insulator. Boundaries A-H, B-C, D-E, F-G and G-H are considered as artificial boundaries.

Physical boundaries can be divided into three categories.

$$\partial D_P = \partial D_O \cup \partial D_S \cup \partial D_I, \qquad (8.1A.8)$$

where ∂D_O indicate the ohmic contact, ∂D_S denotes Schottky contact and ∂D_I is the insulator interface.

Summary of boundary conditions for all three dependent variables of the system (8.1A.4) is given in the following tab. **8.1A.3**, where **n** denotes normal vector perpendicular to ∂D and $u_{n,p}$ denotes the speed of thermal recombination.
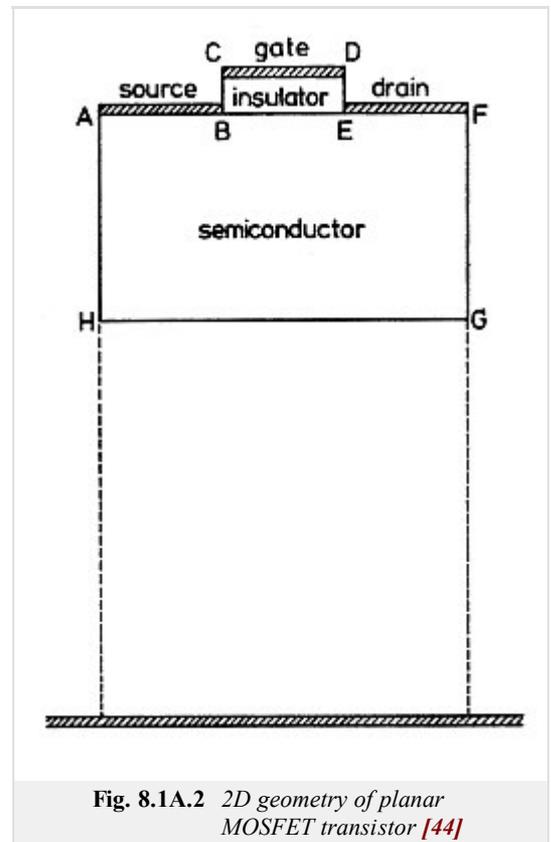

Fig. 8.1A.2 *2D geometry of planar MOSFET transistor [44]*

	ohmic contact	Shottky contact	semiconductor - insulator	artifical interface
Tab. 8.1A.3 *Boundary conditions typical semiconductor interface structures.*				
electric potential	$\psi_b = \frac{kT}{q} ar \sinh\left(\frac{C}{2n_i}\right)$	$\psi_b = \frac{kT}{q} ar \sinh\left(\frac{C}{2n_i}\right)$	$\frac{\partial \psi}{\partial \mathbf{n}} = 0$	$\frac{\partial \psi}{\partial \mathbf{n}} = 0$
electron concentration	$n = \frac{\sqrt{C^2 + 4n_i^2} + C}{2}$	$\mathbf{J}_n \mathbf{n} = -qu_n\left(n - \frac{\sqrt{C^2 + 4n_i^2} + C}{2}\right)$	$\mathbf{J}_n \mathbf{n} = 0$	$\frac{\partial n}{\partial \mathbf{n}} = 0$
hole concentration	$p = \frac{\sqrt{C^2 + 4n_i^2} - C}{2}$	$\mathbf{J}_p \mathbf{n} = -qu_p\left(p - \frac{\sqrt{C^2 + 4n_i^2} - C}{2}\right)$	$\mathbf{J}_p \mathbf{n} = 0$	$\frac{\partial p}{\partial \mathbf{n}} = 0$

8.1 Drift diffusion model

Advanced theory

1   Basic Semiconductor Equations

The fundamentals of semiconductor physic are well described by tools of quantum mechanic. This point of view gives us a model of particle behavior at atomistic level where electrons are described by wave functions whose frequency and wavelength are related to the particle energy and momentum. Electrons in semiconductor are well described by superposition of different wave functions called wave packet. The wave packet energy and momentum are the solution of the Schrödinger equation, the heart of quantum mechanic. So the motion of the particle is described by a single wave packet subject to the external forces and to the interaction with the crystal lattice.

The device modeling represents the problem of charge transport in macroscopic scale in comparison with the semiconductor lattice dimensions. In this approach we assume that the movement of electron between two collisions is described by Newton law and only the interactions with lattice are described by quantum mechanics. So, the carriers can be described statistically thought the distribution function that is given by solving of Boltzmann transport equation. However, this is a difficult task to accomplish, thus through several idealistic simplification of Boltzmann equation we obtain the practical system of equations called the drift-diffusion model. These equations are also called the basic semiconductor equations. They can be easily deduced from Maxwell's equations (8.1B.1) as will be shown in following paragraphs.

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \qquad (\ 8.1B.1a \ )$$

$$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \qquad (\ 8.1B.1b \ )$$

$$\nabla \cdot \mathbf{D} = \rho \qquad (\ 8.1B.1c \ )$$

$$\nabla \cdot \mathbf{B} = 0 \qquad (\ 8.1B.1d \ )$$

In those equations $\mathbf{E}$ denotes electric field and $\mathbf{H}$ magnetic field, $\mathbf{D}$ is electric induction and $\mathbf{B}$ magnetic induction. $\mathbf{J}$ is the current density, ρ is the space charge density and t is the time.

To complete the description we have to introduce the material relations

$$\mathbf{D} = \varepsilon \, \mathbf{E}, \qquad (\ 8.1B.2a \ )$$

$$\mathbf{B} = \mu \, \mathbf{H}, \qquad (\ 8.1B.2b \ )$$

where ε denotes permittivity and μ permeability of medium. Generally these are second order tensors (anisotropic character), but the common materials used in semiconductor devices can be considered as scalars (isotropic character). Values of relative permittivity for the most common materials are listed in tab. **8.1B.1**. Relations (8.1B.2) are valid only in semiconductor materials with time-invariant permittivity and permeability and without considering of the mechanical strain due to polarization forces. For this reason, are relations (8.1B.2) are insufficient to describe the piezoelectric and ferroelectric phenomena.

Material	ε_r []
Si	11,7
SiO$_2$	3,9
Si$_3$N$_4$	7,2
GaAs	12,5
Ge	16,1

1.1   Poisson Equation

The third Maxwell's equation (8.1B.1c) is in the form of Poisson partial differential equation. By using of relations (8.1B.2a) and (8.1B.3) can be formulated for eletric potential Y (8.1B.4)

$$\mathbf{E} = - \nabla \, \Psi, \qquad (\ 8.1B.3 \ )$$

$$\nabla \cdot \left(\varepsilon \nabla \, \Psi \right) = - \rho. \qquad (\ 8.1B.4 \ )$$

To use equation (8.1B.4) in analysis of semiconductor devices we define the space charge density ρ as a product of elementary charge q and the concentration of charge carriers, ie, the sum of the concentrations of positively charged holes p and the concentration of negatively charged electrons n, which are compensated by a fixed charge concentration of impurities C (donors or acceptor) in the lattice of semiconductor.

$$\rho = q(p - n + C) \qquad (\ 8.1B.5 \ )$$

where $C = N_D - N_A$ corresponds to the concentration of impurities defined as sum of the concentration of ionized donors N_D and ionized acceptors N_A, including signs of compensated charge.

Now we can combine (8.1B.4) and (8.1B.5) to obtain the final form of Poisson equation (8.1B.6) suitable for modeling the electrostatic field in semiconductor structures.

$$\nabla \cdot \left(\nabla \, \Psi \right) = \frac{q}{\varepsilon}(n - p - C) \qquad (\ 8.1B.6 \ )$$

1.2   Continuity Equation

The continuity equation can be derived directly from the first Maxwell equation (8.1B.1a). Applying the divergence operator to this equation and considering that the divergence of the rotation vector is zero yields

$$\nabla \cdot \nabla \times \mathbf{H} = \nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0. \qquad (\ 8.1B.7 \ )$$

If the vector of the current density $\mathbf{J}$ is divided to the sum of its components due to holes $\mathbf{J}_p$ and electrons $\mathbf{J}_n$

$$\mathbf{J} = \mathbf{J}_p + \mathbf{J}_n \qquad (\ 8.1B.8 \ )$$

and considering that the fixed charge in the semiconductor crystal lattice is time-invariant

$$\frac{\partial C}{\partial t} = 0,$$ (8.1B.9)

yields

$$\nabla \cdot (\mathbf{J}_p + \mathbf{J}_n) + q \frac{\partial}{\partial t}(p - n) = 0.$$ (8.1B.10)

This equation expresses that the current density sources or sinks are fully compensated by the time change of space charge. In order to obtain the continuity equation separately for electrons and holes, we need to establish a common function R, which expresses the rate of generation and recombination of electrons and holes. Then we get

$$\nabla \cdot (\mathbf{J}_n) - q \frac{\partial n}{\partial t} = qR,$$ (8.1B.11a)

$$\nabla \cdot (\mathbf{J}_p) + q \frac{\partial p}{\partial t} = -qR.$$ (8.1B.11b)

The formulation of the rate R requires a good knowledge of the physical mechanisms of generation and recombination in semiconductor materials and is one of the key parameters for obtaining the relevant results of simulations of semiconductor structures.

1.3 Kinetic Equations of Charge Carriers

In the previous chapter we introduced the vector current density divided into two components (8.1B.8). Current density of charged particles can be generally expressed as the product of a charge of single particle, particles concentration and average velocity (drift velocity) of particles. Then the current density of electrons and holes can be writen as

$$\mathbf{J}_n = qn\mathbf{v}_n,$$ (8.1B.12a)

$$\mathbf{J}_p = qp\mathbf{v}_p,$$ (8.1B.12b)

where $\mathbf{v}_n$ is the vector of electrons drift velocity and $\mathbf{v}_p$ id the holes drift velocity. The main problem is relating the drift velocity to forces acting on the carriers. By number of simplifications and idealistic assumptions [44] the drift velocity can be considered as result of Lorenz force and diffusion force. Then we can write

$$\mathbf{J}_n = qn\mu_n\mathbf{E} + qD_n\nabla n,$$ (8.1B.13a)

$$\mathbf{J}_p = qp\mu_p\mathbf{E} - qD_p\nabla p.$$ (8.1B.13b)

In both equations, the first term on the right side is a component of current density caused by the Lorenz force, taking into account only the influence of electric field (magnetic induction effect is neglected), and establishes the effective mobility of electrons μ_n and holes μ_p. Mobility of charge carriers is a quantity which has a large influence on the properties of semiconductor components and therefore its correct formulation is very important. The second term on the right side of equations (8.1B.13) represents the effect of carrier diffusion in the direction of the gradient of carrier concentration and introduces Einstein diffusion constants

$$D_n = \mu_n \frac{kT}{q},$$ (8.1B.14a)

$$D_p = \mu_p \frac{kT}{q},$$ (8.1B.14b)

where k is Boltzmann's constant and T is the temperature.

2 Physical Parameters

The equations (8.1B.6), (8.1B.11) and (8.1B.13) constitute the basic equation system for analyzing and simulation of most semiconductor devices. However, there may be cases when become important physical phenomena that are not sufficiently covered by these equations. That is the case of tunneling currents through very thin layers, which have a significant effect on the properties of the devices. Usually are used the empirical corrections than more complex system of equations to describe such phenomena.

Physical parameters of these equations define the geometry of the device, kind of semiconductor material and used manufacturing technology. Their modeling will be discussed in the following paragraphs.

2.1 Doping Profile

Distribution of impurities in the semiconductor volume, resulting in N type or P type semiconductor, determines the geometry and function of a semiconductor device. It is therefore essential input information to achieve a accurate results of simulations, so the manufacturing processes such as ion implantation, diffusion, thermal oxidation, epitaxial accretion and others should be properly modeled. The scope of following paragraphs will be restricted only to processes of ion implantation and diffusion of impurities in volume of semiconductor substrate. The resulting profiles of these processes are often modeled by Gaussian function

$$G(x) = ae^{-\frac{(x-b)^2}{2c^2}},$$ (8.1B.15)

where x is the spatial coordinate, a, b and c are real constants and e is the Euler number. Gaussian function is even (symmetrical according to the y-axis) and has the shape of the "bell", which represents the distribution of impurities in semiconductors as a result of the locally applied technological process. Then constant a determines the peak concentration of impurities (height of the "bell"), b the position of application of technological process (coordinate of the "bell" peak) and c is proportional to the diffusion of impurities in the substrate (width of the "bell"). In (8.1B.5), we have introduced a fixed charge concentration C, which corresponds to the distribution of impurities in semiconductors. Lets assume that N_D and N_A indicate the peak concentration of acceptors and donors (corresponding to a coefficient), then the one dimensional (1D) P-N junction problem in general can be formulated as

$$C = N_{Di} + N_D G(x_D, b_D, c_D) - N_A G(x_A, b_A, c_A),$$ (8.1B.16)

where N_{Di} is the initial substrate concentration of impurities, which is usually slightly doped by donors, so is type of N. Lets consider the P-N diode reduced to 1D problem with a total length L = 8 μm made of the silicon substrate with a low initial concentration of donors $N_{Di} = 10^{21}$ m^{-3} and the highly doped regions on opposite ends of the substrate with concentrations of donors consisting $N_D = 10^{23}$ m^{-3} and $N_A = 10^{23}$ m^{-3}. The width of the P-N transition is $L_J = 1$ μm. Resulting doping profile is shown in

fig. **8.1B.1** and is expressed by the following formula

$$C = N_{Di} + N_D e^{\left(-\frac{(x-0)^2}{2c^2}\right)} - N_A e^{\left(-\frac{(x-L)^2}{2c^2}\right)},$$

(8.1B.17)

where c is expressed by the desired width of the transition as

$$c = \frac{L_J}{\sqrt{2\ln\left(N_A/N_{Di}\right)}}.$$

(8.1B.18)

Fig. 8.1B.1 *Example of P-N diode doping profile*
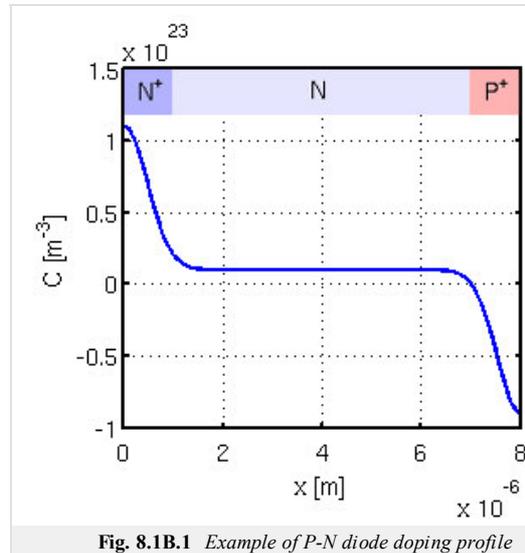
2.2   Carriers Mobility

The mobility μ is an important quantity which represents a relationship between carriers drift velocity **v** and electric field **E** by the formula

$$\mathbf{v} = \mu\,\mathbf{E}.$$

(8.1B.19)

The movement of carriers in semiconductors is disturbed by collisions with the crystal lattice, impurities and defects, therefore, track of the carriers movement between two points is not a straight line. In order of easily imaginable description this phenomenon was introduced the mobility, which is modeled with respect to the mechanisms that affect the carriers collision. These mechanisms may be caused by

- atomic lattice material
- ionized impurities
- carrier-carrier collisions
- neutral impurities
- intensity of the electric field

Often can be used the value $\mu_{n,p}^{300}$ specified for the given material at a temperature of 300K, but a correct model of solid-state structure involve detailed modeling of the carriers mobility because it captures the physical phenomena that have a major effect on the properties of semiconductor devices. In the following paragraphs we focus on the most common models of mobility.

Dependence of Carrier Mobility on Lattice Temperature

One of the fundamental mechanisms affecting the collisions are vibrations of atoms in a lattice of pure crystal semiconductors. The intensity of these vibrations is proportional to the lattice temperature. Modeling of this phenomenon by empirical relations (8.1B.20) is based on the experimentally measured values of carrier mobility at different temperatures, from which the values of the coefficients α_n and α_p are determined.

$$\mu_n^L = \mu_n^{300}\left(\frac{T}{300\text{K}}\right)^{-\alpha_n}$$

(8.1B.20a)

$$\mu_p^L = \mu_p^{300}\left(\frac{T}{300\text{K}}\right)^{-\alpha_p}$$

(8.1B.20b)

The mobility of carriers dependence on the temperature in the crystal Si and GaAs are shown in figure **8.1B.2**. Typical values of mobility and the coefficients are shown in tab. **8.1B.2**.

a) crystal Si   b) crystal GaAs

Fig. 8.1B.2 *The mobility of carriers dependence on the temperature in the semiconductor crystal*
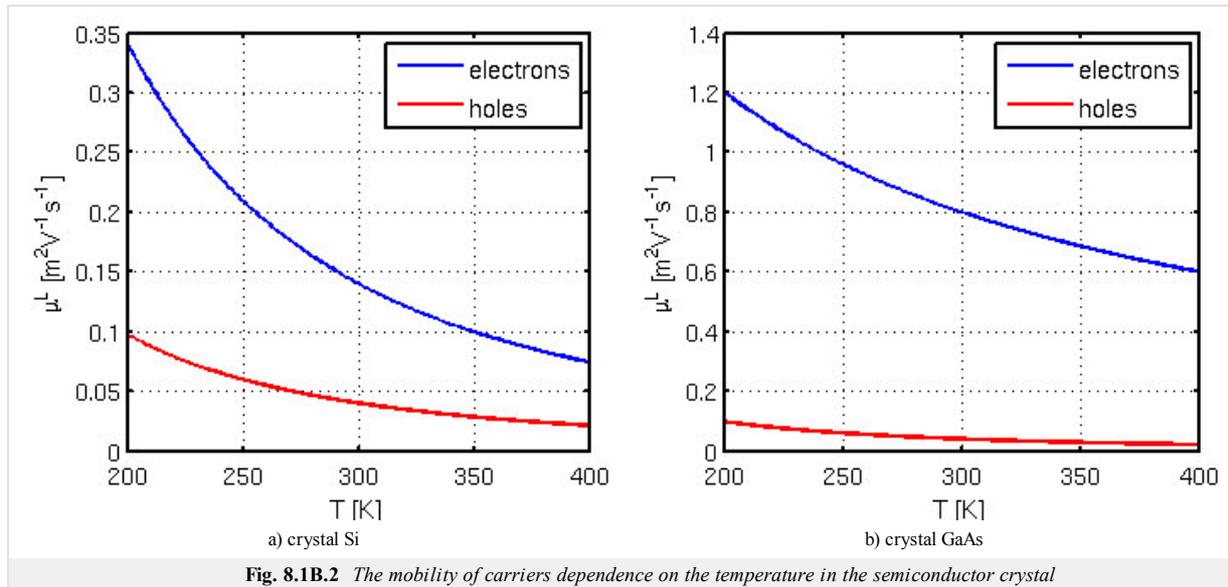
Tab. 8.1B.2 *Values of mobility and the coefficients of temperature dependence [44].*

	Si	GaAs
μ_n^{300} [m^{2}V^{-1}s^{-1}]	0,14	0,8
μ_p^{300} [m^{2}V^{-1}s^{-1}]	0,4	0,4
α_n []	2,2	1
α_p []	2,2	2,1

Dependence of Mobility on Concentration of Ionized Impurities

Ionized impurities are mainly donor and acceptor impurities. Relations describing the effect of their concentration on the mobility of carriers are defined by the means of the total concentration of ionized impurities *CI*. This is defined as the sum of a fixed charge concentration of *m* kinds of ionized impurities, multiplied by the absolute value of their normalized charge *Z*.

$$CI = \sum_{i=1}^{m} |Z_i| \cdot C_i.$$

(8.1B.21)

For the usual impurities is $|Z| = 1$. However, in some applications, such as solar cells doped with a zinc, can be double-ionized acceptors, which corresponds to the value of $|Z| = 2$. Empirical relations (8.1B.22) describe the influence of the concentration of impurities on the carriers mobility in the volume of Si, while the influence of lattice temperature is also included. Figure **8.1B.3** shows the plot of those relations at temperatures 300K and 400K.

$$\mu_n^{LI} = 0.0088\left(\frac{T}{300K}\right)^{-0,57} + \frac{0.1252\left(\frac{T}{300K}\right)^{-2.33}}{1+\frac{CI}{1.432 \cdot 10^{23}\left(\frac{T}{300K}\right)^{2.546}}}$$

(8.1B.22a)

$$\mu_p^{LI} = 0.00543\left(\frac{T}{300K}\right)^{-0,57} + \frac{0.0407\left(\frac{T}{300K}\right)^{-2.33}}{1+\frac{CI}{2.67 \cdot 10^{23}\left(\frac{T}{300K}\right)^{2.546}}}$$

(8.1B.22b)

a) electron mobility   b) hole mobility

Fig. 8.1B.3 *Influence of the impurities concentration on the carriers mobility in the volume of Si at temperatures 300K and 400K*
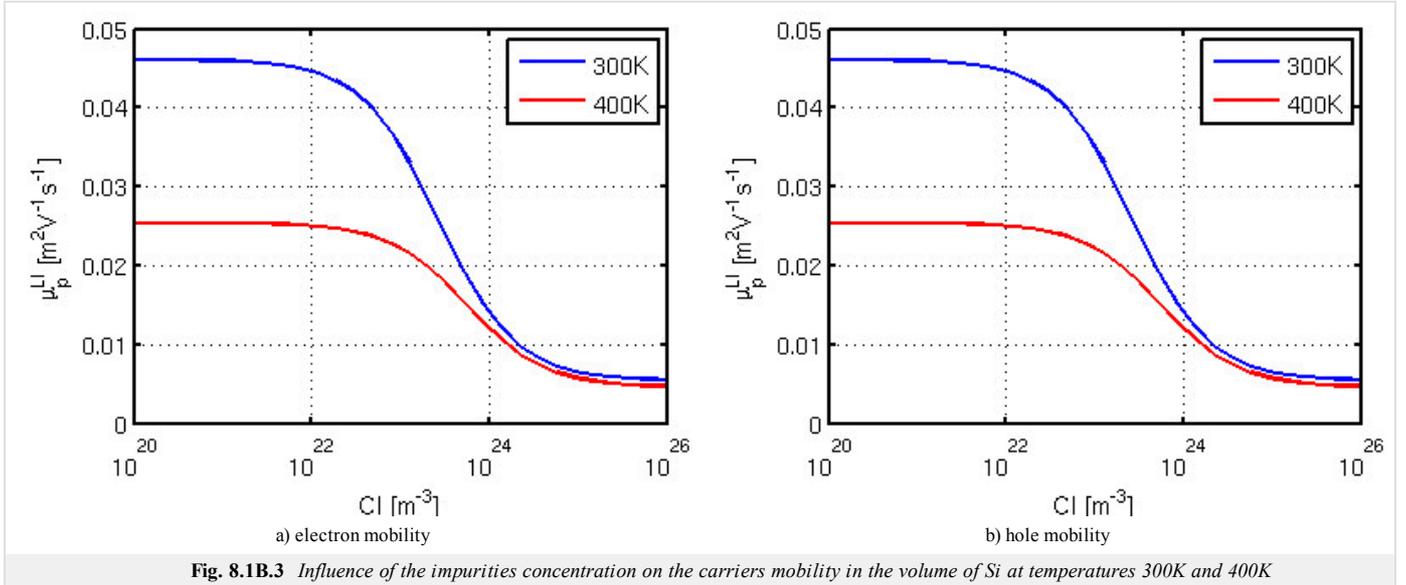
Dependence of Carrier Mobility on Electric Field Intensity

Another important phenomenon that results in a change in carrier mobility is the drift velocity saturation at high electric fields. For Si it is possible to use the following empirical formula

$$\mu_{n,p}^{LE} = \frac{\mu_{n,p}^{L}}{\left(1 + \left(\frac{E_{n,p}}{E_{n,p}^{crit}}\right)^{\beta_{n,p}}\right)^{1/\beta_{n,p}}},$$

(8.1B.23)

where $E_{n,p}$ denotes the magnitude of the electric field component in the direction of the current density vector. Threshold $E_{n,p}^{crit}$ and coefficients $\beta_{n,}$ are chosen to fit experimental data. These coefficients and their dependence on temperature are formulated as

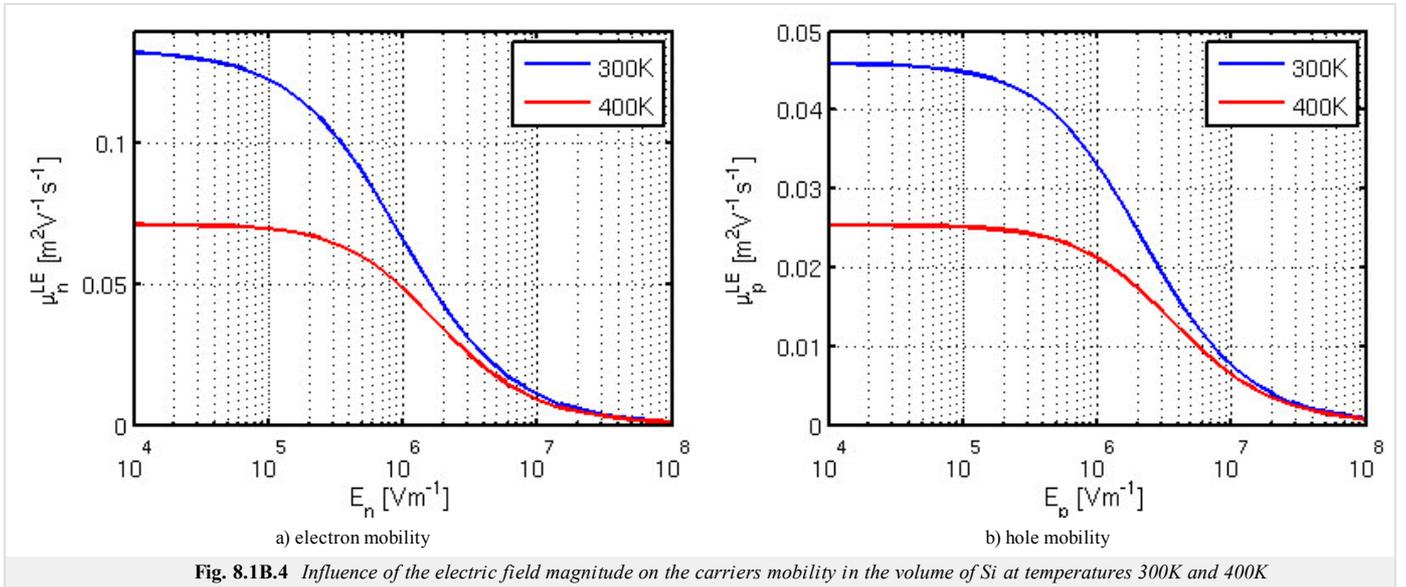$$E_n^{crit} = 857 \cdot 10^3 \left(\frac{T}{300K}\right)^{1.55},$$

(8.1B.24a)

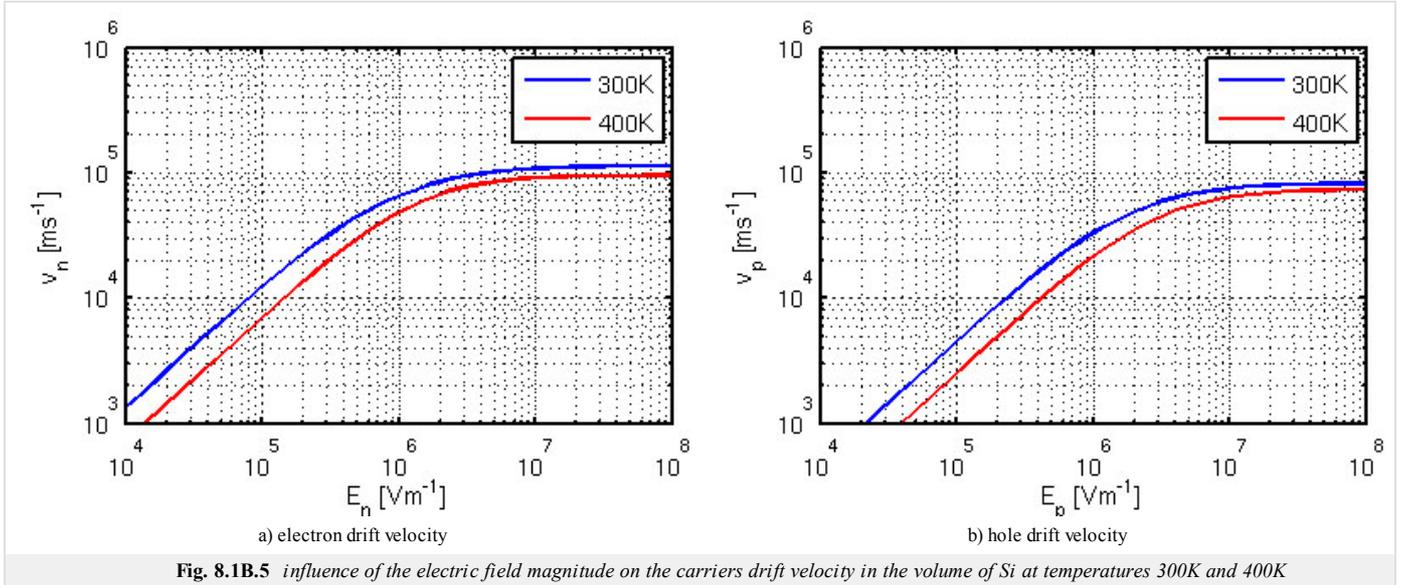$$E_p^{crit} = 1800 \cdot 10^3 \left(\frac{T}{300K}\right)^{1.68},$$

(8.1B.24b)

$$\beta_n = 1.11 \left(\frac{T}{300K}\right)^{0.66},$$

(8.1B.25a)

$$\beta_p = 1.21 \left(\frac{T}{300K}\right)^{0.17}.$$

(8.1B.25b)

Figure **8.1B.4** depicted the evaluated above relationships. Fig. **8.1B.5** shows the effect of drift velocity saturation of carriers due to the magnitude of electric field intensity which is given by

$$v_{n,p} = \mu_{n,p}^{LE} E_{n,p}.$$

(8.1B.26)

a) electron mobility   b) hole mobility

Fig. 8.1B.4 *Influence of the electric field magnitude on the carriers mobility in the volume of Si at temperatures 300K and 400K*

a) electron drift velocity

b) hole drift velocity

Fig. 8.1B.5 *influence of the electric field magnitude on the carriers drift velocity in the volume of Si at temperatures 300K and 400K*
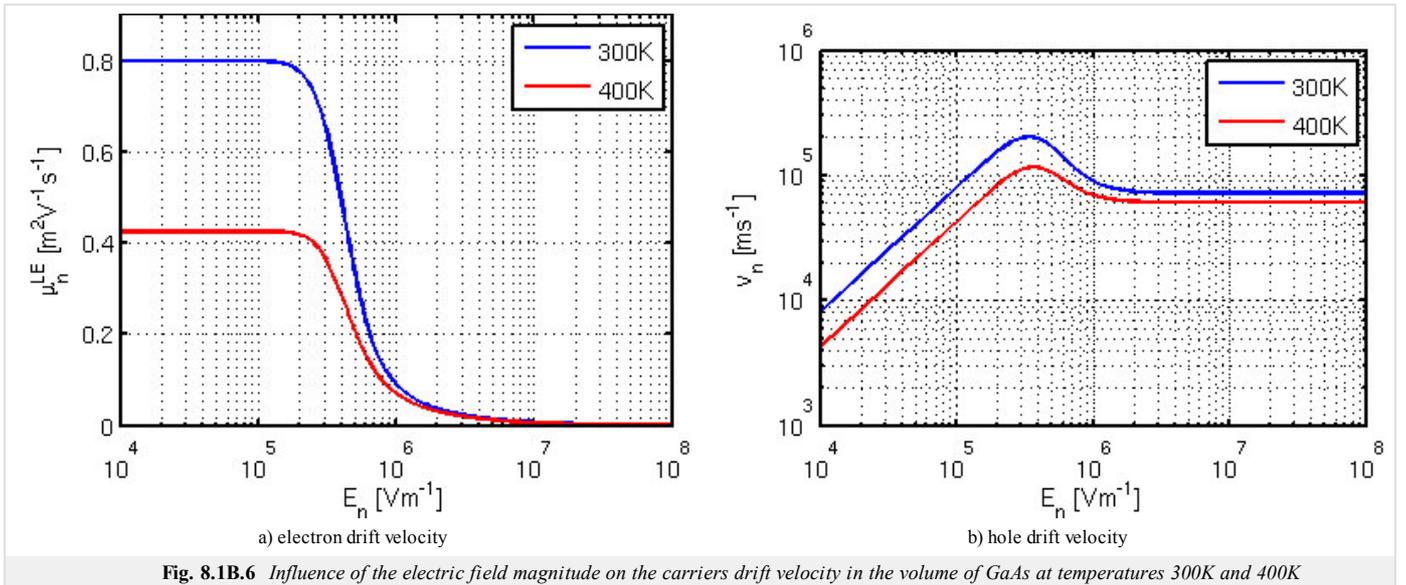
For electrons in GaAs this phenomenon can be modeled by the following formula

$$\mu_n^{LE} = \frac{\mu_n^L + v_{sat} \dfrac{(E_n)^3}{(E_n^{crit})^4}}{1 + \left(\dfrac{E_n}{E_n^{crit}}\right)^4},$$

(8.1B.27)

where E_n denotes the magnitude of the electric field component in the direction of the vector current density E_n^{crit} is its critical value and v_{sat} is a temperature dependent saturation velocity of electrons, which is modeled by

$$v_{sat} = \frac{v_{sat}^{300}}{(1-A_n)+A_n\left(\dfrac{T}{300\text{K}}\right)},$$

(8.1B.28)

where the coefficient A_n is determined experimentally. Typical values of coefficients are $E_n^{crit} = 400\times10^3$ Vm^{-1}, $A_n = 0{,}56$ a $v_{sat}^{300} = 72\times10^3$ ms^{-1}. In fig. **8.1B.6** is depicted the dependence of mobility and drift velocity of electrons on the magnitude of the electric field intensity. When we realize that by (8.1B.12) is the drift velocity proportional to the current size, we can observe the negative differential resistance as a result of Gunn's effect which is the key phenomenon of the many active microwave devices made of GaAs.

a) electron drift velocity

b) hole drift velocity

Fig. 8.1B.6 *Influence of the electric field magnitude on the carriers drift velocity in the volume of GaAs at temperatures 300K and 400K*

2.3   Generation and Recombination of Carriers

In the previous paragraphs was mentioned function R, which describes the rate of generation and recombination of electrons and holes. Consider a uniformly doped semiconductor, which is in thermal equilibrium, ie. the processes of generation and recombination of excited thermal energies are in dynamic equilibrium. For this situation we can define the equilibrium concentration of electrons n_0 and p_0 holes, which are bound by the following formula

$$n_0 p_0 = n_i^2,$$

(8.1B.29)

where n_i is the intrinsic concentration. If semiconductor is excited by external stimulus, the balance between generation and recombination processes will be unbalanced and the concentration of electrons and holes will be disturbed from their equilibrium values. In the continuity equation (8.1B.11) is this effect expressed by R function.

Depending on the physical phenomenon which is the cause of generation and recombination event of carriers, we distinguish

- R^{SRH} (Shockley-Read-Hall) transitions of phonons
- R^{OPT} photon transitions
- R^{SURF} surface phenomena
- R^{AU} triple Auger transitions
- R^{II} impact ionization.

Individual formulations can be combined by simple addition $R = R^{SRH} + R^{OPT} + R^{AU} + R^{SURF} + R^{II}$.

The dominant mechanism is Shockley-Read-Hall recombination and generation, which is modeled as a trap between valence and conductive band of semiconductor. The resulting function has the following form

$$R^{SRH} = \frac{np - n_i^2}{\tau_p(n+n_1) + \tau_n(p+p_1)},$$ (8.1B.30)

where τ_n and τ_p are carriers lifetime, n_1 and p_1 are the concentrations defined by trap energy level. If the level of trap is in the middle of the forbidden zone, concentrations are equal intrinsic, $p_1 = n_1 = n_i$.

3 Analysis of Basic Equations and Boundary Conditions

In the first section was given a set of basic equations (8.1B.6), (8.1B.11) and (8.1B.13). It is important to note that in the current densities in the relations (8.1B.13) have been omitted the currents caused by a narrowing of the forbidden zone and temperature gradient, because their effect is considered to be negligible. However, the (8.1B.13) may become incorrect if any of these phenomenon become significant.

Using (8.1B.6) and by substituting the current density equations (8.1B.13) into the continuity equation (8.1B.11) we obtain a system of three partial differential equations (8.1B.31) with variables Ψ, n and p.

$$\nabla \cdot \left(\nabla \psi \right) - \frac{q}{\varepsilon}(n - p - C) = 0$$ (8.1B.31a)

$$\nabla \cdot (D_n \nabla n - n\mu_n \nabla \psi) - R = \frac{\partial n}{\partial t}$$ (8.1B.31b)

$$\nabla \cdot (D_p \nabla p + p\mu_p \nabla \psi) - R = \frac{\partial p}{\partial t}$$ (8.1B.31c)

For mathematical analysis we need to know the initial estimation of Ψ, n and p and boundary conditions in analyzed domain. The bounded domain D is generally three-dimensional, like practically all semiconductor structures. However, in many cases, this domain can be considered as two-dimensional, or even one-dimensional, which make the analyzed the problem significantly easier.

Lets ∂D denotes partial boundary of domain D. It can be divided into two parts

$$\partial D = \partial D_p \cup \partial D_a$$ (8.1B.32)

where ∂D_p denotes those parts of the boundaries that are real, physical boundaries, such as contacts and interfaces. ∂D_a indicates artificial interfaces that are introducing, for example, to exclusion of sub-structure on the large substrate or to the separation of neighboring devices on a common substrate.

The illustrative description of idealized 2D geometry of MOS transistor is depicted in fig. **8.1B.7**. The entire domain is represented by polygon *A-B-C-D-E-F-G-H-B*. Equation (8.1B.31) is valid only in the subdomain *A-B-E-F-G-H-B*. Insulator *B-C-D-E-B* can be characterized either by the Laplace equation for electrostatic potential (8.1B.33) or by assumption of absence of charge carriers (8.1B.34).

$$\nabla \cdot \nabla \left(\Psi \right) = 0$$ (8.1B.33)

$$n = p = C = 0$$ (8.1B.34)

However, the use of these equations makes it impossible to determine the current passing through the base and the influence of the charge on insulating oxide layer.

The boundaries *A-B*, *E-F*, *C-D* and *B-E* can be considered as physical boundaries representing three contacts and interface between the semiconductor and insulator. Boundaries *A-H*, *B-C*, *D-E*, *F-G* and *G-H* are considered as artificial boundaries.

Physical boundaries can be divided into three categories.

$$\partial D_P = \partial D_O \cup \partial D_S \cup \partial D_I$$ (8.1B.34)

where ∂D_O indicate the ohmic contact, ∂D_S denotes Schottky contact and ∂D_I is the insulator interface.
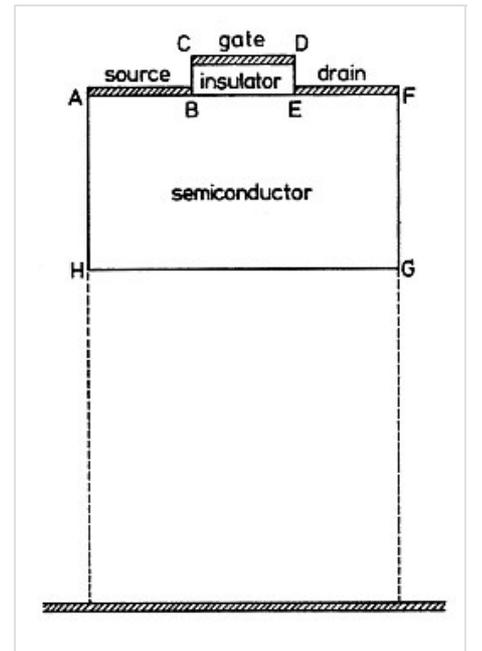

Fig. 8.1B.7 *2D geometry of planar MOSFET transistor* **[44]**

3.1 Ohmic Contact

The simplest boundary condition for the ohmic contact is the Dirichlet boundary condition for the electrostatic potential

$$\psi(t) - \psi_b - \psi_D(t) = 0, \tag{8.1B.36}$$

where $\Psi_D(t)$ denotes the external applied voltage and Ψ_b denotes a built-in potential directly related to impurities in the volume of semiconductor and can be expressed by the following formula

$$\psi_b = \frac{kT}{q} \, ar \sinh\left(\frac{C}{2n_i}\right). \tag{8.1B.37}$$

If the semiconductor is doped only by donors or acceptors, we can use the simplified relations

$$N_D > > N_A \rightarrow \psi_b \cong \frac{kT}{q} \ln\left(\frac{N_D}{2n_i}\right), \tag{8.1B.38a}$$

$$N_A > > N_D \rightarrow \psi_b \cong \frac{kT}{q} \ln\left(\frac{N_A}{2n_i}\right). \tag{8.1B.38b}$$

Furthermore, we consider boundary conditions for carriers concentration. If we consider the thermal balance (8.1B.28) and the absence of space charge on the ohmic contact, the following relations are valid.

$$np - n_i{}^2 = 0 \tag{8.1B.39}$$

$$n - p - C = 0 \tag{8.1B.40}$$

These two equations can be expressed as Dirichlet boundary conditions for electrons (8.1B.41a) and holes (8.1B.41b)

$$n = \frac{\sqrt{C^2 + 4n_i{}^2} + C}{2}, \tag{8.1B.41a}$$

$$p = \frac{\sqrt{C^2 + 4n_i{}^2} - C}{2}. \tag{8.1B.41b}$$

3.2 Schottky Contact

The physical description of Shottkyho contact is not trivial, so for the purpose of device simulation is used a very simplified model. For the electrostatic potential can write the Dirichlet boundary condition

$$\psi(t) - \psi_b - \psi_s - \psi_D(t) = 0, \tag{8.1B.42}$$

where $\Psi_D(t)$ denotes the external applied voltage and Ψ_s is the size of Schottky barrier, which is characterized by kind of metal and semiconductor in junction. Ψ_b again denotes a built-in potential. Boundary condition for the concentration of carriers is based on the current density flowing through the contact. By using the theory of thermal emission and diffusion can be derived the following boundary conditions for the reverse polarized contact

$$\mathbf{J}_n \, \mathbf{n} = -qu_n\left(n - \frac{\sqrt{C^2 + 4n_i{}^2} + C}{2}\right), \tag{8.1B.43a}$$

$$\mathbf{J}_p \, \mathbf{n} = -qu_p\left(p - \frac{\sqrt{C^2 + 4n_i{}^2} - C}{2}\right), \tag{8.1B.43b}$$

where $\mathbf{n}$ denotes the normal vector perpendicular to ∂D and $u_{n,p}$ is the speed of thermal recombination.

3.3 Semiconductor Insulator Interface

Another category consists of the interface between semiconductor and insulating material. At this interface must be valid Gauß's law

$$\varepsilon_{sem} \frac{\partial \psi}{\partial \mathbf{n}}\bigg|_{sem} - \varepsilon_{ins} \frac{\partial \psi}{\partial \mathbf{n}}\bigg|_{ins} = Q_{\text{int}}, \tag{8.1B.44}$$

where ε_{sem} and ε_{ins} indicates the permittivity of semiconductor and insulator. Q_{int} is the charge on the interface. For thin layers of insulators such as oxide, can be neglected the component of electric field in an insulator perpendicular to the interface, thus simplifying to

$$\varepsilon_{sem} \frac{\partial \psi}{\partial \mathbf{n}}\bigg|_{sem} = Q_{\text{int}}. \tag{8.1B.45}$$

By neglecting of the charge on the interface Q_{in} can be the previous condition reduced to the Neumann condition for the electrostatic potential

$$\frac{\partial \psi}{\partial \mathbf{n}} = 0. \tag{8.1B.46}$$

For carriers, the current component perpendicular to the interface must equal the rate of surface recombination R^{SURF}

$$\mathbf{J}_n \, \mathbf{n} = -qR^{SURF}, \tag{8.1B.47a}$$

$$\mathbf{J}_p \, \mathbf{n} = qR^{SURF}. \tag{8.1B.47b}$$

Very often is the surface recombination neglected, so we get simpler conditions

$$\mathbf{J}_n \, \mathbf{n} = 0, \tag{8.1B.48a}$$

$$\mathbf{J}_p \, \mathbf{n} = 0. \tag{8.1B.48b}$$

3.4 Artificial Interface

The last case is the artificial interfaces. Here we assume either the boundary condition (8.1B.49a) (8.1B.49b) and (8.1B.49c) to ensure that the domain is enclosed or specific values for the electrostatic potential and carrier concentration of the external sources.

$$\frac{\partial \psi}{\partial \mathbf{n}} = 0 \qquad\qquad\qquad (8.1B.49a)$$

$$\frac{\partial n}{\partial \mathbf{n}} = 0 \qquad\qquad\qquad (8.1B.49b)$$

$$\frac{\partial p}{\partial \mathbf{n}} = 0 \qquad\qquad\qquad (8.1B.49c)$$

8.1 Drift diffusion model

Model in Comsol

The attached file is a model P-N diodes made of silicon for simulation environment COMSOL Multiphysics in version 3.5. 1D model represents the structure of the doping profile identical to fig. **8.1B.1**. The equations (8.1B.31) and boundary conditions for ohmic contact (8.1B.37) and (8.1B.41) are implemented in the model. The SRH recombination model and table the value of mobility carriers at 300K is used.

After opening the model in COMSOL program plots the uniform mesh of the model. In the *Solve* menu, click on *Solve Problem* which run the default simulation, which is based on calculating the potential and the concentration of carriers, depending on the applied voltage on the diode, which is varied in range from 0 to 2 V.

In the *Postprocessing* menu, click on *Plot Parameters*. In the *Line* tab in menu *Predefined Quantities* select variable whose distribution along the diode we want to plot

- *psi* electric potential

- *cn* concentrations of electrons

- *cp* concentration of holes

In the *General* tab in the menu *Predefined Value* select the specific value of applied voltage, for which we want to plot the distribution. The appropriate quantity is plotted by clicking on button $\boxed{\text{Ok}}$.
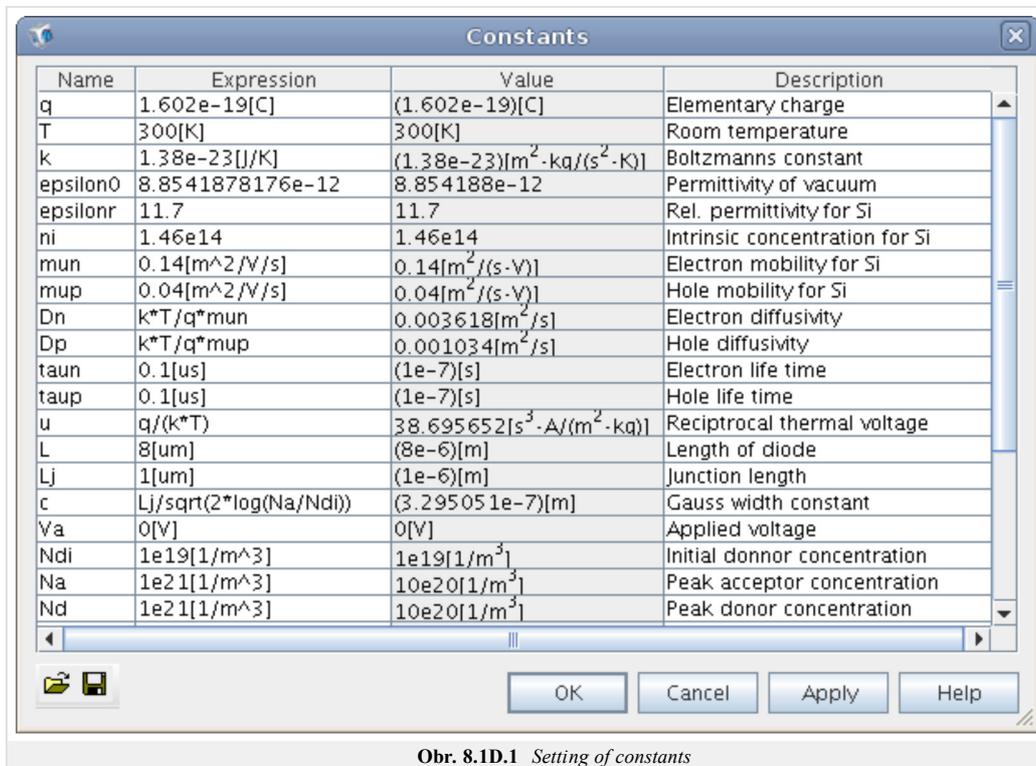
V-A characteristics is available in the menu *Postprocessing*/ *Domain Plot Variables*, where the in the *Point* tab select the *Boundary Selection* No. 1, check if there is the *Ic* term in *Expression* field, and click $\boxed{\text{Ok}}$.
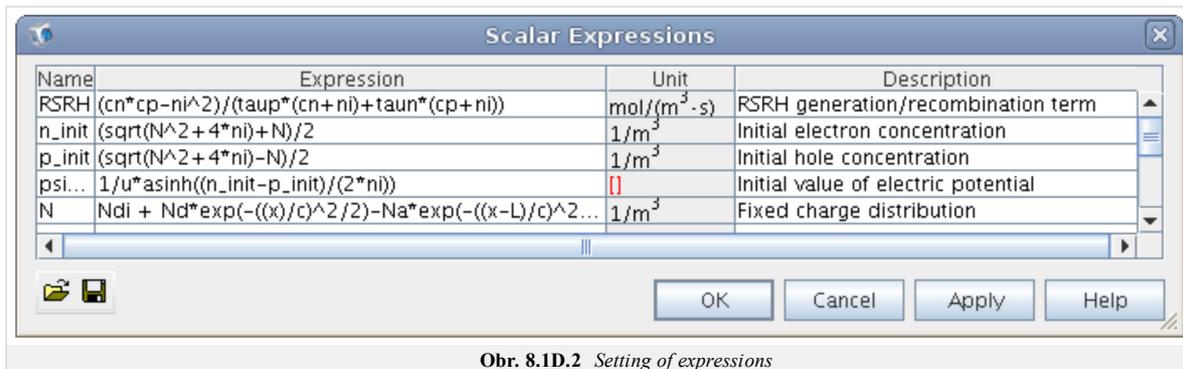
8.1 Drift diffusion model

Developing Comsol model

In the following text is summarized the implementation of a model 1D silicon diodes in the program COMSOL Multiphysics in version 3.5. The basic knowledge of working with the program is assumed, so we set out images of important dialogues that define the model. This model is a 1D structure of the doping profile of an identical as in fig. **8.1B.1**. Are implemented equation (8.1B.31) and boundary conditions for ohmic contact (8.1A.37) and (8.1A.41). SRH recombination model and table the value of mobility carriers at 300K is used. To work with this model read the instructions in the layer C.

The diode consist of a line with the coordinates x1 = 0 and x2 = L = 8 μm. Definition of the necessary constants and expressions is given in fig. **8.1D.1**, fig. **8.1D.2** and fig. **8.1D.3**.

Name	Expression	Value	Description
q	1.602e-19[C]	(1.602e-19)[C]	Elementary charge
T	300[K]	300[K]	Room temperature
k	1.38e-23[J/K]	(1.38e-23)[m^2·kg/(s^2·K)]	Boltzmanns constant
epsilon0	8.8541878176e-12	8.854188e-12	Permittivity of vacuum
epsilonr	11.7	11.7	Rel. permittivity for Si
ni	1.46e14	1.46e14	Intrinsic concentration for Si
mun	0.14[m^2/V/s]	0.14[m^2/(s·V)]	Electron mobility for Si
mup	0.04[m^2/V/s]	0.04[m^2/(s·V)]	Hole mobility for Si
Dn	k*T/q*mun	0.003618[m^2/s]	Electron diffusivity
Dp	k*T/q*mup	0.001034[m^2/s]	Hole diffusivity
taun	0.1[us]	(1e-7)[s]	Electron life time
taup	0.1[us]	(1e-7)[s]	Hole life time
u	q/(k*T)	38.695652[s^3·A/(m^2·kg)]	Reciprocal thermal voltage
L	8[um]	(8e-6)[m]	Length of diode
Lj	1[um]	(1e-6)[m]	Junction length
c	Lj/sqrt(2*log(Na/Ndi))	(3.295051e-7)[m]	Gauss width constant
Va	0[V]	0[V]	Applied voltage
Ndi	1e19[1/m^3]	1e19[1/m^3]	Initial donnor concentration
Na	1e21[1/m^3]	10e20[1/m^3]	Peak acceptor concentration
Nd	1e21[1/m^3]	10e20[1/m^3]	Peak donor concentration

Obr. 8.1D.1 *Setting of constants*

| Name	Expression	Unit	Description
RSRH	(cn*cp-ni^2)/(taup*(cn+ni)+taun*(cp+ni))	mol/(m^3·s)	RSRH generation/recombination term
n_init	(sqrt(N^2+4*ni)+N)/2	1/m^3	Initial electron concentration
p_init	(sqrt(N^2+4*ni)-N)/2	1/m^3	Initial hole concentration
psi...	1/u*asinh((n_init-p_init)/(2*ni))	[]	Initial value of electric potential
N	Ndi + Nd*exp(-((x)/c)^2/2)-Na*exp(-((x-L)/c)^2...	1/m^3	Fixed charge distribution

Obr. 8.1D.2 *Setting of expressions*
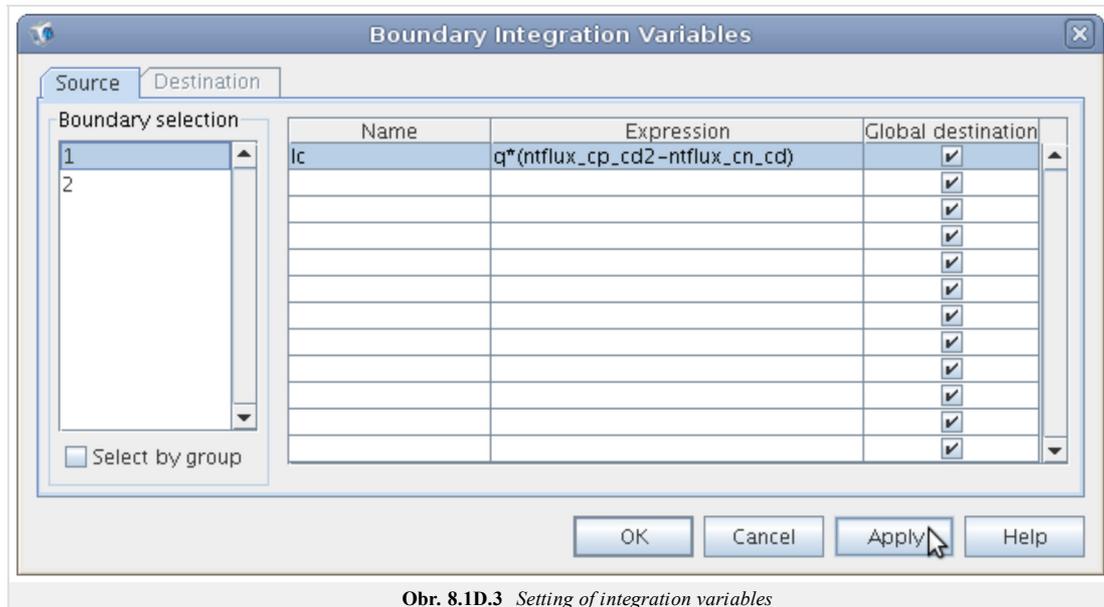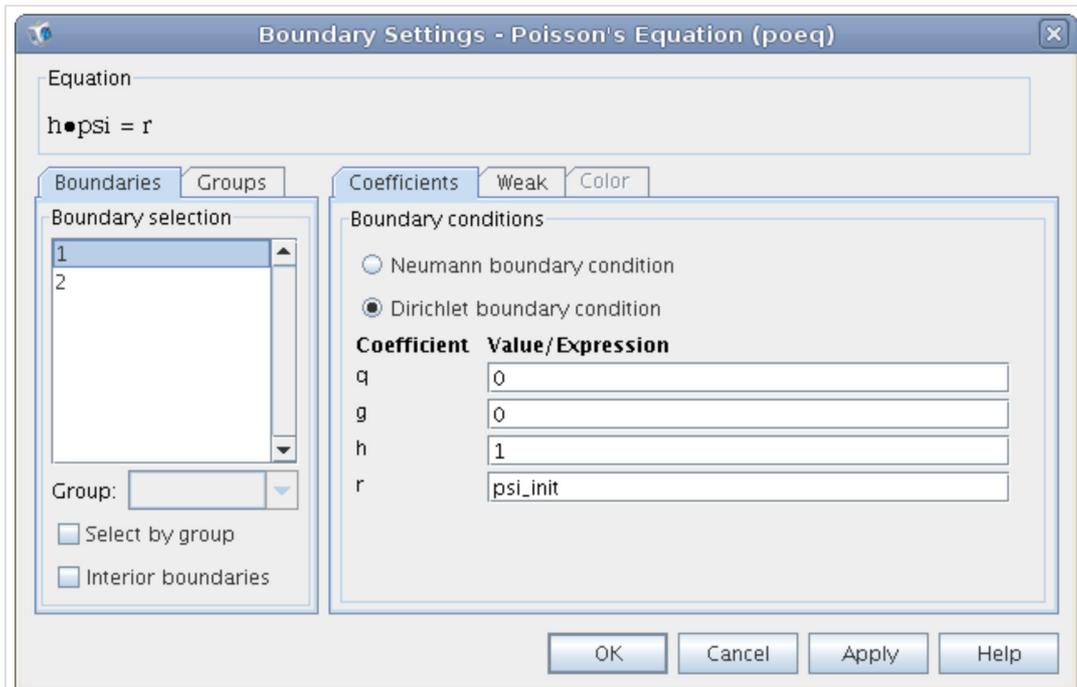
Obr. 8.1D.3 *Setting of integration variables*

Module Poisson equation

In this section is described the electrostatic module. In fig. **8.1D.4** is depicted the settings of preset formula over the diode domain, where the tab *Init* is set to *psi_init*.
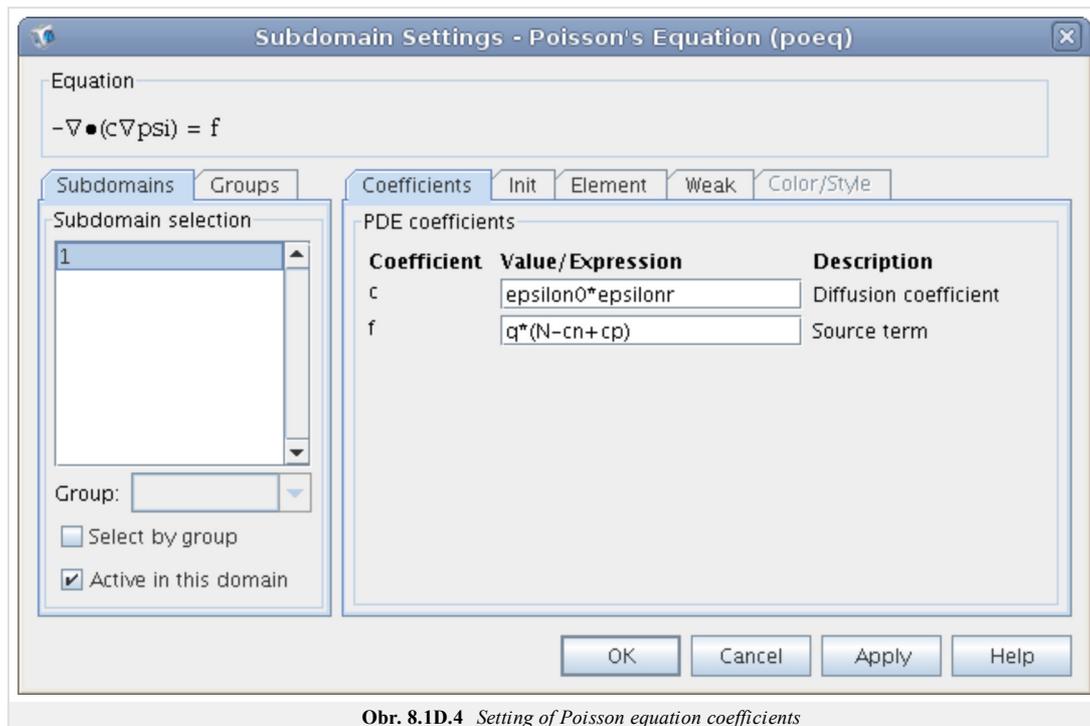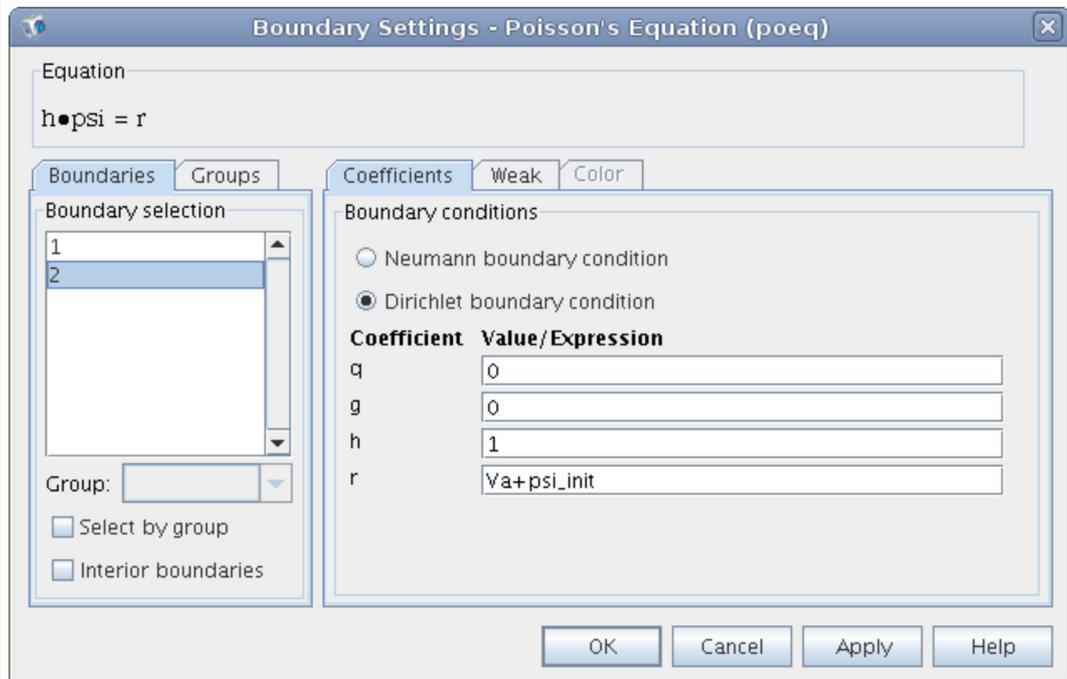

Obr. 8.1D.4 *Setting of Poisson equation coefficients*

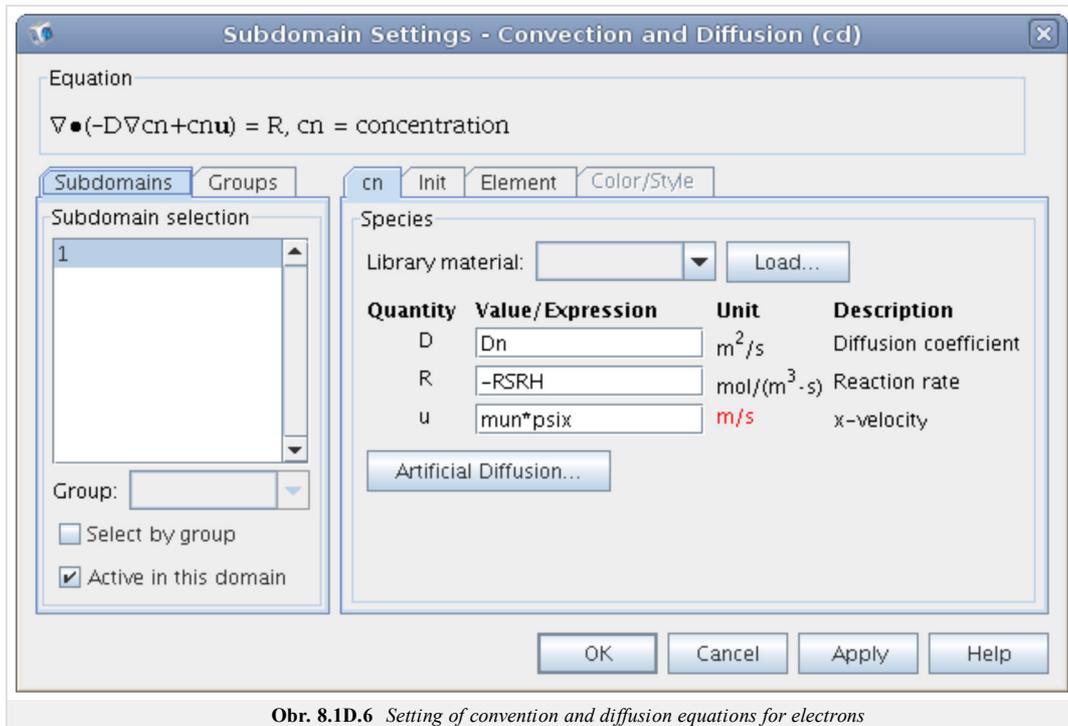The fig. **8.1D.5** shows the set of boundary conditions at points x_1 and x_2.

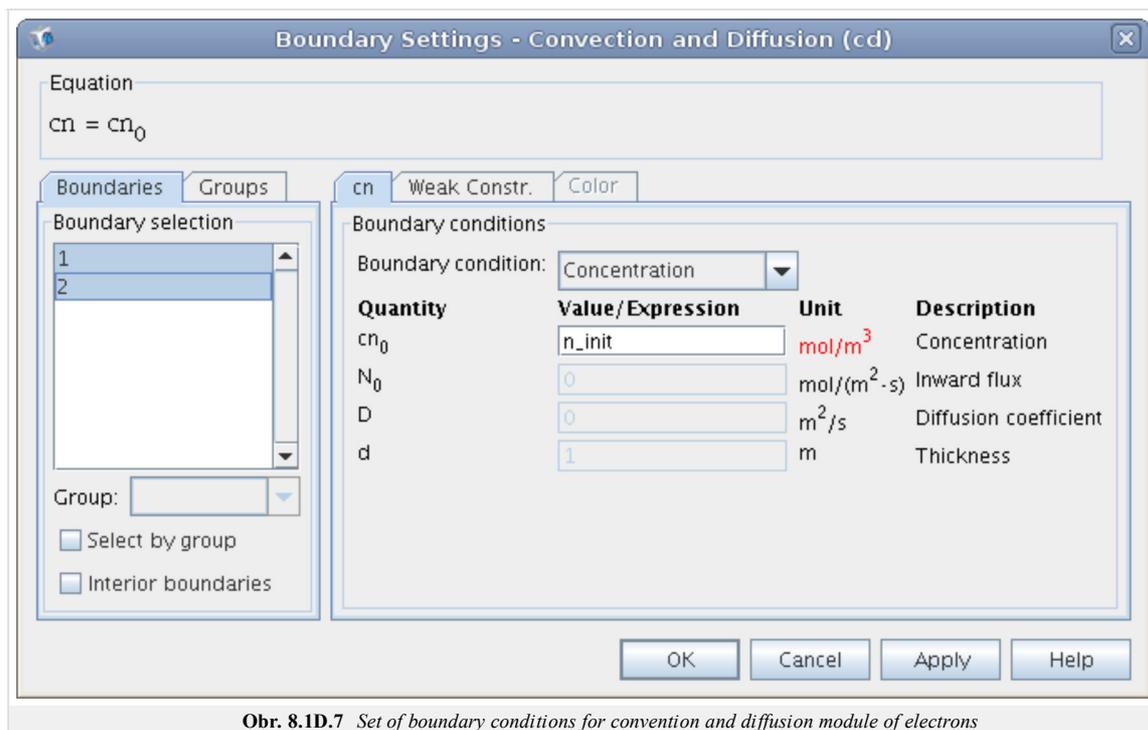a)

b)

Obr. 8.1D.5 *Setting of boundary conditions of the electrostatic module*

Module convention and diffusion of electrons

In fig. **8.1D.6** is depicted the preset formula over the diode domain, where the tab *Init* is set to *n_init*.

Obr. 8.1D.6 *Setting of convention and diffusion equations for electrons*

The fig. **8.1D.7** shows identical set of boundary conditions of points x_1 and x_2.

Obr. 8.1D.7 *Set of boundary conditions for convention and diffusion module of electrons*

Module convention and diffusion of holes

In fig. **8.1D.8** is depicted the preset formula over the diode domain, where the tab *Init* is set to *p_init*.

Obr. 8.1D.8 *Setting of convention and diffusion equations for holes*
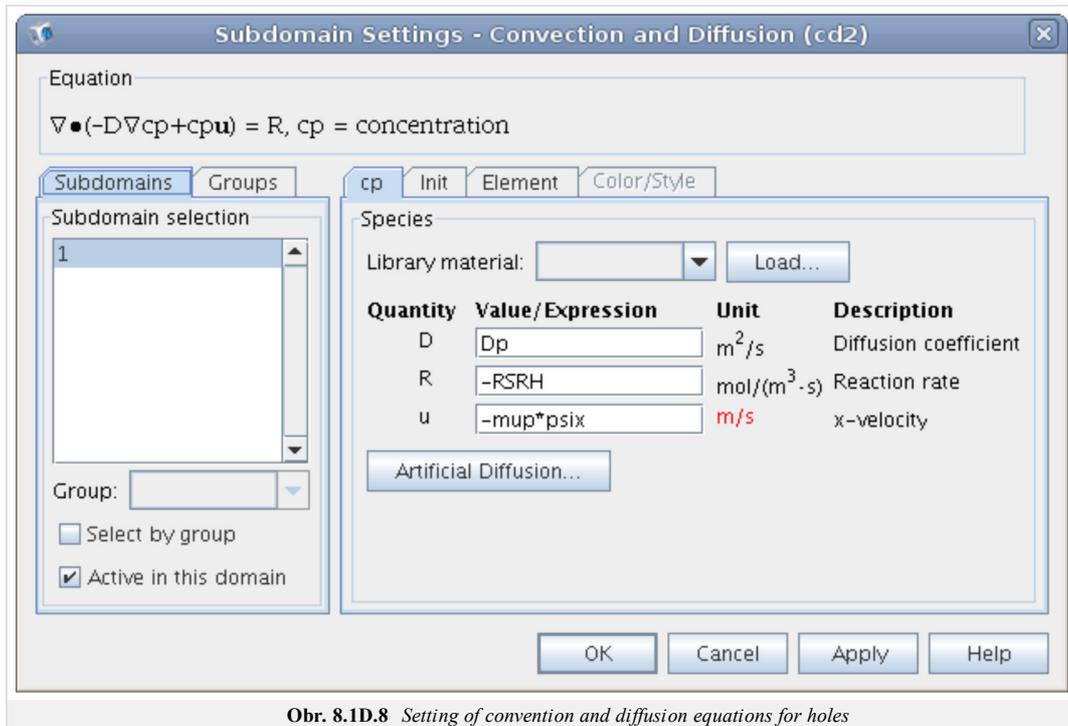
The fig. **8.1D.9** shows identical set of boundary conditions of points x_1 and x_2.

Obr. 8.1D.9 *Setting of boundary conditions for convention and diffusion module of holes*
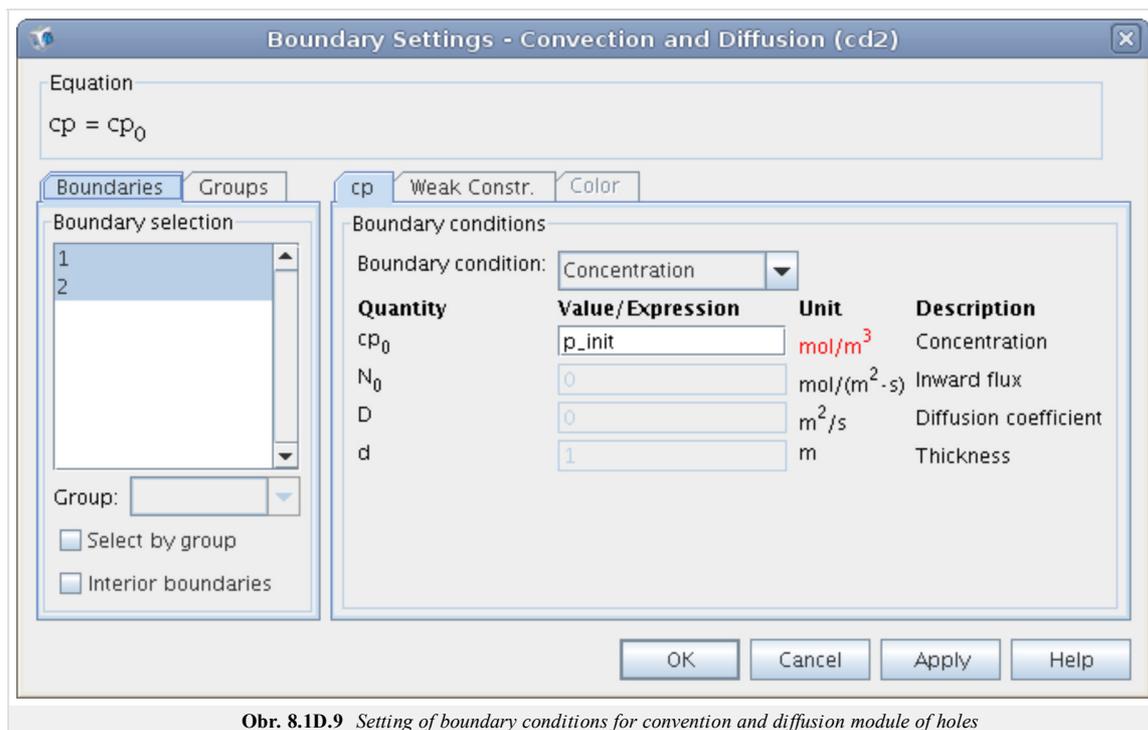
Other settings

Setting of discretization mesh is shown in fig. **8.1D.10** and settings of solver in fig. **8.1D.11**.
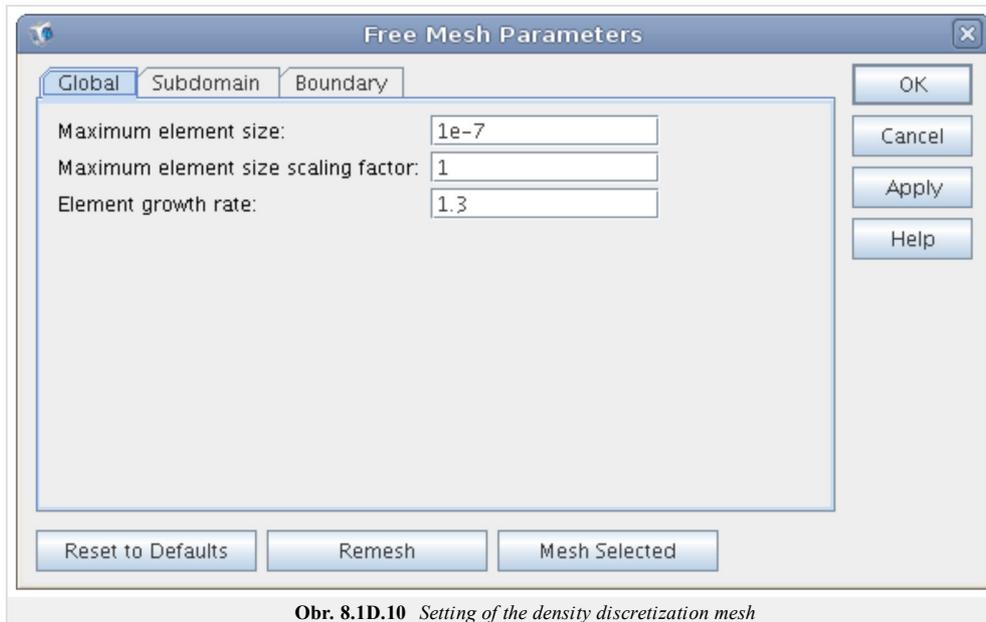
Obr. 8.1D.10 *Setting of the density discretization mesh*
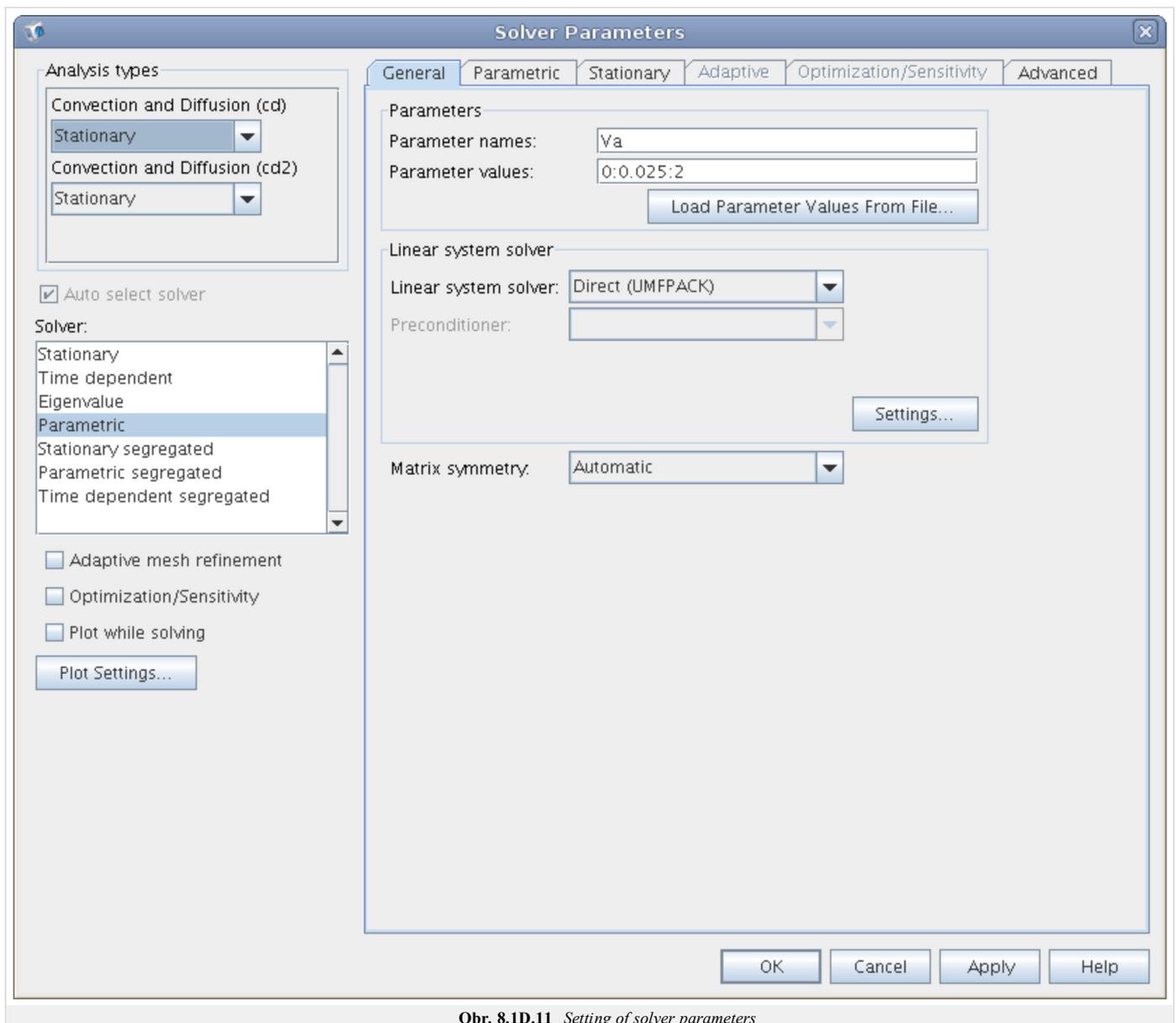
Obr. 8.1D.11 *Setting of solver parameters*

8.1 Drift diffusion model

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**10** of **10**:

Question 1

Transfer of charge in the volume of semiconductors is mediated by...

Possible answers for question 1:

- ⊙ … photons.
- ⊙ … the electrons.
- ⊙ … the electrons and holes.
- ⦿ I won't answer this question

Question 2

What forces acting on charge carriers is considered in the drift and diffusion model?

Possible answers for question 2:

- ⊙ The forces of electric and magnetic fields.
- ⊙ Diffusive and Lorenz force excluding magnetic effects.
- ⊙ Gravitational force.
- ⦿ I won't answer this question

Question 3

What function is used to model doping profiles resulting from the processes of diffusion and ion implantation?

Possible answers for question 3:

- ⊙ Linear.
- ⊙ Gaussian.
- ⊙ Quadratic.
- ⦿ I won't answer this question

Question 4

What is defined in the model drift and diffusion with doping profile and concentration of charge carriers?

Possible answers for question 4:

- ⊙ Concentrations of free charge.
- ⊙ Built-in potential.
- ⊙ Barrier potential.
- ⦿ I won't answer this question

Question 5

Current density in semiconductors doped by donors only is given by...

Question 6

What is the dominant cause of generation and recombination of free charge carriers in semiconductors in thermal equilibrium?

Possible answers for question 6:

Transitions of electrons.

Transitions of photons.

Transitions of phonons.

I won't answer this question

Question 7

Description of the Gunn phenomenon at a macroscopic level, it is possible by...

Possible answers for question 7:

… appropriate modeling electron mobility.

… tools of quantum mechanics.

… empirical corrections.

I won't answer this question

Question 8

Boundary conditionsfor ohmic contact assumed...

Possible answers for question 8:

… zero normal component of current density.

… thermal equilibrium of semiconductor and the absence of free charge at the interface.

… validity of Gauss law.

I won't answer this question

Question 9

Boundary condition for Shottky contact assumed...

Possible answers for question 9:

… apriori knowledge of the Shottky barrier value for the given material.

… infinite speed of thermal recombination.

… negligible influence of thermal recombination.

I won't answer this question

Question 10

What are the dependent variables of the drift and diffusion model?

Possible answers for question 10:

Time and temperature

Drift velocity of charge carriers and electric potential.

Electric potential and concentration of electrons and holes.

I won't answer this question

see result

Chapter 9: Microwave circuits

Introduction

Sorry, this chapter has not been implemented yet.

9.1 Impedance matching

Basic theory

Sorry, this chapter has not been implemented yet.

9.1 Impedance matching

Tips for WinMIDE

Sorry, this chapter has not been implemented yet.

9.2 Passive circuits

Basic theory

Sorry, this chapter has not been implemented yet.

9.2 Passive circuits

Tips for WinMIDE

Sorry, this chapter has not been implemented yet.

9.3 Narrow-band amplifier

Basic theory

Sorry, this chapter has not been implemented yet.

9.3 Narrow-band amplifier

Tips for WinMIDE

Sorry, this chapter has not been implemented yet.

Chapter 10: Short wave wire antennas

Introduction

At present, Radio Amateur transmitting is one of the most popular sports in many countries. Around the word, about 5 million people are interested in radio amateur transmitting. DXing belongs to the most popular part of this hobby (connecting people from other countries).

People use ionosphere propagation in bands 160 m (1.8 MHz), 80 m (3.5 MHz), 40 m (7 MHz), 30 m (10 MHz), 20 m (14 MHz), 17 m (18 MHz), 15 m (21 MHz), 12 m (24 MHz) and 10 m (28 MHz).

The antenna is one of most important part of the technical equipment. The antenna must be of a mono-band or multi-band design, the radiation must be omni-directional or directional in the horizontal plane.

The text presents most popular short wave antennas. The antennas were designed in the freeware 4NEC program.

10.1 History of antennas and wireless communication

Basic theory

The history of the wireless communication started with the discovery of a coherer which is an original detection device for the detection of electromagnetic waves in long-wave frequency band. To the coherer, a long-wave wire antenna was connected. In first experiments, the wire antenna was unmatched. When Marconi implemented first tests in 1896, he used the transmatched long-wire hang up on kite (see fig. **10.1A.1**).

Consequently, he used wire antenna with the defined length depended on the wavelength. For the first radio connection across the Atlantic Ocean, the quarter-wavelength monopole was used (see fig. **10.1A.2**).
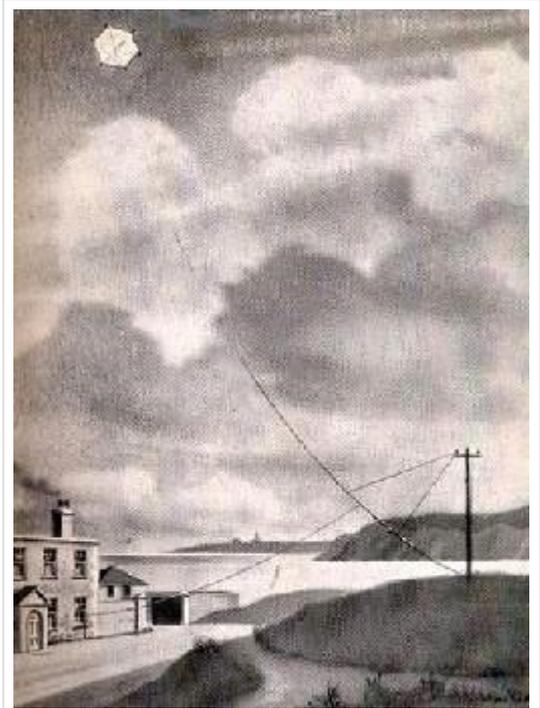
Fig. 10.1A.1 *Historical experiment of Marconi*
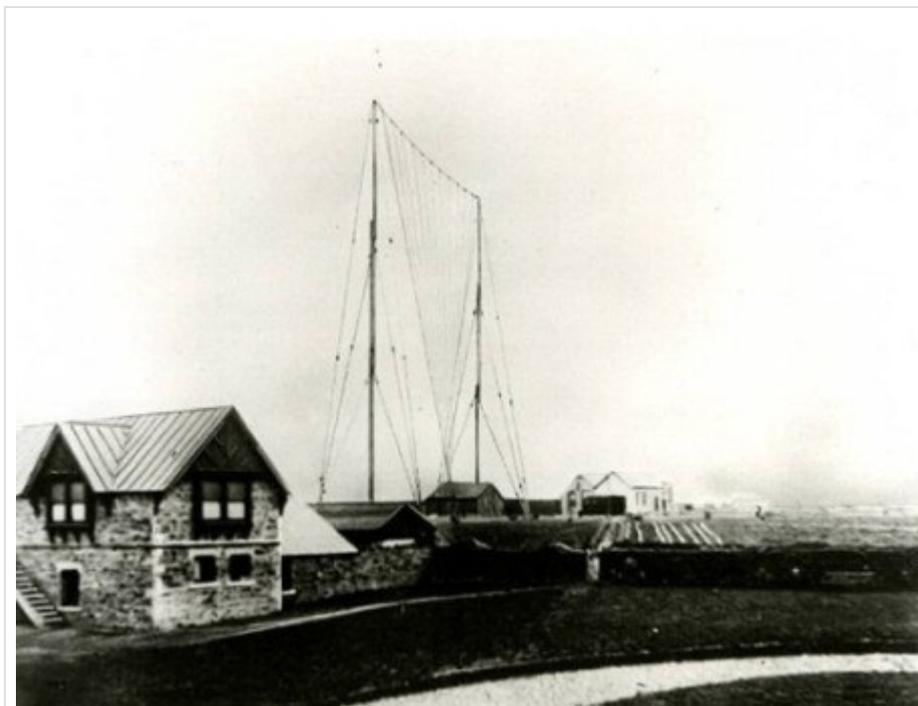
Fig. 10.1A.2 *Quarter-wavelength dipole antenna of Marconi*

After World War II, the antennas with a single fed element and parasitic elements were designed. Yagi-Uda antenna is the most famous antenna of this type. Yagi-Uda antenna was originated in 1926, and was practically used after World War II. The microwave circuits shifted the operation frequencies to higher bands. The wavelength of electronics was decreased, and new possibilities of the design were created.

In the following chapters, wire antennas, which are frequently used in various communication applications, are described.

10.1 History of antennas and wireless communication

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

First message on long-wave across Atlantic was transsmited by...

Possible answers for question 1:
- ○ … Marconi.
- ○ … Popov.
- ○ … prof. Raida.
- ● I won't answer this question

Question 2

Koherer is device...

Possible answers for question 2:
- ○ … in transmit part of trasciever.
- ○ … to receive EM wave.
- ○ … which amplified the EM wave behind detector.
- ● I won't answer this question

Question 3

The expansion of directional antennas and applications are...

Possible answers for question 3:
- ○ … before World War 1.
- ○ … during World War 2.
- ○ … during "Cold" War.
- ● I won't answer this question

see result

10.2 Half-wavelength dipole

Basic theory

A dipole antenna, developed by Heinrich Rudolph Hertz around 1886 is an antenna with a center-fed driven element for transmitting or receiving radio frequency energy. These antennas are the simplest practical antennas from a theoretical point of view. Typically, a dipole antenna is formed by two quarter wavelength conductors or elements placed back to back for a total length of $\lambda/2$.

Fig. 10.2A.1 *Wire model of dipole*
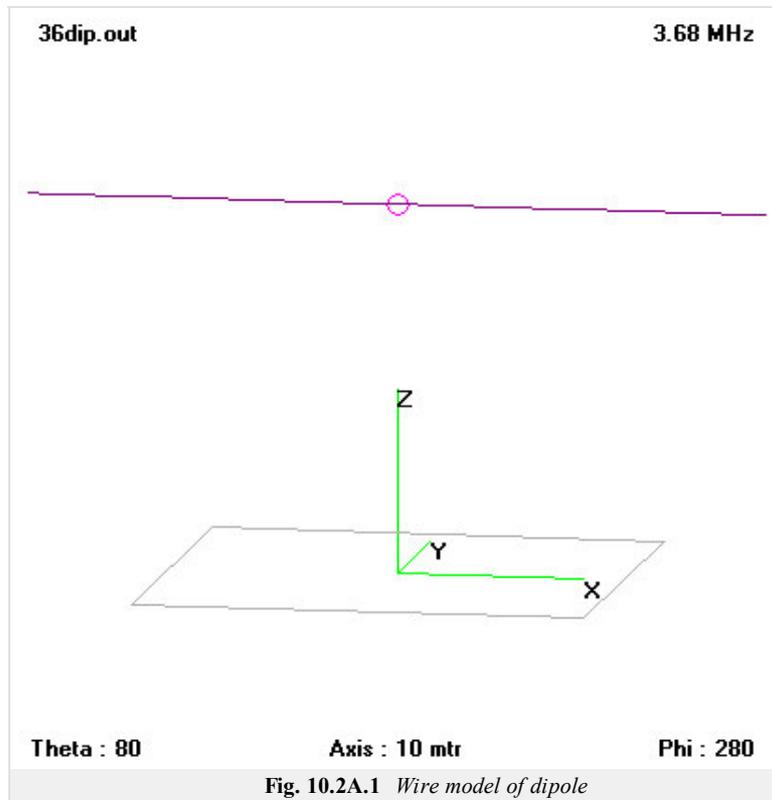
A standing wave on an element of a length $\sim \lambda/4$ yields the greatest voltage differential, as one end of the element is at a node while the other element is at an antinode of the wave. Assuming a sinusoidal distribution, the current impressed by this voltage differential is given by

$$I = I_0 e^{j\omega t} \cos(kl).$$
(10.2A.1)

Here, I_0 is the current in the maximum, denotes angular frequency and t time, k is wavenumber and l is the length of the dipole arm.

In the far-field zone, the electric field of a radiating electromagnetic wave is given by:

$$E_\theta = \frac{-jI_0}{2\pi\varepsilon_0 cr} \frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} e^{j(\omega t - kr)}.$$
(10.2A.2)

Here, E is the respective component of the electric field, ε_0 is the permitivity of vacuum, c is the speed of light in vacuum, r is the distance from the observation point, is the angle measured from the dipole axis, and other symbols were already explained.

In figures below, frequency responses of standing wave ration and input impedance of a dipole ($l = 19.64$ m) is depicted. The half-wavelength resonance appears at the frequency $f = 3.68$ MHz.
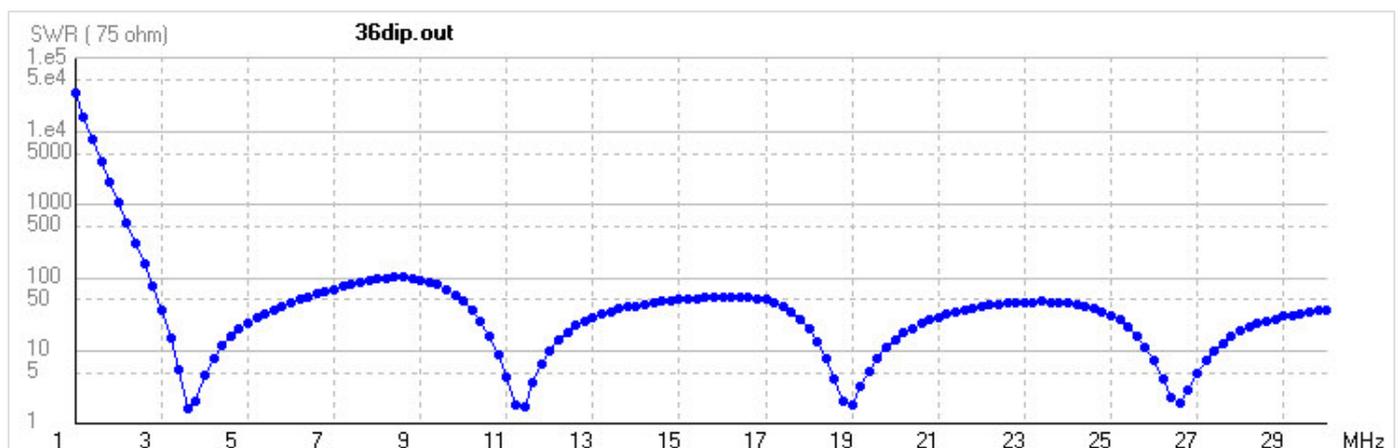
Fig. 10.2A.3 *Frequency response of input impedance of half-wavelength dipole (l = 19.64 m)*
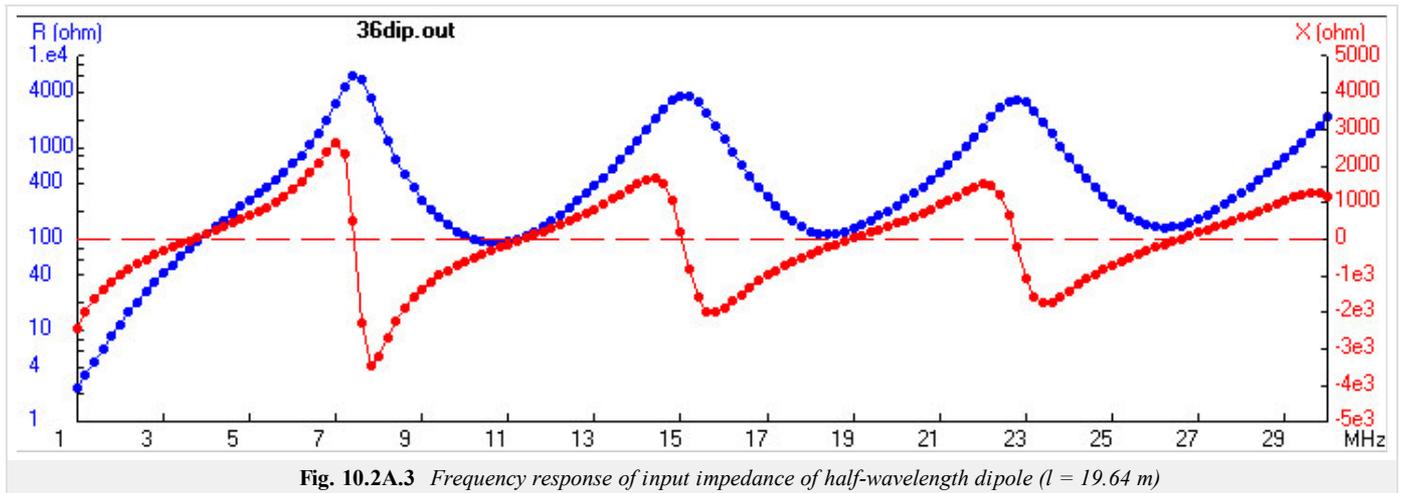
In fig. **10.2A.4**, far field radiation pattern of a dipole in the half-wavelength resonance is depicted. The maximum radiation appears in the direction perpendicular to the axis of the dipole. The antenna does not radiate in the axis of the dipole. Since the dipole is placed above the conducting plane, the antenna radiates into a half-space only.
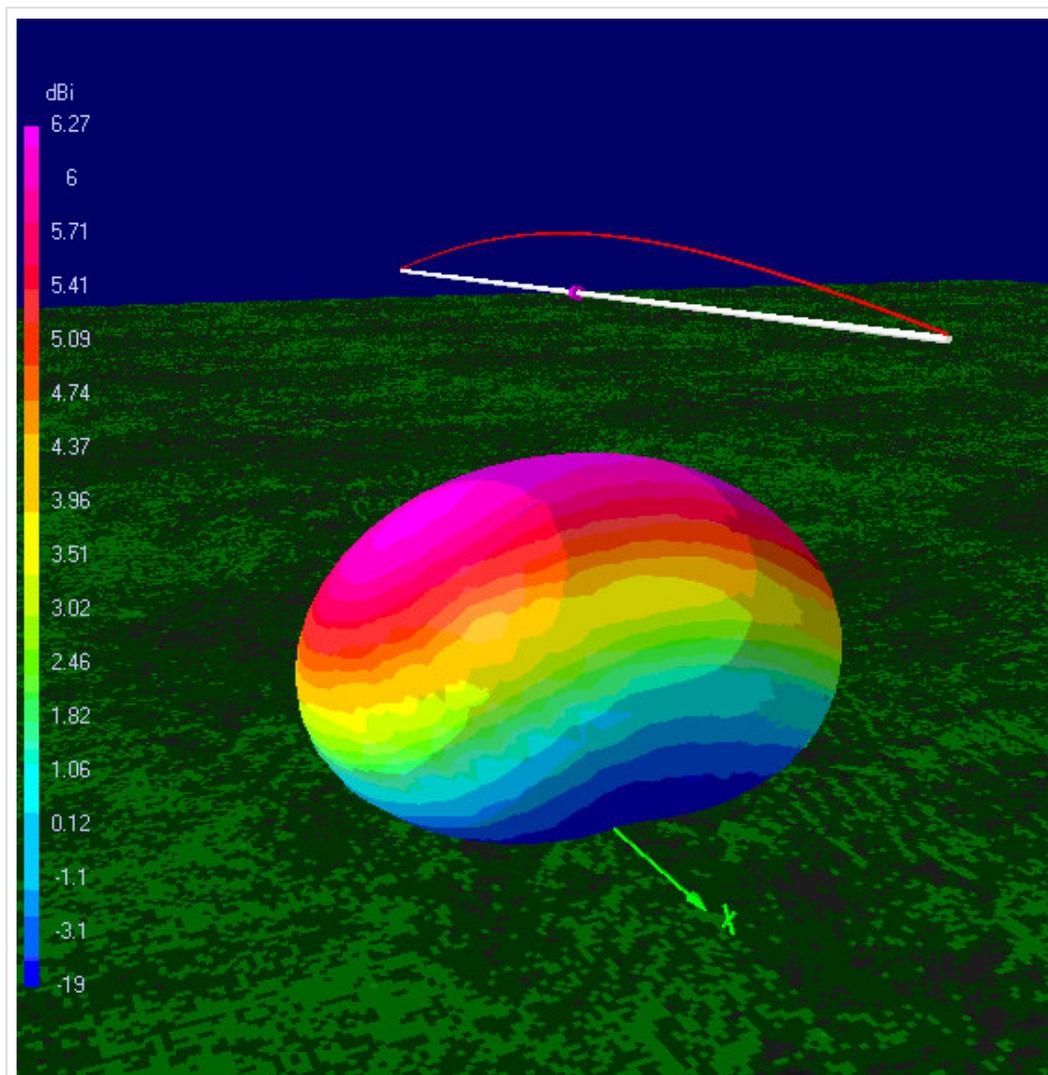
Fig. 10.2A.4 *Far field radiation pattern of half-wavelength dipole above the conducting plane in the height h = 20 m (l = 19.64 m, f = 3.68 MHz)*

10.2 Half-wavelength dipole

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

Input impedance of dipole antenna in resonance is...

Possible answers for question 1:

- ○ … 50 Ω.
- ○ … 75 Ω.
- ○ … 73 + j42 Ω.
- ● I won't answer this question

Question 2

Maximum gain of half-wave dipole is...

Possible answers for question 2:

- ○ … 2,13 dBi.
- ○ … 1,64 dBd.
- ○ … 2,13 dBd.
- ● I won't answer this question

Question 3

Half-wave dipole is...

Possible answers for question 3:

- ○ … symmetrical antenna .
- ○ … wideband antenna.
- ○ … asymmetrical antenna.
- ● I won't answer this question

see result

10.3 T2FD Tilted Terminated Folded Dipole

Basic theory

The T2FD (Tilted Terminated Folded Dipole) is a general-purpose shortwave antenna developed in the late 1940s by the United States Navy **[45]**. It performs reasonably well over a broad frequency range, without marked dead spots in terms of frequency, direction, or angle of radiation above the horizon.

Although inferior in electrical terms **[46]** to antennas specifically designed for given frequency bands, or optimized for directionality, its all-around performance, relatively modest size, low cost, and the fact that it does not require any complex electronic matching to operate with a standard shortwave transmitter, have made it popular in professional shortwave communications.

Also, since the late 1980s, amateur radio operators and hobby shortwave listeners have rediscovered this antenna, especially for broadcast receiving and for amateur two-way modes such as Morse code and PSK31 where brute force performance is not as important as a steady signal. There have also been (disputed) claims that this antenna is comparatively insensitive to man-made radio interference, making it useful in urban environments, where a low noise floor is often more beneficial than high received signal strength. The T2FD is useful for hidden indoor systems, or where several optimized frequency-specific antennas cannot be accommodated. For example: an indoor antenna only 7 m long will allow operation on all amateur HF bands above 14 MHz on transmit, and down to 7 MHz on receive.

Fig. 10.3A.1 *Schematic of T2FD antenna*

Fig. 10.3A.2 *5 to 30 meter band antenna design*
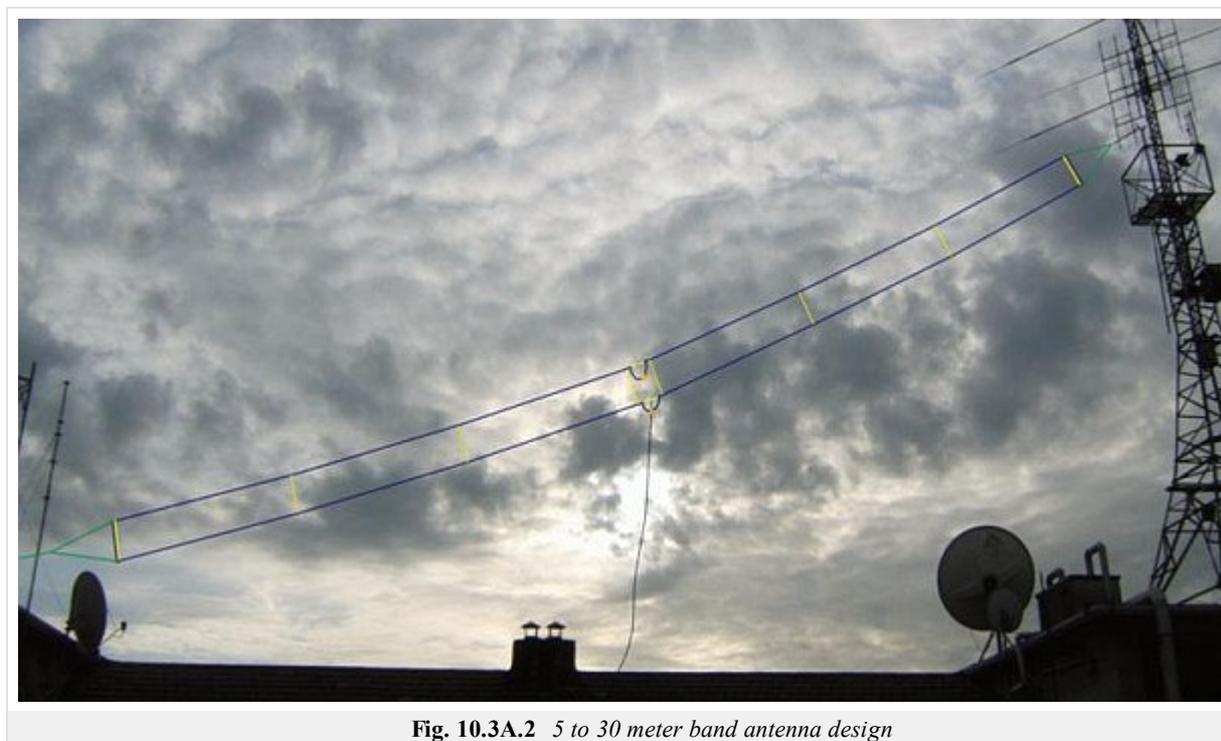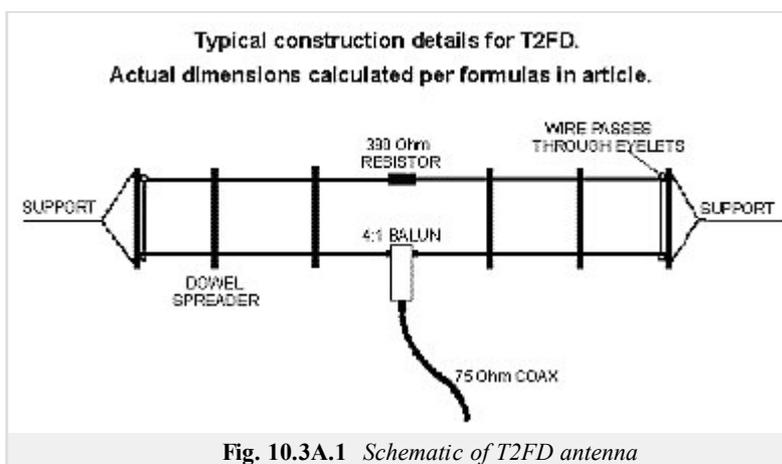
A typical T2FD is built as follows, out of two parallel wire conductors:

- Span equals to 1/3 of the lowest required wavelength.
- Distance between upper and lower conductors equal to 1/100 of the wavelength. This distance is maintained by a number of insulating
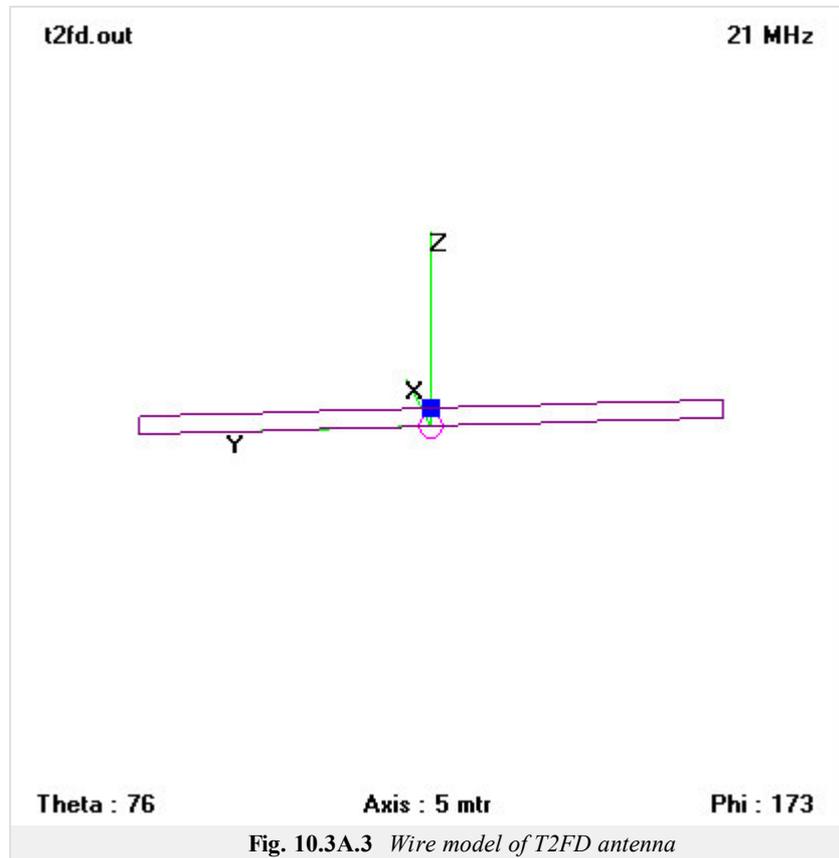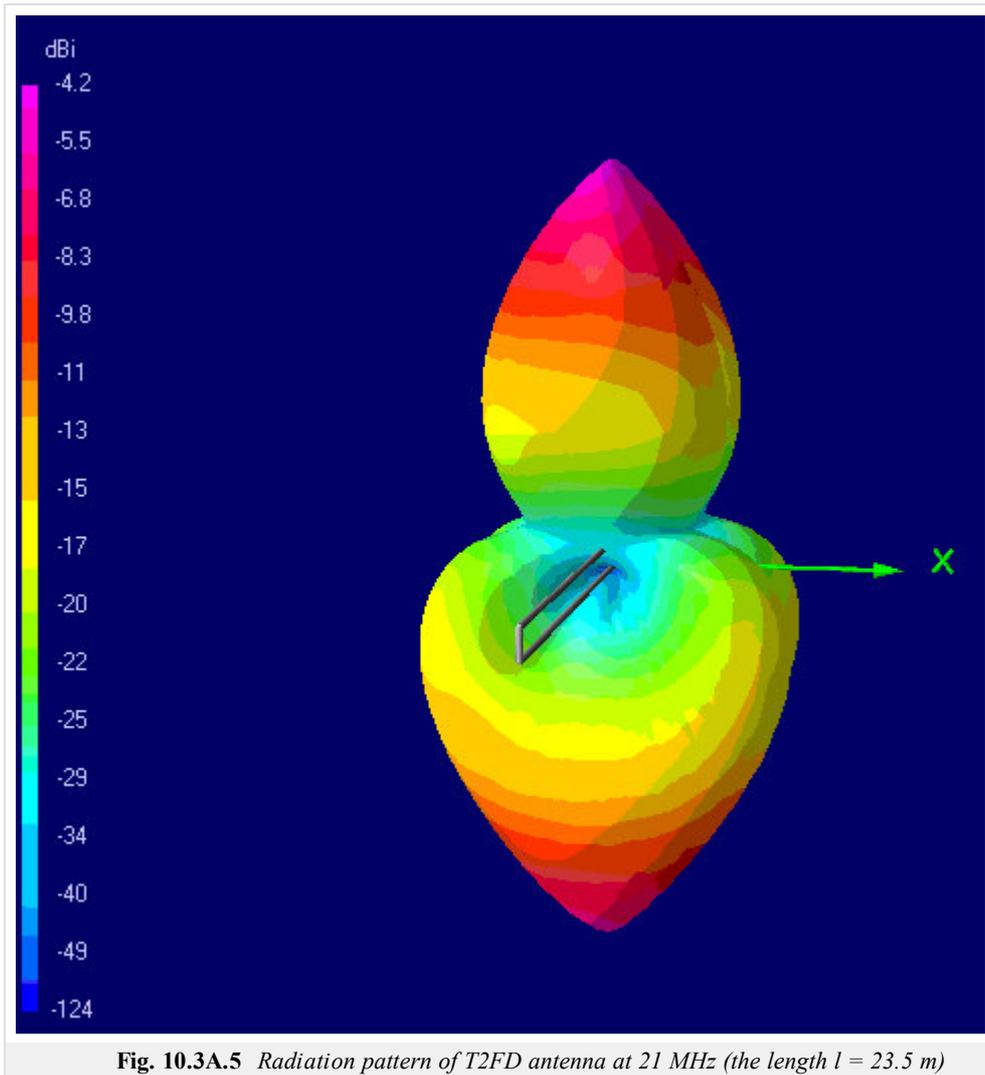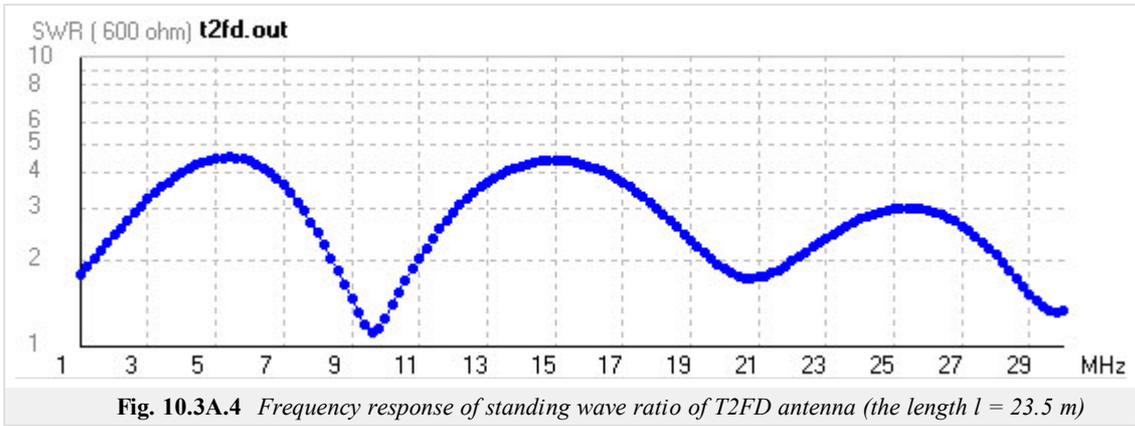
dowels.

- Two dowels at the ends are tied to non-conducting ropes, in turn tied to supports.

- Upper and lower conductors are connected at the ends, by wire sections that follow the two end dowels.

- Fed in the middle of the lower conductor, with an impedance in the order of 300 Ω, balanced, through a standard 4:1 balun. This provides an acceptable all-frequency match to commonly available 75 Ω coaxial cable.

- Terminated in the middle of the upper conductor with a 400 Ω non-inductive resistor, rated to safely absorb at least 1/3 of the applied transmitter power. The resistor absorbs a growing portion of the RF power (either captured from the air or supplied by a transmitter) as the operating frequency nears the lower limit of the design range.

- In order to make it roughly omnidirectional, it is ideally strung sloping at an angle of 20 to 40 degrees from horizontal, but will also function satisfactorily if mounted horizontally, as long as it is pulled-out in a reasonably straight line.

Such an antenna is usable for both local and medium-long distance communication across a frequency range of about 1:6. For example, an antenna for the lower portion of shortwave (say, 3 to 18 MHz) will be roughly 33 m long, with conductors spaced 1 m. For the higher portion of shortwave (5 to 30 MHz), this antenna will be roughly 20 m long, with a spacing of 60 cm. If such long spans cannot be accommodated, smaller antennas will still give adequate receive-only performance down to about half of their lowest design frequency. Transmit performance, however, degrades rapidly below the lowest design frequency.

As a broadband antenna, the T2FD will normally display a reasonably low standing wave ratio across its entire frequency range. However; at some frequencies there may be a moderately reactive element within the loading, so the use of an antenna tuner may be beneficial when using modern, solid-state transmitters at anything approaching their rated power output.

Many ready-made commercial versions of the T2FD are available for the professional / military, amateur radio, and hobby listening markets.

Fig. 10.3A.3 *Wire model of T2FD antenna*

Fig. 10.3A.4 *Frequency response of standing wave ratio of T2FD antenna (the length l = 23.5 m)*

Fig. 10.3A.5 *Radiation pattern of T2FD antenna at 21 MHz (the length l = 23.5 m)*

10.3 T2FD Tilted Terminated Folded Dipole

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

The T2FD antenna is terminated by resistor with impedance about...

Possible answers for question 1:

- ○ … 50 Ω.
- ○ … 300 Ω.
- ○ … 400 Ω.
- ◉ I won't answer this question

Question 2

The T2FD antenna is antenna with...

Possible answers for question 2:

- ○ … high gain.
- ○ … low profile against half-wave dipole.
- ○ … high efficiency of radiation.
- ◉ I won't answer this question

Question 3

Input impedance of T2FD antenna is about...

Possible answers for question 3:

- ○ … 300 Ω.
- ○ … 50 Ω.
- ○ … 400 Ω.
- ◉ I won't answer this question

see result

10.4 Quarter wavelength monopole antenna − ground-plane antenna
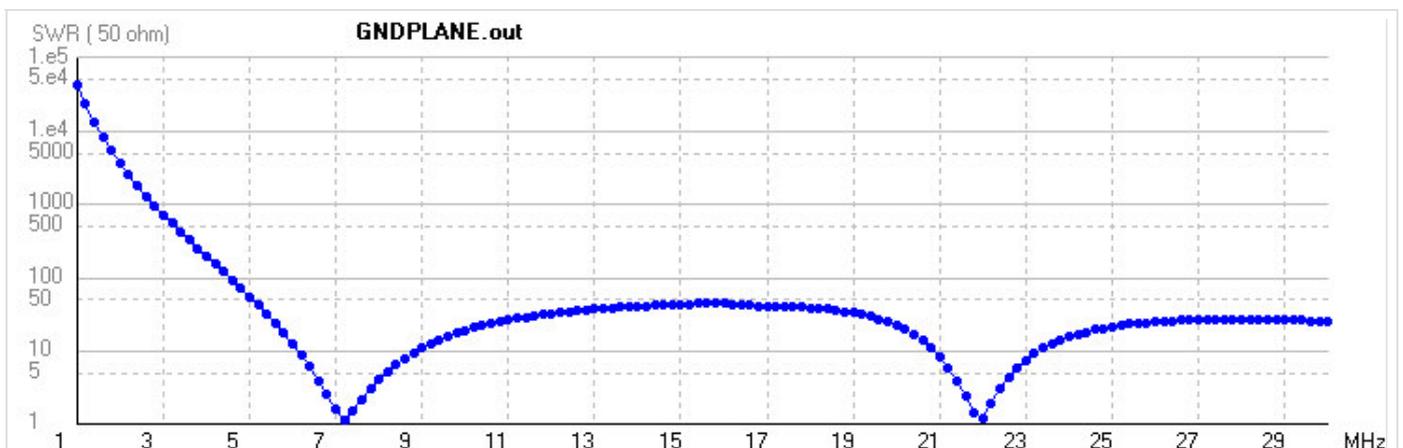
Basic theory

The quarter wave antenna is a single element antenna fed at one end that behaves as a dipole antenna. It is formed by a conductor in length $\lambda/4$. It is fed in the lower end, which is near a conductive surface which works as a reflector. The current in the reflected image has the same direction and phase that the current in the real antenna. The set quarter-wave plus image forms a half-wave dipole that radiates only in the upper half of space.
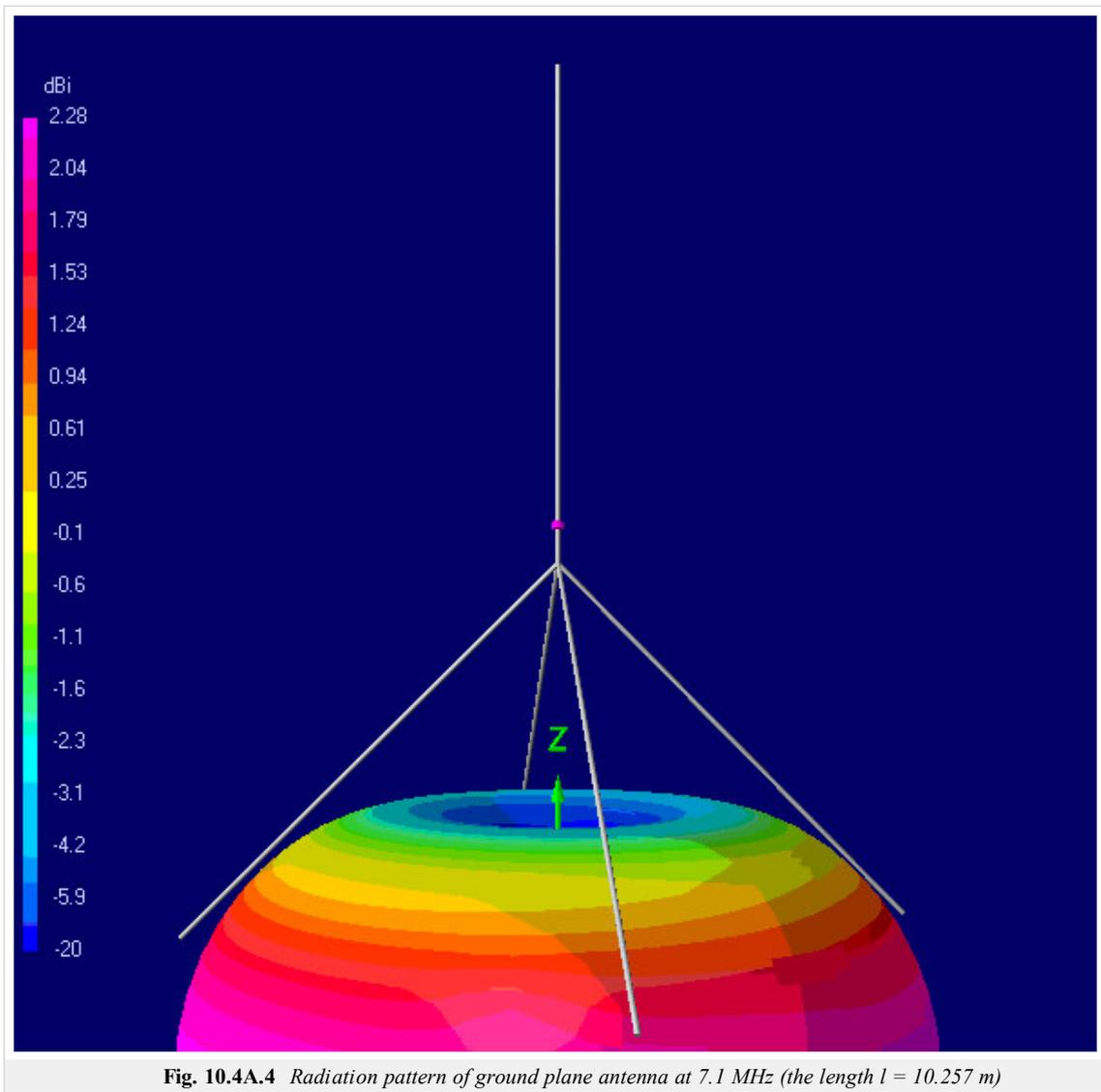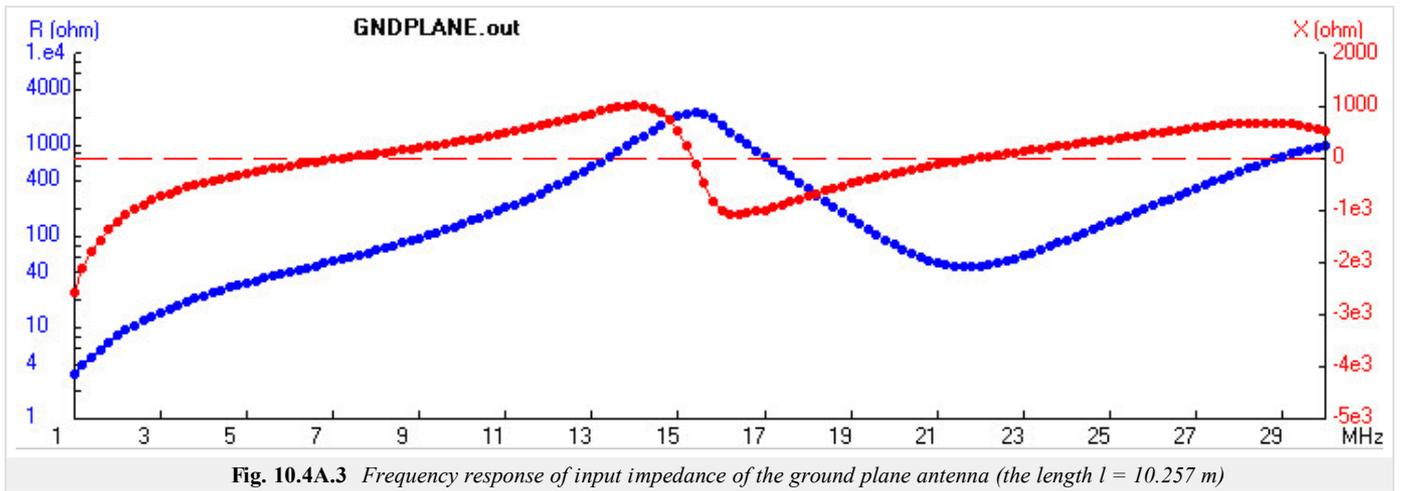
In this upper side of space the emitted field has the same amplitude of the field radiated by a half-wave dipole fed with the same current. Therefore, the total emitted power is one-half the emitted power of a half-wave dipole fed with the same current. As the current is the same, the radiation resistance (real part of series impedance) will be one-half of the series impedance of a half-wave dipole. As the reactive part is also divided by 2, the impedance of a quarter wave antenna is $36 + j21$. The gain is the same as that for a half-wave dipole ($\lambda/2$) that is 2.14 dBi.

Fig. 10.4A.1 *Wire model of ground plane antenna*
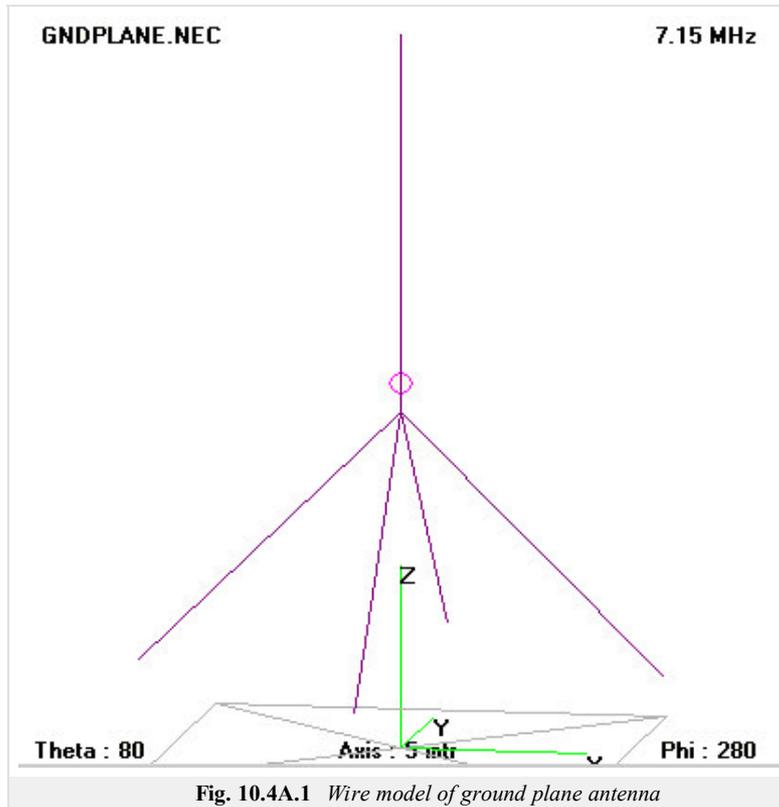
Far from the antenna and near the ground, electromagnetic fields and radiation patterns are the same as for a half-wave dipole. The impedance is not the same a with a good conductor ground plane. Conductivity of earth surface can be improved with an expensive copper wire mesh.

When ground is not available, as in a vehicle, other metallic surfaces can serve a ground plane, for example the roof of the vehicle. In other situations, radial wires placed at the foot of the quarter-wave wire can simulate a ground plane.

Fig. 10.4A.2 *Frequency response of standing wave ratio of ground plane antenna (the length l = 10.257 m)*

Fig. 10.4A.3 *Frequency response of input impedance of the ground plane antenna (the length l = 10.257 m)*

Fig. 10.4A.4 *Radiation pattern of ground plane antenna at 7.1 MHz (the length l = 10.257 m)*

10.4 Quarter wavelength monopole antenna − ground-plane antenna

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

Quarter-wave monopole has input impedance...

Possible answers for question 1:
- ○ … similar as half-wave dipole.
- ○ … 2 times higher then half-wave dipole.
- ○ … 3 times smaller then half-wave dipole.
- ◉ I won't answer this question

Question 2

The quarter-wave monepole is...

Possible answers for question 2:
- ○ … asymmetrical antenna.
- ○ … symmetrical antenna.
- ○ … antenna with circular polarization.
- ◉ I won't answer this question

Question 3

The angle of wires of ground part of quarter-wave monopole depends...

Possible answers for question 3:
- ○ … the input impedance of the antenna.
- ○ … the gain of the antenna.
- ○ … the input impedance and gain of the antenna.
- ◉ I won't answer this question

see result

10.5 Discone antenna

Basic theory

A discone antenna is a version of a biconical antenna in which one of the cones is replaced by a disc. It is usually mounted vertically, with the disc at the top and the cone beneath.

Omnidirectional, vertically polarized and exhibiting unity gain, it is exceptionally wideband, offering a frequency range ratio of up to ~10:1. The radiation pattern in the vertical plane is quite narrow, making its sensitivity highest in the plane parallel to the Earth.

The discone's wideband coverage makes it attractive in commercial, military, amateur radio and radio scanner applications.

When employed as a transmitting antenna, it is often less efficient than an antenna designed for a more limited frequency range. SWR (standing wave ratio) is typically ~2:1 over the range of the design frequency to the second harmonic and ~3.1 thereafter.

A discone antenna typically has at least three major components: the disc, the cone, and the insulator.

The disc should have an overall diameter of 0.7 times a quarter wavelength of the antenna's minimum frequency. The antenna's feed point is at the center of the disc. It is usually fed with 50 Ω coaxial cable, with the center conductor connected to the disc, and the outer conductor to the cone.

Fig. 10.5A.1 *Discone antenna on the roof*

The length of the cone should be a quarter wavelength of the antenna's minimum operating frequency. The cone angle is generally from 25 to 40 degrees.

The disc and cone must be separated by an insulator, the dimensions of which determine some of the antenna's properties.

In order to extend low-frequency response, a vertical whip may be placed affixed vertically to the disc. But this may reduce efficiency at higher frequencies. In this configuration, at lower frequencies the discone may more closely resemble a ground plane antenna or a coaxial dipole.
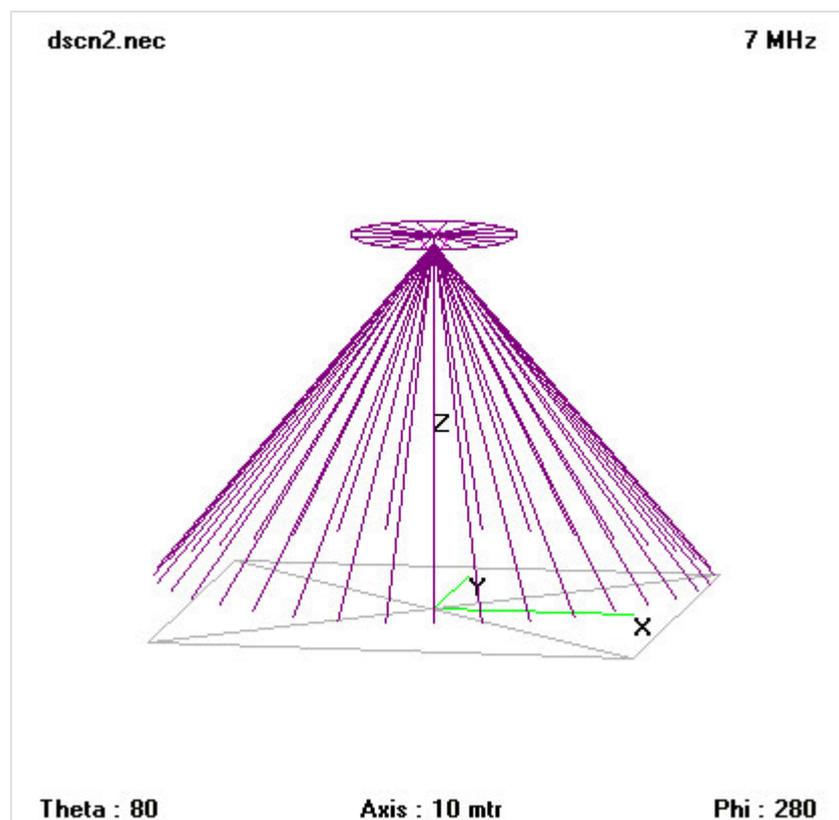
Fig. 10.5A.3 *Frequency response of standing wave ratio of a discone antenna (radius of cone d = 4.2 m, the length l = 16.8 m)*
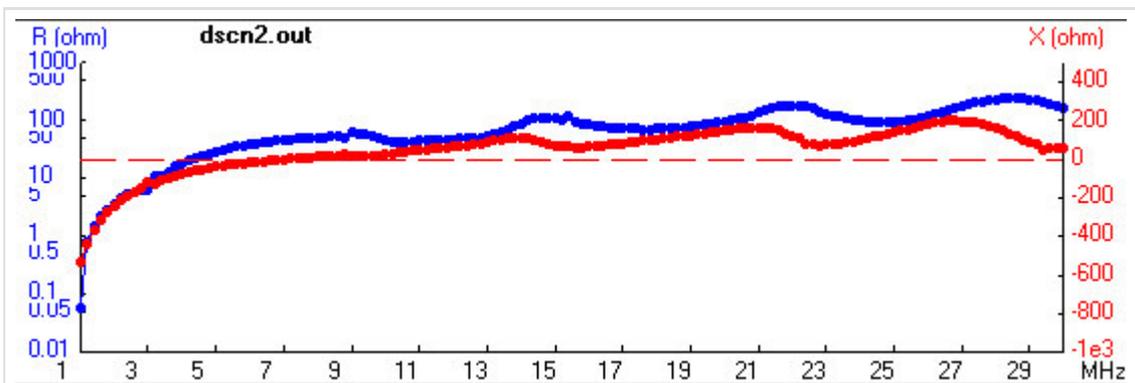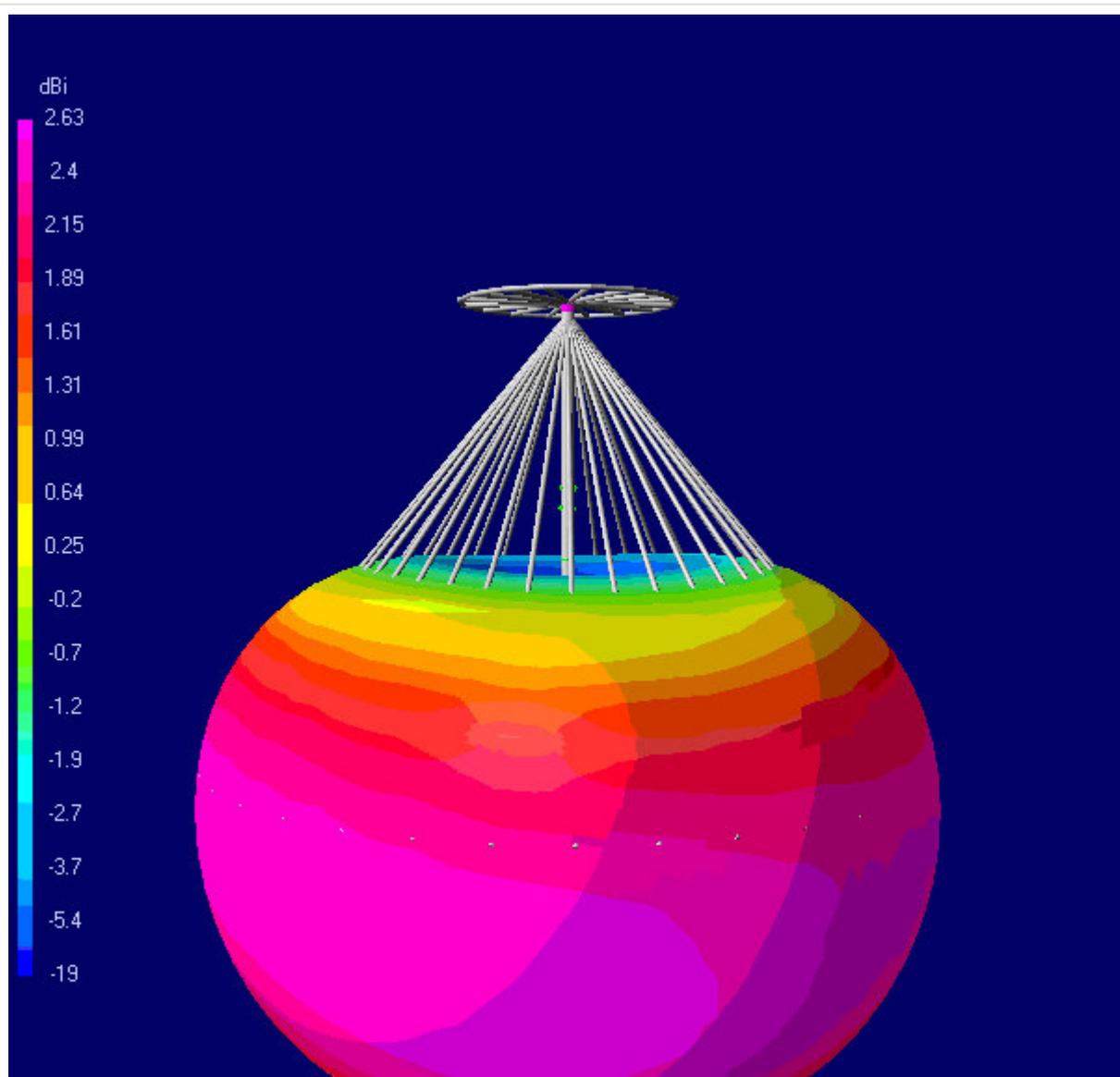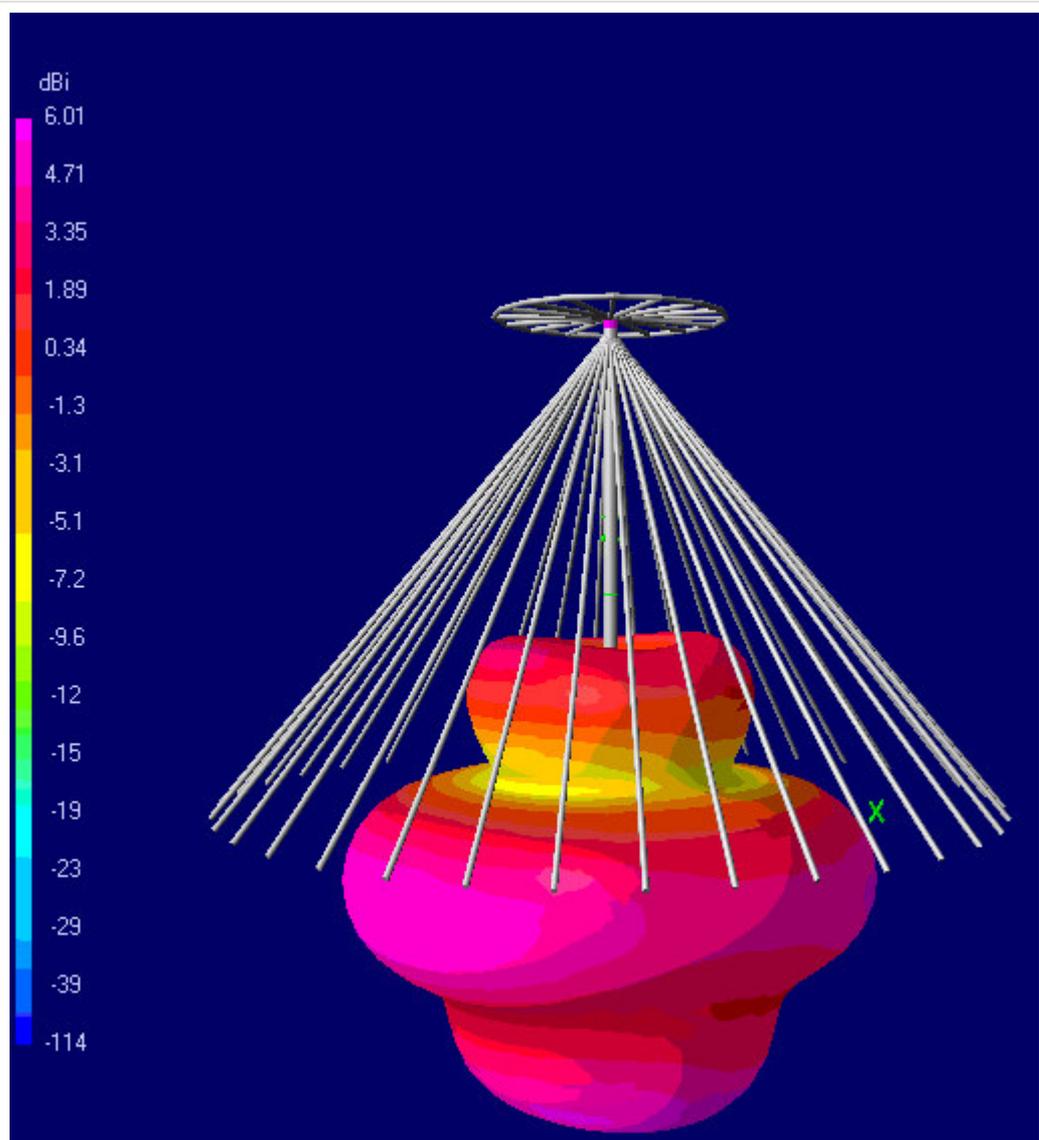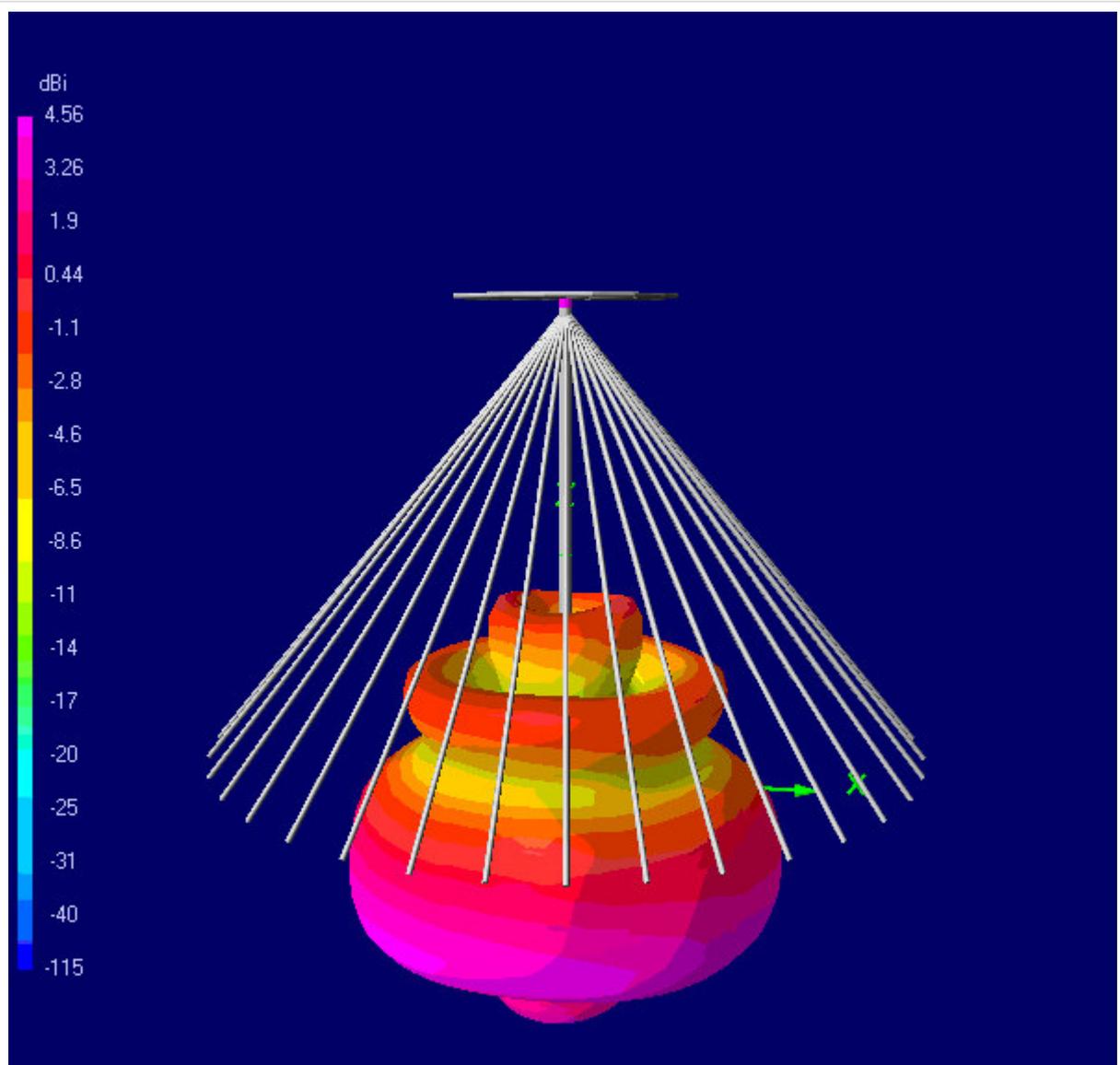

Fig. 10.5A.4 *Frequency response of input impedance of a discone antenna (radius of cone d = 4.2 m, the length l = 16.8 m)*

Fig. 10.5A.5 *Radiation pattern of a discone antenna at 7 MHz (radius of cone d = 4.2 m, the length l = 16.8 m)*
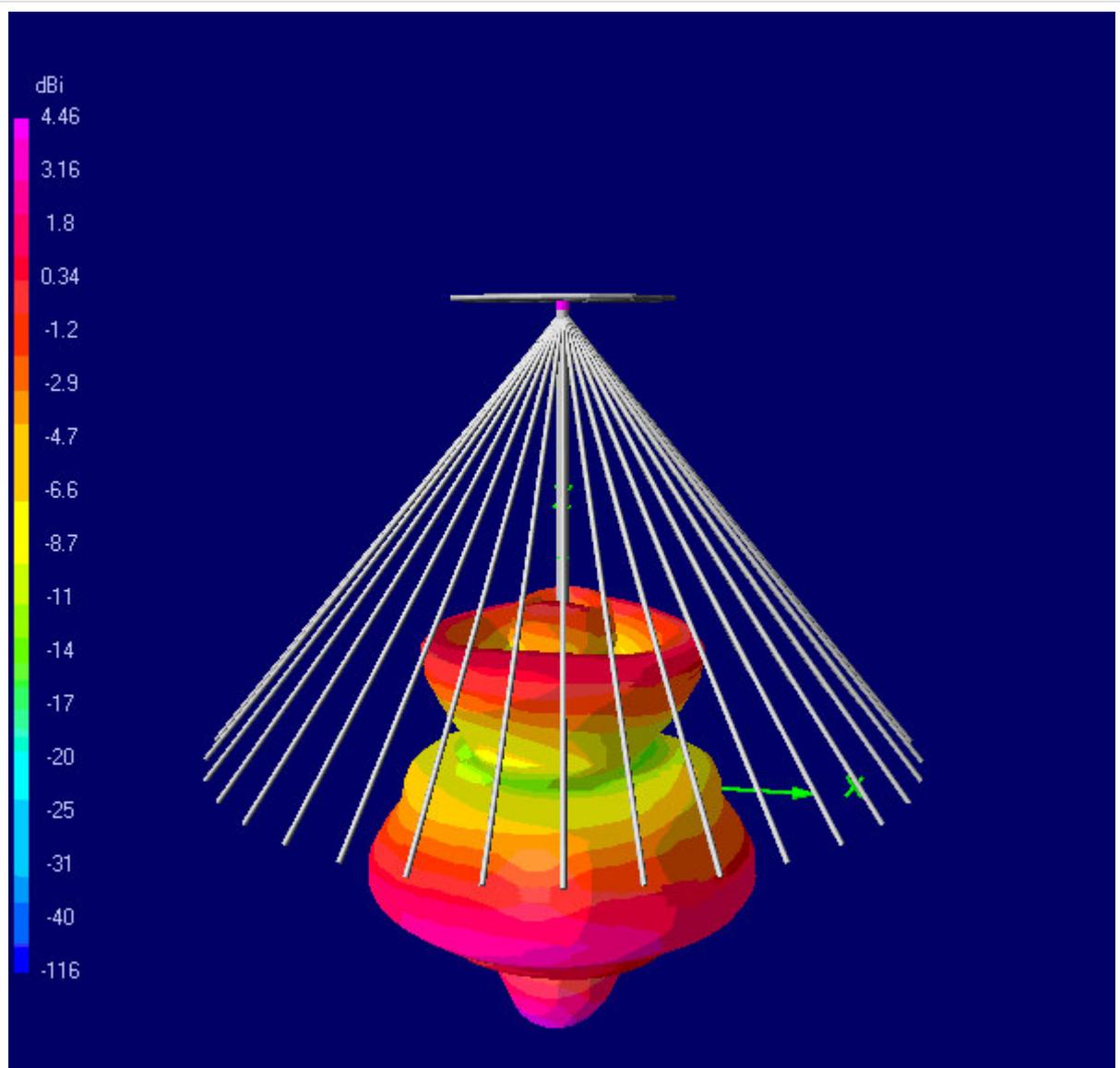
Fig. 10.5A.6 *Radiation pattern of a discone antenna at 14 MHz (radius of cone d = 4.2 m, the length l = 16.8 m)*

Fig. 10.5A.7 *Radiation pattern of a discone antenna at 21 MHz (radius of cone d = 4.2 m, the length l = 16.8 m)*

Fig. 10.5A.8 *Radiation pattern of a discone antenna at 28 MHz (radius of cone d = 4.2 m, the length l = 16.8 m)*

10.5 Discone antenna

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

Discone antena is...

Possible answers for question 1:

- ○ … horizontally polarized antenna with high gain.
- ○ … vertically polarized wide-band antenna.
- ○ … vertically polarized antenna with high gain.
- ● I won't answer this question

Question 2

Bandwidth of discone antenna is usually...

Possible answers for question 2:

- ○ … 1:1.
- ○ … 1:10.
- ○ … 1:3.
- ● I won't answer this question

Question 3

Input impedance of discone antenna is about...

Possible answers for question 3:

- ○ … 50 Ω and can be loaded by coaxial cable.
- ○ … 75 Ω and can be loaded by symmetrical feed-line.
- ○ … 300 Ω and must be loaded by symmetrical balun.
- ● I won't answer this question

see result

10.6 Inverted Vee

Basic theory

An inverted Vee antenna is a modified dipole antenna supported in the center with the ends lower than the center. Inverted Vee antennas are commonly used by many amateur radio stations, and aboard sailing vessels requiring better HF performance than available with a short whip antenna.
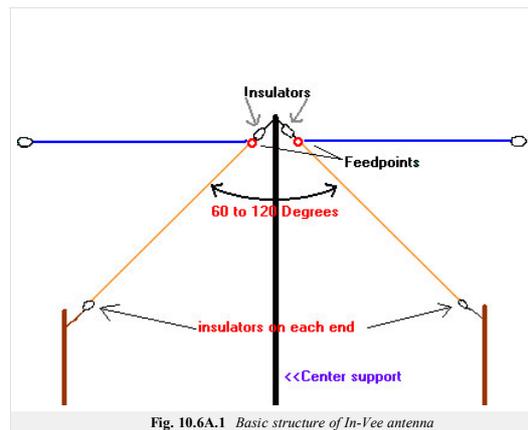

Fig. 10.6A.1 *Basic structure of In-Vee antenna*

Typically, the inverted Vee antenna requires only a single, tall support at the center (often a convenient tree limb), and the ends can be insulated and secured to anchors at or near ground level. This simplified arrangement has several advantages, including a shorter ground distance between the ends. For example, a dipole antenna for the 80 meter band requires a ground length of about 43 m from end to end. An inverted Vee with 12 m apex elevation requires only 35 m. For radio amateurs living on small parcels of property, such savings can make it possible to use the lower frequency amateur bands.

In theory, the gain of an inverted Vee is similar to that of a dipole at the same elevation because most of the radiation is from the high-current portion of the antenna, which is near the center. Since the center of both antennas is the same height, there is little difference in performance. Antenna modeling software predicts maximum gain of 2.15 dBi for the dipole and 1.9 dBi for the inverted Vee.

However, in practice, ground proximity and ground conductivity as well as end effects reduce the efficiency of the inverted Vee considerably compared to the dipole. In the 40-foot example above, considering a useful take-off angle of 40 degrees above the horizon, the inverted Vee produces a maximum gain of 1 dBi in a circular pattern, whereas the dipole produces an oval pattern ranging from 6 dBi toward the sides down to 1.2 dBi toward the ends.
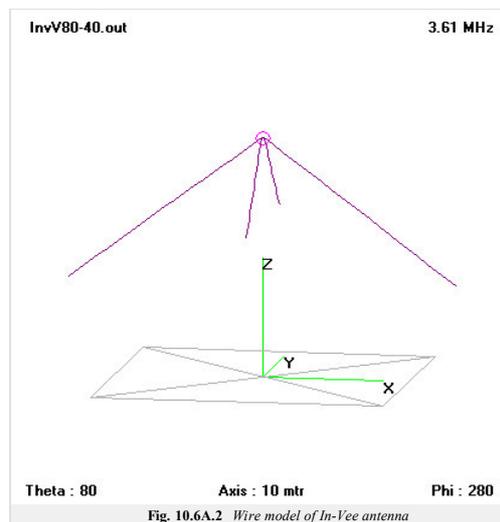
Fig. 10.6A.2 *Wire model of In-Vee antenna*
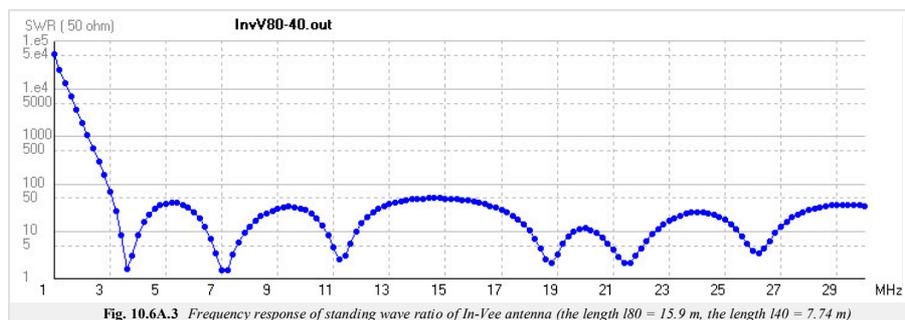

Fig. 10.6A.3 *Frequency response of standing wave ratio of In-Vee antenna (the length l80 = 15.9 m, the length l40 = 7.74 m)*
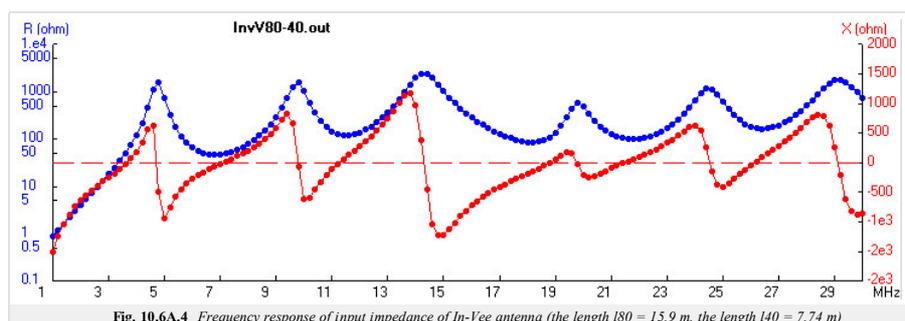

Fig. 10.6A.4 *Frequency response of input impedance of In-Vee antenna (the length l80 = 15.9 m, the length l40 = 7.74 m)*
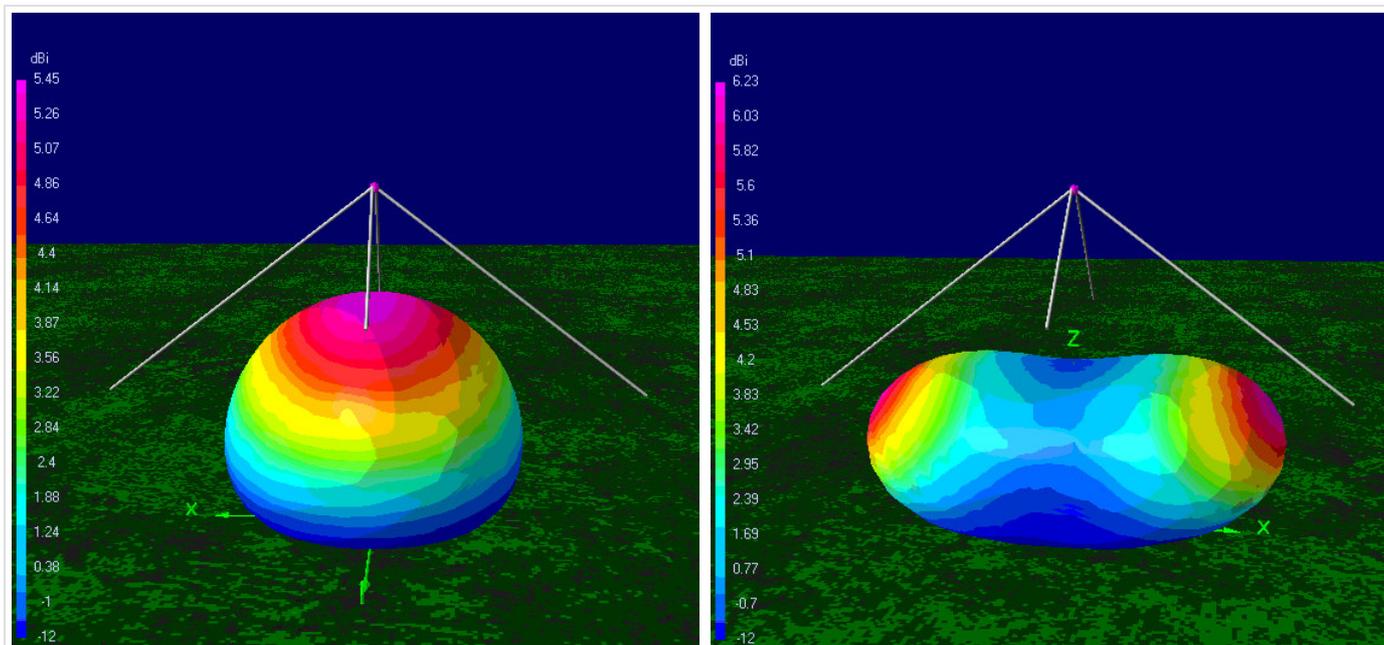
Fig. 10.6A.5 *Radiation pattern of In-Vee antenna at 3.5 MHz and 7.05 MHz (the length l80 = 15.9 m, the length l40 = 7.74 m)*

10.6 Inverted Vee

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

One of most adventages of Inverted-Vee antenna is...

Possible answers for question 1:

- ○ … low length of antenna against half-wave dipole.
- ○ … wideband of antenna.
- ○ … low weight of antenna.
- ● I won't answer this question

Question 2

Inverted-Vee antenna can be majority used for...

Possible answers for question 2:

- ○ … satellite communication.
- ○ … ionosphere propagation.
- ○ … direct wave propagation.
- ● I won't answer this question

Question 3

Input impedance of Inverted-Vee antenna is about...

Possible answers for question 3:

- ○ … 50 Ω and can be loaded by coaxial cable.
- ○ … 50 Ω and can be loaded by symmetrical feed-line.
- ○ … 300 Ω and must be loaded across symmetrical balun.
- ● I won't answer this question

[see result]

10.7 Yagi antenna

Basic theory

A Yagi-Uda Antenna, commonly known simply as a Yagi antenna or Yagi, is a directional antenna system consisting of an array of a dipole and additional closely coupled parasitic elements (usually a reflector and one or more directors). The dipole in the array is driven, and another element, typically 5% longer, effectively operates as a reflector. Other parasitic elements shorter than the dipole may be added in front of the dipole and are referred to as directors. This arrangement gives the antenna increased directionality compared to a single dipole. Directional antennas, such as the Yagi-Uda, are also commonly referred to as beam antennas [46] or high-gain antennas (particularly for transmitting). Many common television antennas are Yagi antennas with added corner reflectors.

Yagi-Uda antennas are directional along the axis perpendicular to the dipole in the plane of the elements, from the reflector through the driven element and out via the director(s). Typically, all elements are arranged at approximately a one-quarter-wavelength mutual spacing. All elements usually lie in the same plane,

Fig. 10.7A.1 *Practical implementation of Yagi antenna*

supported on a single boom or crossbar; however, they do not have to assume this coplanar arrangement. For example, some commercially available Yagi-Uda antennas for television reception have several reflectors arranged to form a corner reflector behind the dipole.

The bandwidth of a Yagi-Uda antenna, which is usually defined as the frequency range for which the antenna provides a good match to the transmission line to which it is attached, is determined by the length, diameter and spacing of the elements. For most designs, bandwidth is low, typically only a few percent of the design frequency.

Yagi-Uda antennas can be designed to operate on multiple bands. Such designs are more complicated, using pairs of resonant parallel coil and capacitor combinations (called a "trap" or LC) in the elements. The trap serves to isolate the outer portion of an element from the inner portion at the trap design frequency. In practice, the higher frequency traps are located closest to the boom of the antenna. Typically, a triband beam will have two pairs of traps per element. For example, a triband design covering the 10, 15 and 20 meter bands would have traps for the 10 and 15 meter bands. The use of traps is not without cost, as they reduce the bandwidth of the antenna on each band and reduce its overall efficiency.

There are no simple formulas for designing Yagi-Uda antennas due to the non-linear relationships between physical parameters such as element length, diameter and position and electrical characteristics such as input impedance and gain. Consequently, designs are found experimentally either through direct measurement or computer simulation, or by modifying existing designs.

Yagi-Uda antennas are widely used by amateur radio operators worldwide for communication on frequencies from shortwave, through VHF/UHF, and into microwave bands. Hams often homebrew this type of antenna, and have provided many technical papers and software to the engineering community.

Fig. 10.7A.2 *Wire model of Yagi antenna*

Fig. 10.7A.3 *VSWR of Yagi antenna (reflector = 10.534 m, dipole = 10.06 m, director = 9.464 m, spacing dipole-reflector = 3.187 m and dipole-director = 3.6847 m, h = 21 m)*

Fig. 10.7A.4 *Frequency response of input impedance of Yagi antenna (reflector = 10.534 m, dipole = 10.06 m, director = 9.464 m, spacing dipole-reflector = 3.187 m and dipole-director = 3.6847 m, h = 21 m)*
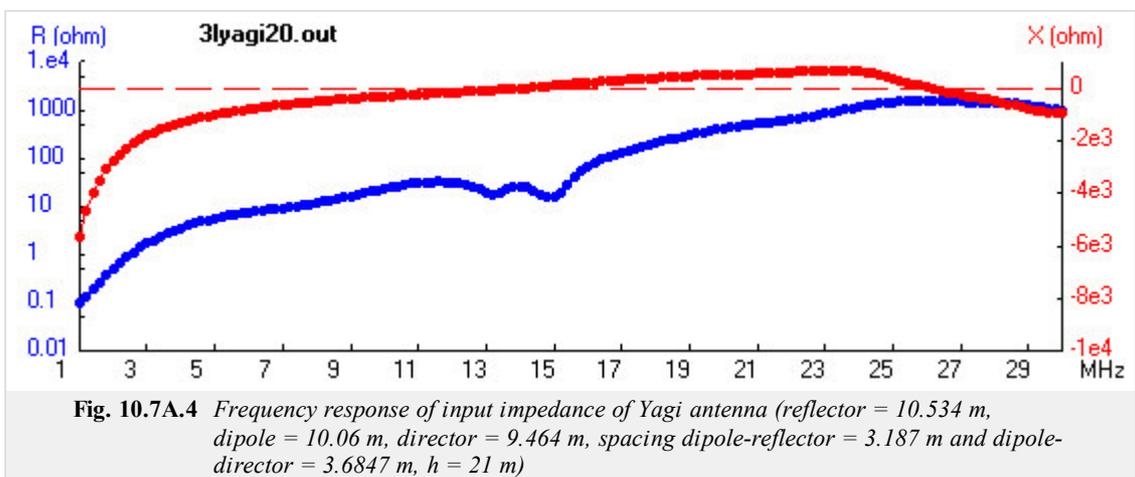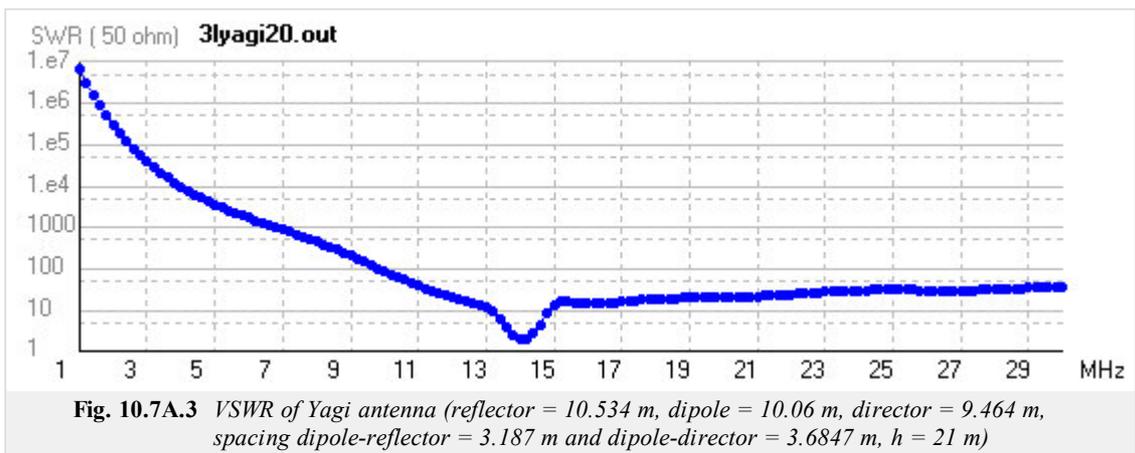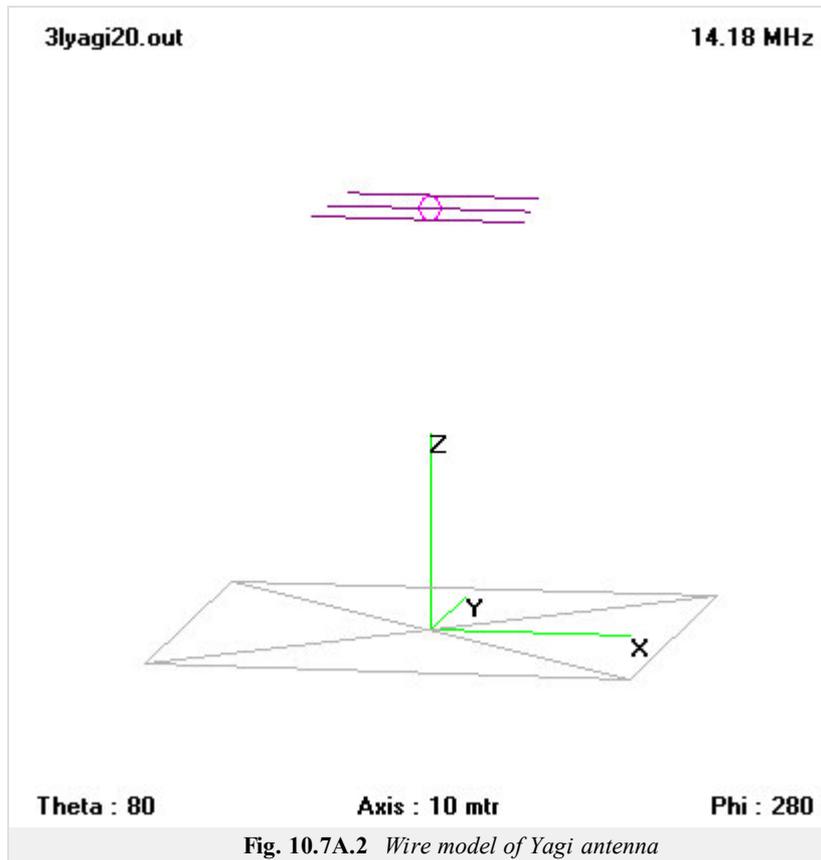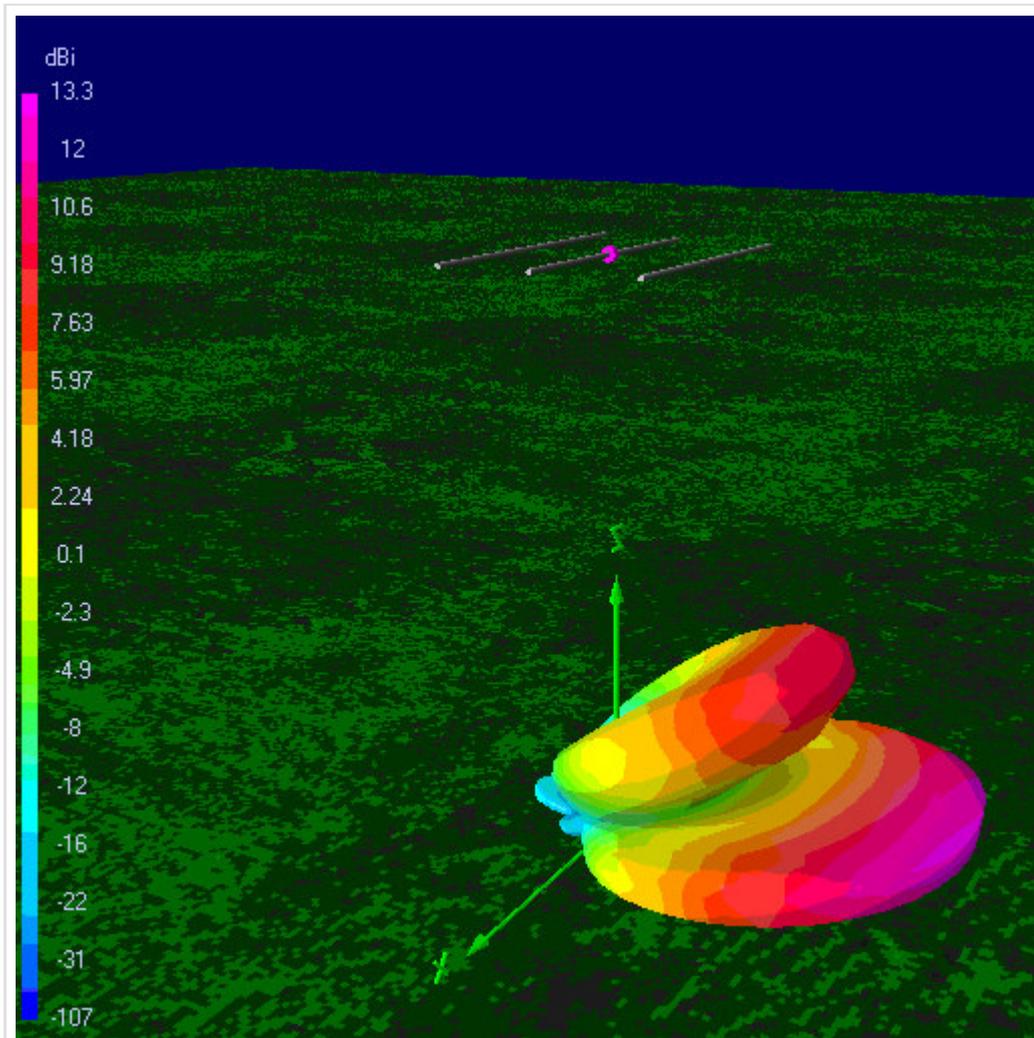
Fig. 10.7A.5 *Directivity pattern of Yagi antenna on 14.05 MHz (reflector = 10.534 m, dipole = 10.06 m, director = 9.464 m, spacing dipole-reflector = 3.187 m and dipole-director = 3.6847 m, h = 21 m)*

10.7 Yagi antenna

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

Yagi antenna is antenna with...

┌─ Possible answers for question 1: ─────────────────────────────────────┐
| ○ … circular polarization. |
| ○ … with medium gain of radiation. |
| ○ … omni-directional radiation. |
| ◉ I won't answer this question |
└──┘

Question 2

Yagi antenna can be built from...

┌─ Possible answers for question 2: ─────────────────────────────────────┐
| ○ … several active elements, one reflector and one director. |
| ○ … several directors, several active elements and several reflectors. |
| ○ … one active element, one reflector and one director. |
| ◉ I won't answer this question |
└──┘

Question 3

Wideband of Yagi antenna is usually...

┌─ Possible answers for question 3: ─────────────────────────────────────┐
| ○ … 1:10. |
| ○ … 1:5. |
| ○ … few percents. |
| ◉ I won't answer this question |
└──┘

see result

10.8 Logarithmic-periodical antenna (LPA)

Basic theory

In telecommunication, a log-periodic antenna (LP, also known as a log-periodic array) is a broadband, multielement, unidirectional, narrow-beam antenna that has impedance and radiation characteristics that are regularly repetitive as a logarithmic function of the excitation frequency. The individual components are often dipoles, as in a log-periodic dipole array (LPDA). Log-periodic antennas are designed to be self-similar and are thus also fractal antenna arrays.

It is normal to drive alternating elements with 180° (π radians) of phase shift from one another. This is normally done by connecting individual elements to alternating wires of a balanced transmission line.

The length and spacing of the elements of a log-periodic antenna increase logarithmically from one end to the other. A plot of the input impedance as a function of logarithm of the excitation frequency shows a periodic variation.
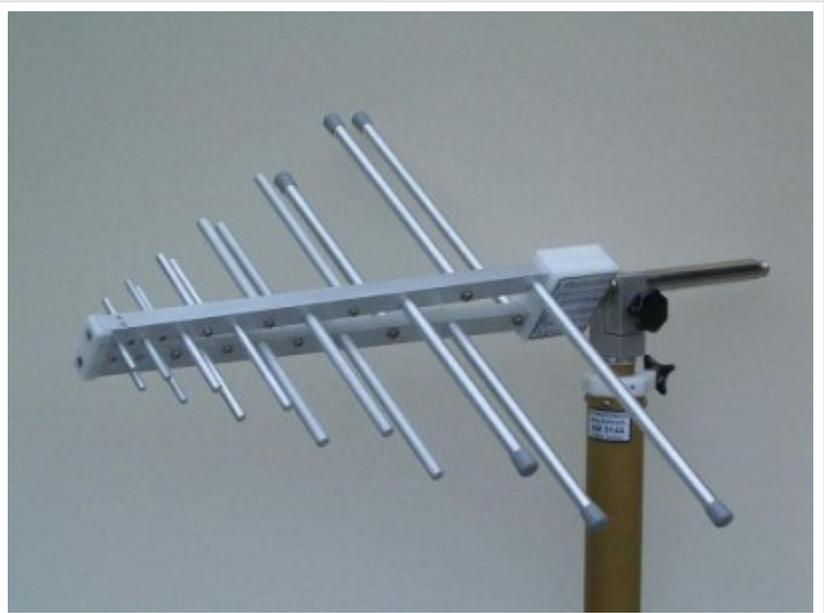
Fig. 10.8A.1 *Practical implementation of LPA antenna*

This antenna design is used where a wide range of frequencies is needed while still having moderate gain and directionality. It is sometimes used for a (VHF/UHF) television antenna.

Fig. 10.8A.2 *Wire model of LPA antenna*
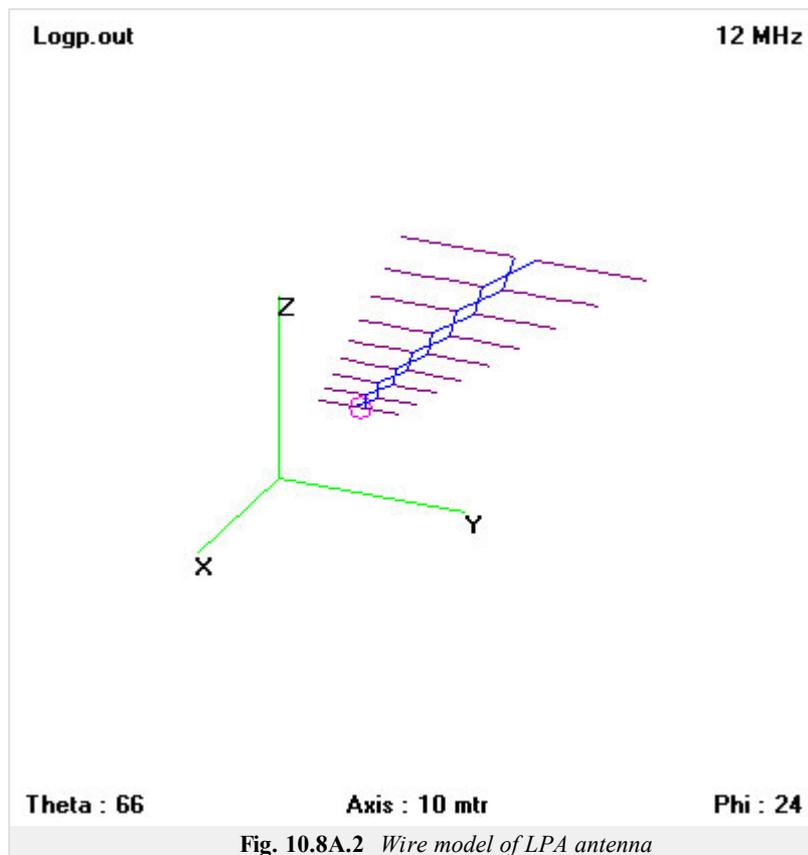
Fig. 10.8A.3 *Frequency response of standing wave ratio of LPA antenna (length d1 = 2.142 m, d2 = 2.463 m, spacing s1 = 1.143 m, d3 = 2.833 m, s2 = 2.803 m, d4 = 3.255 m, s3 = 4.713 m, d5 = 3.741 m, s4 = 6.903 m, d6 = 4.299 m, s5 = 9.833 m, d7 = 4.941 m, s6 = 12.313 m, d8 = 5.682 m, s7 = 15.633 m, d9 = 6.531 m, s8 = 19.463 m)*
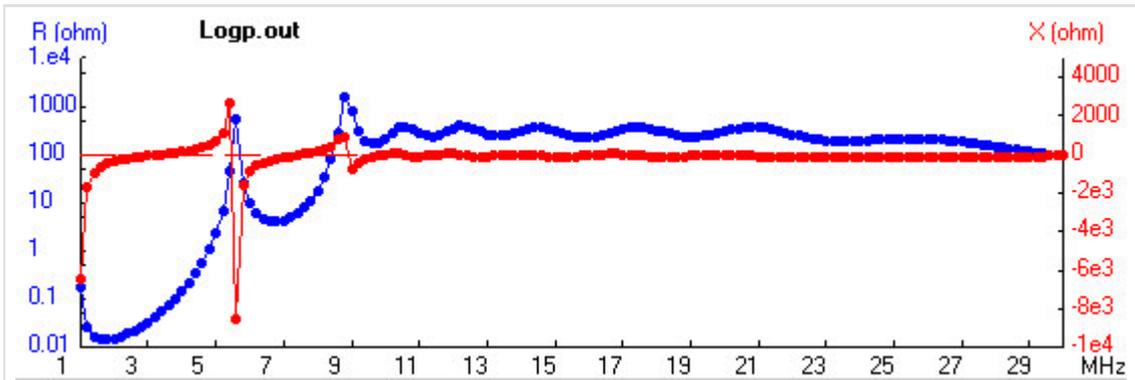
Fig. 10.8A.4 *Frequency response of input impedance of LPA antenna (length d1 = 2.142 m, d2 = 2.463 m, spacing s1 = 1.143 m, d3 = 2.833 m, s2 = 2.803 m, d4 = 3.255 m, s3 = 4.713 m, d5 = 3.741 m, s4 = 6.903 m, d6 = 4.299 m, s5 = 9.833 m, d7 = 4.941 m, s6 = 12.313 m, d8 = 5.682 m, s7 = 15.633 m, d9 = 6.531 m, s8 = 19.463 m)*
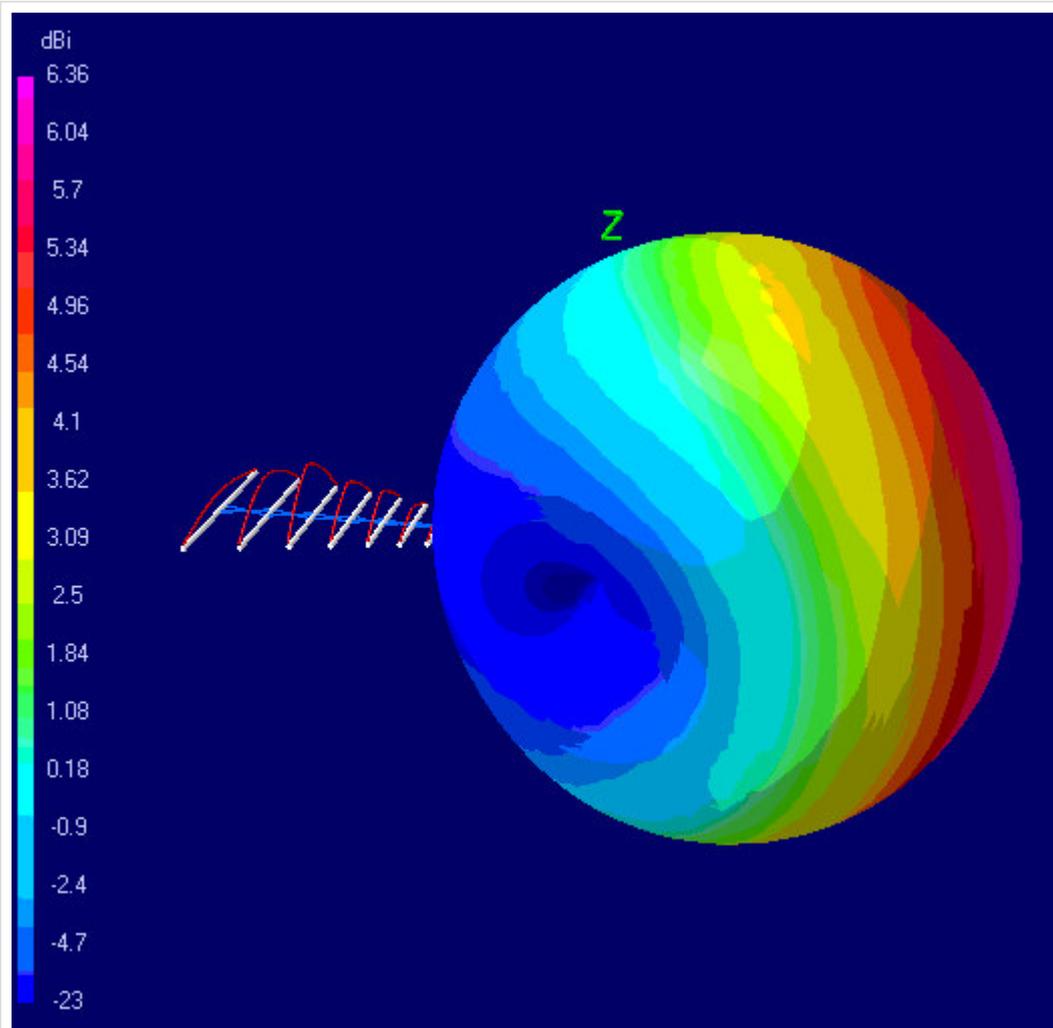
Fig. 10.8A.5 *Radiation pattern of LPA antenna at 14 MHz (length d1 = 2.142 m, d2 = 2.463 m, spacing s1 = 1.143 m, d3 = 2.833 m, s2 = 2.803 m, d4 = 3.255 m, s3 = 4.713 m, d5 = 3.741 m, s4 = 6.903 m, d6 = 4.299 m, s5 = 9.833 m, d7 = 4.941 m, s6 = 12.313 m, d8 = 5.682 m, s7 = 15.633 m, d9 = 6.531 m, s8 = 19.463 m)*
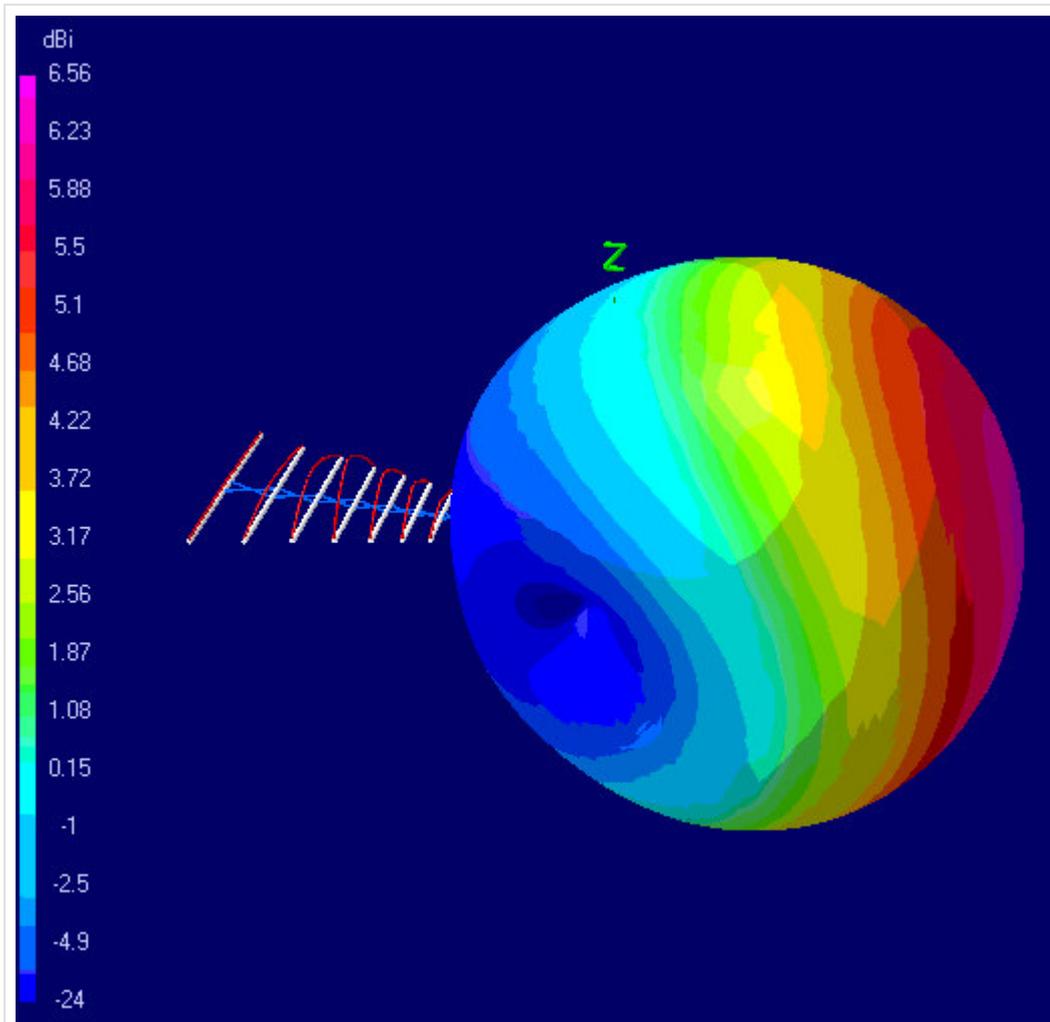
Fig. 10.8A.6 *Radiation pattern of LPA antenna at 21 MHz (length d1 = 2.142 m, d2 = 2.463 m, spacing s1 = 1.143 m, d3 = 2.833 m, s2 = 2.803 m, d4 = 3.255 m, s3 = 4.713 m, d5 = 3.741 m, s4 = 6.903 m, d6 = 4.299 m, s5 = 9.833 m, d7 = 4.941 m, s6 = 12.313 m, d8 = 5.682 m, s7 = 15.633 m, d9 = 6.531 m, s8 = 19.463 m)*

Fig. 10.8A.7 *Radiation pattern of LPA antenna at 28 MHz (length d1 = 2.142 m, d2 = 2.463 m, spacing s1 = 1.143 m, d3 = 2.833 m, s2 = 2.803 m, d4 = 3.255 m, s3 = 4.713 m, d5 = 3.741 m, s4 = 6.903 m, d6 = 4.299 m, s5 = 9.833 m, d7 = 4.941 m, s6 = 12.313 m, d8 = 5.682 m, s7 = 15.633 m, d9 = 6.531 m, s8 = 19.463 m)*

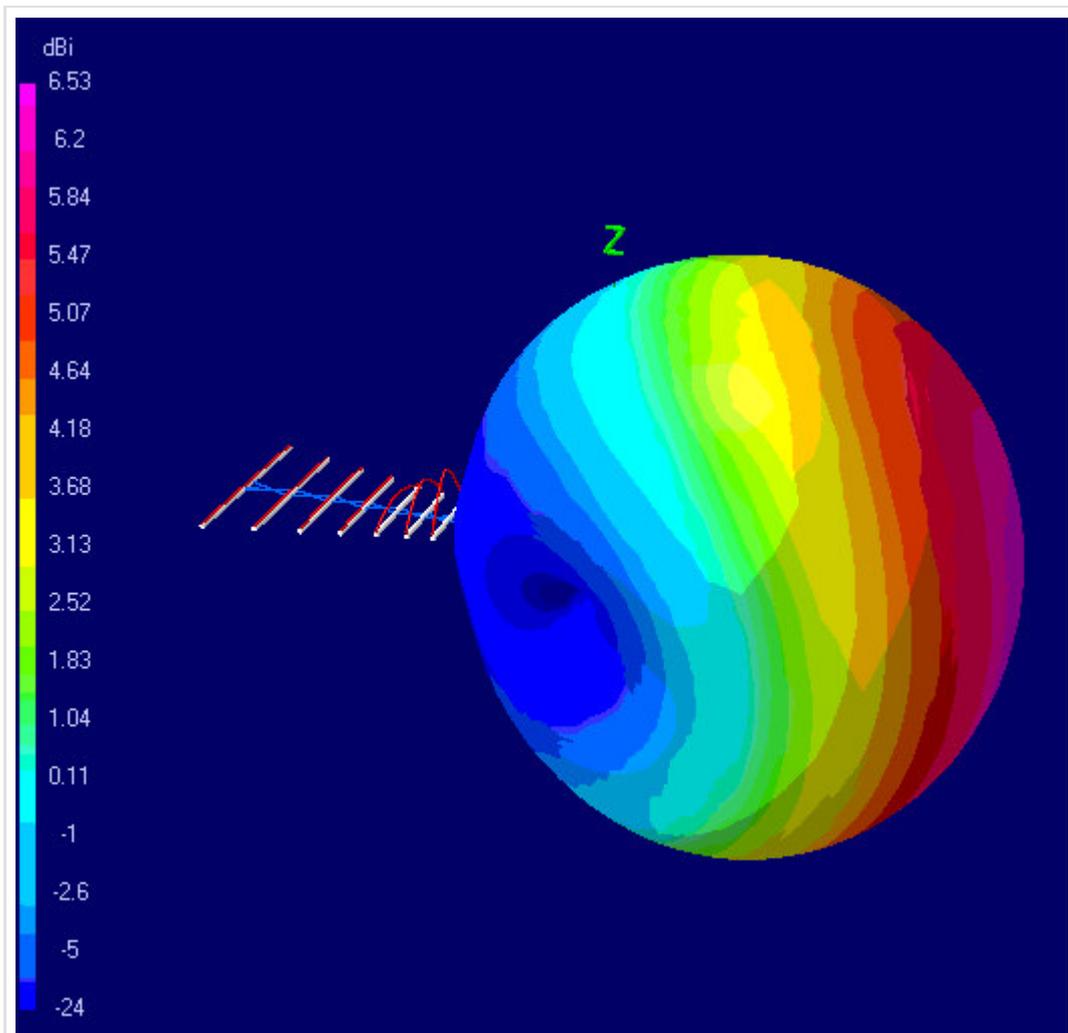# 10.8 Logarithmic-periodical antenna (LPA)

Quiz

Answer these questions to get feedback on how well you understand the course. Only one of the answers is correct. You don't have to answer every question. If you don't know the answer you can just leave it blank (default option: "I won't answer this question") and this won't affect your score. Answering **correctly** will **add 2 points** to your score but on the other hand you'll **lose 1 point** if your answer is **wrong**. The questions are divided in groups of five questions.

Press **See result** after you have finished answering.

Displaying questions **1**..**3** of **3**:

Question 1

LPA antenna is antenna...

Possible answers for question 1:
- ⦿ … with circular polarization.
- ⦿ … with medium value of gain.
- ⦿ … with omni-directional propagation.
- ⦾ I won't answer this question

Question 2

Two neighbour elements are feeded by phase shift...

Possible answers for question 2:
- ⦿ … 180°.
- ⦿ … 225°.
- ⦿ … 0°.
- ⦾ I won't answer this question

Question 3

Wideband of LPA antenna is usually...

Possible answers for question 3:
- ⦿ … 1:10.
- ⦿ … 1:5.
- ⦿ … few percents.
- ⦾ I won't answer this question

[see result]

Huygens principle

Mathematical expression of Huygens principle, as usually presented in classical textbooks, is not of a complete form: its exploitation is limited to the case when the intensity of the total magnetic field **H** is directly related to the intensity of the total electric field **E** in every point of the space (i.e., we can write $\mathbf{H} = \mathbf{E} / Z$). The described situation appears at the open end of a waveguide where intensity of electric field is related (with sufficient accuracy) to the intensity of magnetic field according to the following expressions.

$$\mathbf{E}^{tot} = \mathbf{E}^{inc}\left(1 + \rho_0\right) \quad \mathbf{H}^{tot} = \mathbf{H}^{inc}\left(1 + \rho_0\right) \quad \mathbf{H}^{inc} = \frac{\mathbf{E}^{inc}}{Z_g}. \tag{1}$$

Here, the upper index *inc* denotes intensity of incident wave arriving to the open end of a waveguide and the upper index *tot* denotes intensity of the total field (summation of the incident wave and the reflected one). Further, ρ_0 is reflection coefficient and

$$Z_g = \frac{Z}{\sqrt{1 - \left(\dfrac{\lambda_0}{\lambda_m \sqrt{\varepsilon_r}}\right)^2}} \tag{2}$$

is impedance of the mode TE$_{10}$ in a rectangular waveguide; Z is characteristic impedance of the environment inside a waveguide, $\lambda_m = 2a$ is critical wavelength for a waveguide without dielectrics and a is the dimension of a longer side of a waveguide. Next, λ_0 denotes wavelength in the vacuum and ε_r is the relative permittivity in the waveguide. Considering the mode TE$_{10}$, only the component $\mathbf{E}_y$ of the intensity of electric field, and only the component $\mathbf{H}_x$ of the intensity of magnetic field are non-zero..

Substituting $\mathbf{E}^{tot}$ and $\mathbf{E}^{tot}/Z_g$ the complete relation of Huygens principle, we obtain a well-known, simplified expression of Huygens principle (relation for re-computing magnitude of the intensity of electric field on an elementary surface $E^{(S)}$ to an arbitrary point, which is located in the far-field region)

$$E^{(P)} = \frac{j}{\lambda} \int\limits_{-a/2}^{+a/2} \int\limits_{-b/2}^{+b/2} E^{(S)} \cos(n,\, r_2) \frac{\exp(-jkr)}{r_2} \, dy \, dx. \tag{3}$$

Here, a and b denote dimensions of an elementary irradiated surface, λ is wavelength in a given environment, $E^{(S)}$ is magnitude of the intensity of electric field on an elementary surface, k denotes wavenumber, n is a vector perpendicular to an elementary surface and r_2 is the position vector between an elementary surface S and an observation point P.

The complete relation can be derived from Maxwell equations for an elementary surface, where electric current density and magnetic current density are known. For the far-field region, we get

$$E_{\vartheta}^{S} = -j\frac{k}{4\pi}\left(\eta\,\mathbf{J} \cdot \mathbf{T}_{xyz\vartheta} + \mathbf{K} \cdot \mathbf{T}_{xyz\varphi}\right)\frac{\exp(-jkr)}{r}, \tag{4a}$$

$$E_{\varphi}^{S} = -j\frac{k}{4\pi}\left(\eta\,\mathbf{J} \cdot \mathbf{T}_{xyz\varphi} + \mathbf{K} \cdot \mathbf{T}_{xyz\vartheta}\right)\frac{\exp(-jkr)}{r}. \tag{4b}$$
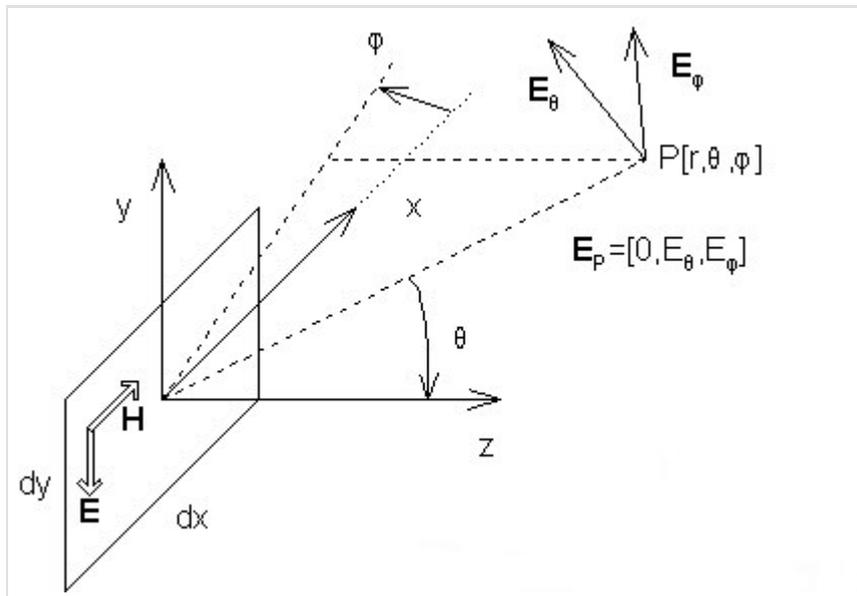
Eqn. (4a,b) is explained in fig. **1**. Here, an elementary surface of dimensions dx and dy, where electric current **J** and magnetic current **K** are flowing, is depicted. Electric intensity of scattered field E^s, i.e. intensity created by the irradiated elementary surface in the point **r**, is expressed in spherical coordinates. Next, current densities **J** and **K** are supposed being known in a Cartesian coordinate system. Relations for E_ϑ^s and E_φ^s contain vectors **T**, which perform transformation from Cartesian coordinate system to spherical one:

$$\mathbf{T}_{xyz\varphi} = \begin{bmatrix} -\sin(\varphi) & \cos(\varphi) & 0 \end{bmatrix}, \tag{5a}$$

$$\mathbf{T}_{xyz\vartheta} = \begin{bmatrix} -\cos(\vartheta)\cos(\varphi) & \cos(\vartheta)\sin(\varphi) & -\sin(\vartheta) \end{bmatrix}. \tag{5b}$$

Currents **J** and **K** do not need to physically flow on an elementary surface (we speak about so called equivalent currents, which are induced by electric field and magnetic one on the elementary surface). These currents can be evaluated using

$$\mathbf{J} = \mathbf{n} \times \mathbf{H}, \tag{6a}$$

$$\mathbf{K} = -\mathbf{n} \times \mathbf{E}. \tag{6b}$$

Here, **n** is perpendicular to the elementary surface (identical with the unitary vector $\mathbf{u}_z$).

Using eqns. (6a,b), current densities flowing on electric and magnetic walls can be computed in a real case. Unfortunately, an examined irradiated surface does not need being neither electric nor magnetic one.

Direct application of the described principle is conditioned by the requirement that no other objects are present in a half-space defined by the elementary surface and the positive direction of vector **z**. If this condition is not met, Huygens approach and principle of equivalent currents can be applied, but the situation is complicated by reflections from those objects which further influences irradiation of the elementary surface. This case is not analyzed here.

Thanks to the fact that vector products $\mathbf{n} \times \mathbf{E}$, $\mathbf{n} \times \mathbf{H}$ (**n** is perpendicular to the surface) are computed when evaluating equivalent currents, only those components of currents **J** and **K** have to be known, which are tangential to the elementary surface.

The last symbol, which has not been explained yet, is free-space characteristic impedance η.

That is the end of a brief note on the complete form of Huygens principle. In the following paragraphs, we show the simplification of the complete form in the case when electric intensity is known on an elementary surface and magnetic intensity is related to it via impedance (this approach can be used for an open end of a waveguide but it cannot be applied to a surface of a microstrip antenna because its surface impedance is not constant).

Before performing the described simplification, let us turn our attention to the signs in eqns. (4a,b). Note that the right-hand sides of those relations start by the sign minus. The sign minus in (4a) is changed to plus if **H** is expressed as $-\mathbf{E}/Z_g$ (see direction of **E** and **H** in fig. **1**). This is further explained by the following example.

Assume an elementary electric dipole with current density **J**. The current is expected to flow in the direction x. Considering such component of **E**, which is oriented in parallel with the current **J**, then the relation $\mathbf{E} \sim +j\mathbf{J}$ has to be valid in the far-field region. Fig. **2** illustrates the situation.
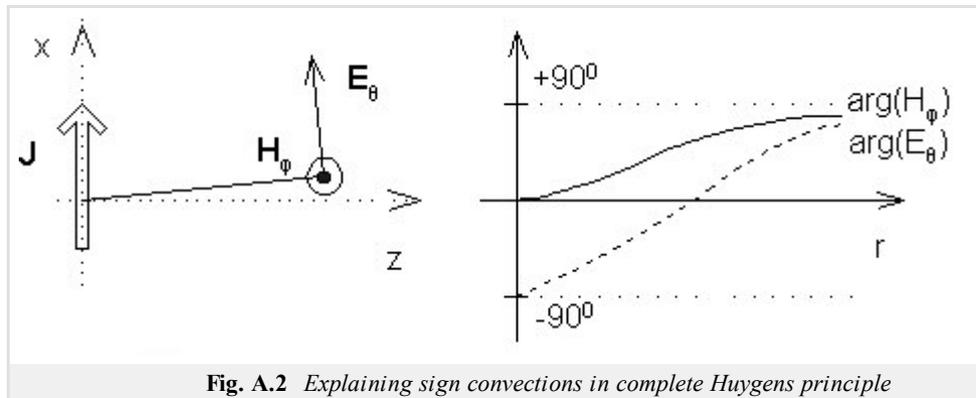
Fig. A.2 *Explaining sign convections in complete Huygens principle*

Now, let us investigate the situation of magnetic intensity and the behavior of **E**, **H** in the near-field zone of an elementary dipole. In fig. **2**, both the vector **E** and the vector **H** are depicted. Complex amplitude of components E_ϑ and H_φ depend on the distance r by the following way:

$$E_\vartheta \approx \frac{j}{kr} + \frac{1}{(kr)^2} - \frac{j}{(kr)^3}, \; ,$$

$$H_\varphi \approx \frac{j}{kr} + \frac{1}{(kr)^2}.$$

Magnetic field on the surface of the dipole (thanks to the small distance r, the term $1/k^2 r^2$ is dominant) is of the same phase than the current. On the contrary, phase of electric field (the term $-j/k^3 r^3$ is dominant) is shifted for 90°. In the far-field zone, (thanks to the large distance r, the terms j/kr are dominant) phase of both components is identical (90°). Phase course of E_ϑ and H_φ is depicted in fig. **2**.
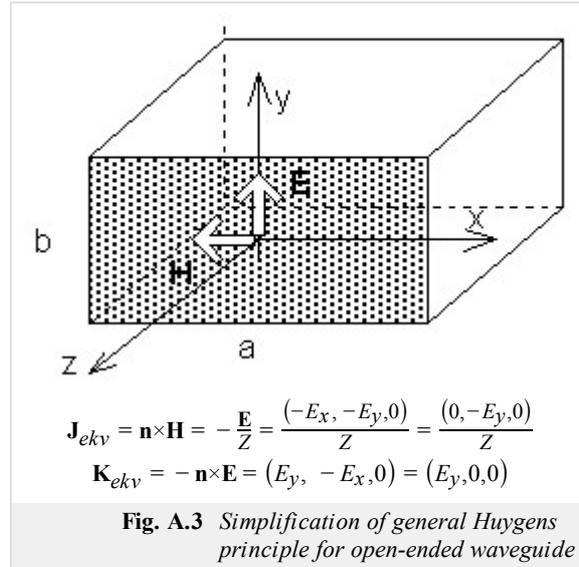
Dealing with the near-field zone of an elementary dipole - the zone is given by the distance $\lambda_0/(2\pi)$, where the amplitude of the longitudinal component E_r and the amplitude of the transversal one E_j are of the same magnitude.

Now, let us concentrate on the derivation of the simplified Huygens principle (relations for E_ϑ and E_φ).

Assume an open end of a waveguide (fig. **3**). In the waveguide, mode TE_{10} propagates. The electric intensity of the total field at the end of the waveguide is denoted by **E**, the magnetic intensity by **H**. In fig. **3**, even relations among intensities **E** and **H** and current densities **J** and **K** are given.

$$\mathbf{J}_{ekv} = \mathbf{n} \times \mathbf{H} = -\frac{\mathbf{E}}{Z} = \frac{(-E_x, -E_y, 0)}{Z} = \frac{(0, -E_y, 0)}{Z}$$

$$\mathbf{K}_{ekv} = -\mathbf{n} \times \mathbf{E} = (E_y, -E_x, 0) = (E_y, 0, 0)$$

Fig. A.3 *Simplification of general Huygens principle for open-ended waveguide*

Let us substitute for **J** and **K** to the complete relation for E_ϑ and E_φ, and assume this moment that **E** and **H** are mutually perpendicular and their components x and y are non-zero. At the end of the derivation, this assumption will be cancelled.

The described substituting and integrating waves from all the elementary surfaces yield

$$E_\vartheta^S = -j\frac{k}{4\pi}\frac{\exp(-jkr)}{r} \iint_S \left\{ \eta\begin{bmatrix} -E_x & -E_y \end{bmatrix}\begin{bmatrix} \cos(\vartheta)\cos(\varphi) \\ \cos(\vartheta)\sin(\varphi) \end{bmatrix} + \begin{bmatrix} E_y & -E_x \end{bmatrix}\begin{bmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{bmatrix} \right\} dS, \quad (7a)$$

$$E_\vartheta^S = \frac{j}{2\lambda}\frac{\exp(-jkr)}{r}\left[1 + \frac{\eta}{Z}\cos(\vartheta)\right]\left[N_x\cos(\varphi) + N_y\sin(\varphi)\right], \quad (7b)$$

Here, N_x, N_y are double integrals, which form own radiation pattern of the open end of the waveguide

$$N_x = \int_S E_x \exp(-jk\Delta r)dS,$$

$$N_y = \int_S E_y \exp(-jk\Delta r)dS.$$

Analogically for the component E_φ

$$E_\varphi^S = -j\frac{k}{4\pi}\frac{\exp(-jkr)}{r} \iint_S \left\{ \eta\begin{bmatrix} -E_x & -E_y \end{bmatrix}\begin{bmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{bmatrix} + \begin{bmatrix} -E_y & E_x \end{bmatrix}\begin{bmatrix} \cos(\vartheta)\cos(\varphi) \\ \cos(\vartheta)\sin(\varphi) \end{bmatrix} \right\} dS, \quad (8a)$$

$$E_\varphi^S = \frac{j}{2\lambda}\frac{\exp(-jkr)}{r}\left[1 + \frac{\eta}{Z}\cos(\vartheta)\right]\left[N_x\sin(\varphi) - N_y\cos(\varphi)\right]. \quad (8b)$$

When deriving eqns. (8a,b), the observation point was expected in the far-field region. We could therefore take the exponential term out of the integral and we integrated space shifts Δr only.

In the case of the open end of the waveguide, the situation becomes even simpler because the component E_x is zero, and therefore even N_x is zero.

And finally, two concluding notes:

1. Simplified Huygens principle expects that intensity of incident field only $\mathbf{E}^{inc}$ (or $\mathbf{H}^{inc}$) is substituted into it, and that the intensity of the total field is approximately twice higher (the total field is exactly twice higher for perfectly conducting wall and approximately twice higher for the open end of the waveguide, e.g.).

2. Simplified Huygens principle contains the term $\cos(\vartheta)$ instead of the correct $(1 + \eta \cos(\vartheta)/Z$. Therefore, the simplified principle introduces an approximation of the real directivity pattern. This approximation can be used when beam-width is going to be determined, but it is infeasible when side-lobe level is investigated.

Matlab examples

Chapter 2.: **EM wave propagation**

- Fresnel diffraction: fresnel.zip
- General diffraction: gen_diffra.zip
- Layered media: layers.zip

Chapter 3.: **Transmission lines**

- Waveguides: waveguide.zip
- Microstrips: microstrip.zip

Chapter 4.: **Antennas**

- Wire dipole: dipol.zip
- Mutual impedance: impedace.zip
- Yagi antenna: yagi.zip
- Microstrip dipole: m_dipol.zip
- Microstrip antenna: rpatch1.m
- Multiband patch antenna: patchant.m
- Planar ultra wideband antenna: U_antena.m

Chapter 5.: **Time-domain modeling of antennas**

- Drátový dipól: Implicit_EN.m

Chapter 6.: **Selective surfaces**

- Analysis of FSS: FSS_en.zip

Chapter 7.: **Radio optics**

- Gaussian beam: gauss.zip

Java applets

Here are *.ZIP- & *.RAR- files to the source code of the applets for the people who understand the JAVA programming language.

Chapter 4.: **Antennas**

Wire dipole:	ZIP - RAR
Yagi antenna:	ZIP - RAR
Microstrip dipole:	ZIP - RAR

References

[1] ČERNOHORSKÝ, D., NOVÁČEK, Z., RAIDA, Z. *Elektromagnetické vlny a vedení*. 2. rozšířené a přepracované vydání. Brno: Nakladatelství VUTIUM, 1999. ISBN 80-214-1261-5.

[2] JORDAN, E.C., BALMAIN, K. G. *Electromagnetic waves and radiating systems*. 2nd edition. Englewood Cliffs: Prentice Hall, 1968.

[3] ČERNOHORSKÝ, D., NOVÁČEK, Z., RAIDA, Z. *Elektromagnetické vlny a vedení: příklady pro cvičení a domácí projekty*. Dotisk 2. vydání. Brno: FEI VUT v Brně, 2001. ISBN 80-214-1254-2.

[4] REKTORYS, K. a kol. *Přehled užité matematiky I, II*. Dotisk 6. vydání. Praha: Prometheus, 1995. ISBN 80-858-4972-0.

[5] HARRINGTON, R. F. *Field computation by moment methods*. Piscataway: IEEE Press, 1993.

[6] AJZENBERG, G. Z. *Antěny ultrakorotkych voln*. Moskva: Svjazizdat, 1957 (v ruštině).

[7] HUDÁK, I. *Analýza drátových antén momentovou metodou*. Diploma thesis. Brno: FEI VUT v Brně, 1996.

[8] SAINATI, R. A. *CAD of microstrip antennas for wireless applications*. Norwood: Artech House, 1996.

[9] HARRINGTON, R. F. *Field computation by moment methods*. 2nd ed. Piscataway: IEEE Press, 1993.

[10] MOSIG, J. R., GARDIOL, F. E. Analytical and numerical techniques in the Green's function treatment of microstrip antennas and scatterers. *IEE Proceedings H*. 1982, vol. 130, no. 2, p. 172 - 182.

[11] ALATAN, L., AKSUN, M. I., LEBLEBICIOGLU, K., BIRAND, M. T. Use of computationally efficient method of moments in the optimization of printed antennas. *IEEE Transactions on Antennas and Propagation*. 1999, vol. 47, no. 4, p. 725 - 731.

[12] MOSIG, J. R., GARDIOL, F. E. General integral equation formulation for microstrip antennas and scatterers. *IEE Proceedings H*. 1985, vol. 132, no. 7, p. 424 - 432.

[13] MOSIG, J. R., GARDIOL, F. E. A dynamical radiation model for microstrip structures. In HAWKES, P. *Advances in Electronics and Electron Physics*. New York: Academic Press, 1982, p. 139 - 237.

[14] POZAR, D.M., SCHAUBERT, D.H. *Microstrip Antennas*. Piscataway: IEEE Press. 1995

[15] ČERNOHORSKÝ, D., RAIDA, Z., ŠKVOR, Z., NOVÁČEK, Z. *Analýza a optimalizace mikrovlnných struktur*. Brno: VUTIUM Publishing. 1999.

[16] SALEH, B.E.A., TEICH, M.C. *Základy fotoniky I*. Praha: Matfyzpress, 1994

[17] MUNK, B.A. *Frequency Selective Surfaces: Theory and Design*. Chichester: John Wiley and Sons, 2000.

[18] WU, T.K. *Frequency Selective Surfaces and Grid Arrays*. Chichester: John Willey and Sons, 1995.

[19] SCOTT, C. *The Spectral Domain Method in Electromagnetics*. Norwood: Artech House, 1989.

[20] GUPTA, K. C., GARG, R., BAHL, I., BHARTIA, P. Microstrip lines and slotlines. 2nd ed. Norwood: Artech House, 1996.

[21] SILVESTER, P. P., FERRARI, R. F. Finite elements for electrical engineers. 3rd ed. Cambridge: Cambridge University Press, 1996.

[22] Wilkinson, E. *An N-Way Hybrid Power Divider*. IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-8, January 1960, pp. 116-118.

[23] PARAD, L. I., Moynihan, R. L. *Split TEE Power Divider*. IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-13, January 1965, pp. 91-95.

[24] HOFFMANN, K. *Planární mikrovlnné obvody*. Skriptum ČVUT, Praha 2000.

[25] MAAS, S. *Microwave Cookbook*. In Czech Rep. is available at the Czech section library of IEEE.

[26] REED, J., Wheeler, G. J. *A Method of Analysis of Symetrical Four-Port Networks* IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-4, No. 4., October 1956, pp. 246-52

[27] HOFFMANN, K. *Nepublikované sdělení*.

[28] REHNMARK, S. *High Directivity CTL-Couplers and a New Technique for the Measurements of CTL-Coupler Parameters*. IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-25, No. 12., December 1977, pp. 1116-1121.

[29] CRISTAL, E. G., YOUNG, L. *Theory and Tables of Optimum Symmetrical TEM Mode Coupled Transmission Line Directional Couplers*. IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-13, No. 5., September 1965, pp. 544-8.

[30] ŠKVOR, Z. *CAD pro vf. techniku*. Skriptum ČVUT, Praha 1998.

[31] LEVY, R. *General Synthesis of Asymetric Multi-Element Coupled Transmission Line Directional Couplers*. IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-11, No. 4., July 1963, pp. 226-37.

[32] LEVY, R. *Tables for Asymetric Multi-Element Coupled Transmission Line Directional Couplers*. IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-12, No. 3., July 1964, pp. 275-79.

[33] WONG, K.-L. *Compact and broadband microstrip antennas*. John Wiley and Sons, Inc., 2002.

[34] VOLAKIS, J. L. *Antenna engineering handbook*. McGraw-Hill, 2007.

[35] MARKLEIN, R. *The finite integration technique as a general tool to compute acoustic, electromagnetic, elastodynamic, and coupled wave fields*. Available at: http://www.unikassel.de/fb16/tet/marklein/publikationen.html

[36] CLEMENS, M., WEINLAND, T. *Discrete electromagnetism with the finite integration technique*. Progress In Electromagnetics Research, PIER 32, 65–87, 2001.

[37] RAO, M. S., *Time Domain Electromagnetics*. London: Academic Press, 1999.

[38] ČERNOHORSKÝ, D., NOVÁČEK, Z. *Antény a šíření rádiových vln*. Skripta FEKT VUT, Brno 2003, ISBN 80-86056-47-3.

[39] GARG, R., BHARTIA, P., BAHL, I., ITTIPIBOON, A. *Microstrip antenna design handbook*. ARTECH HOUSE, INC., Boston, London 2001, ISBN 0-89006-513-6.

[40] *Ultra-Wideband (UWB) Technology*. Available at WWW: http://www.intel.com/technology/comms/uwb/

[41] CHEN, Z. N., CHIA, M. Y. W. *Broadband Planar Antennas: Design and Applications*. John Wiley and Son, Ltd. West Sussex, England 2006, ISBN 0-470-87174-1.

[42] WEIGAND, S., HUFF, G. H., PAN, K. H., BERNHARD, J. T. *Analysis and design of broad-band single-layer rectangular U-slot*

microstrip patch antennas. IEEE Transaction on Antennas and Propagation. 2003, vol. 51, no. 3, p. 457-468.

[43] MARŠÁLEK, R., *Teorie rádiové komunikace*. Elektronická skripta FEKT VUT, Brno 2005.

[44] SELBERHERR, S. *Analysis and Simulation of Semiconductor Devices*. Hiedelberg: Springer-Verlag, 1984.

[45] COUNTRYMAN, G. L. *An Experimental All-Band Nondirectional Transmitting Antenna*. W1RBK, (W3HH), QST, June 1949, page 54.

[46] CEBIK, L. B. *Modeling the T2FD*. W4RNL. Available at WWW: http://www.cebik.com/content/a10/wire/t2fd.html

Register

Ampér law

Ampér law expresses the fact that the summation of induced conductive current I_{ind}, of enforced conductive current I_{vn} and of displacement current $d\psi/dt$, which pass the surface bounded by the closed curve l in the positive direction, equals to the circulation of the magnetic field intensity vector along this curve.

$$\int_l \mathbf{H} \cdot d\mathbf{l} = I_{ind} + I_{vn} + \frac{\partial \psi}{\partial t}.$$

Circulation of vector $\mathbf{A}$ along the closed curve l can be imagined as a summation of scalar products of the vector $\mathbf{A}$ and a vector of elementary curve $d\mathbf{l}$ in the certain number of points on l. More points are considered, more accurate result is obtained.

Antenna efficiency, efficiency of antenna radiation

Antenna efficiency is defined as a ratio of a radiated power and a supplied one. Efficiency can be evaluated exploiting the following relation:

$$\eta = \frac{R_{\Sigma vst}}{R_{\Sigma vst} + R_{ztr}}$$

($R_{\Sigma vst}$ is radiation resistance related to input current, and R_{ztr} is lossy resistance related to input current). Good antenna efficiency is conditioned by a small lossy resistance and (or) by a high radiation resistance. On the contrary, low efficiency is exhibited by antennas of small radiation resistance, which is typical for antennas much shorter that the wavelength.

Computing radiation efficiency, a lossles antenna is assumed (lossy resistance is zero). Then radiation efficiency is given as a ratio of a real power radiated by the antenna (we integrate Poynting vector in the far-field region over the whole half-space) and a real power at the input terminal of the antenna.

Antenna gain

As the gain of the antenna, the maximum value of the directivity factor (in decibels) is understood

$$G_{abs} = 10\log(D_{max}).$$

This gain is called the *absolute one*. The *relative gain* tells us, how much is the maximum directivity factor higher that the directivity factor of the reference antenna:

$$G_{rel} = 10\log\left(\frac{D_{max}}{D_{ref}}\right).$$

As the reference antenna, the half-wavelength long dipole is usually considered.

Directivity factor $D(\varphi, \vartheta)$ describes the directivity of the antenna radiation: directivity factor is higher than one in those directions, where the radiation is concentrated, and is smaller than one in those directions, where the radiation is suppressed. Directivity factor of the isotropic source equals to one.

Aperture-fed microstrip antenna

Microstrip antenna elements are excited via slots in the ground plane. The feeding transmission line is etched on a substrate, which is positioned below the substrate with the antenna. Both the substrates share the common ground plane with a slot. The transmission line can be found on the *true side* of the lower substrate.

Basic mode, parasitic modes

Studying wave propagation, the mode of the lowest critical frequency is called the dominant (basic) mode. Studying wave propagation through periodic structures, the mode of the lowest spatial frequency is called the basic mode. The term mode denotes here a certain direction of the radiated wave, for which waves radiated from all the parts of the antenna reach the observation point with the same phase. In analogy to the antenna array, the basic mode can be understood as a wave propagating in the main lobe direction. Wave propagating in the direction, which is nearly perpendicular to the structure, is of the lowest spatial frequency (for perpendicular direction, spatial frequency equals to zero). In analogy to the antenna array, even the periodic structure can have a side-lobe (an intensive radiation to another direction, significantly declining the normal). This radiation is usually undesired (parasitic mode).

Basis function

Basis function forms a basis of approximating an unknown, sought function. Piecewise constant functions (functions are of unitary value over a single discretization element and are zero over the rest of elements) and piecewise linear functions are the most frequently used basis functions.

Beam

Beam is a curve, which tangent is of the same direction as the gradient of eiconale in every point. Beam is of the direction of the steepest phase variation and of the direction of Poynting vector, i.e. the direction of the energy flow.

Bessel function, Neumann function

Solving wave equation in cylindrical and spherical coordinates by the method of variables separation, Bessel differential equation is obtained for a single variable (radial coordinate)

$$y'' + \frac{1}{x} y' + \left(1 - \frac{v2}{x2}\right) y = 0.$$

If n is not an integer number, linear combination of Bessel functions of order n and $-n$ is the solution of the equation

$$y = C_1 J_v(x) + C_1' J_{-v}(x).$$

If n = n is an integer number, solutions $J_v(x)$ and $J_{-v}(x)$ are linearly dependent and Neumann function $N_n(x)$ has to be used as the second particular integral.

$$y = C_2 J_n(x) + C_2' N_n(x).$$

Biot-Savart law

The force between electric charges *in motion* is called the magnetic force. The force field, which is created by a moving charge, is called the magnetic field.

Magnetic field of the charge Q_1, which moves by the velocity $\mathbf{v}_1$, is quantified by magnetic displacement $\mathbf{B}$. Magnetic displacement is given by Biot-Savart law

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0 Q_1}{4\pi r^3} \mathbf{v}_1 \times \mathbf{r}$$

Magnetic displacement tells us the force, which is caused by the charge Q_1 moving by velocity $\mathbf{v}_1$ and which affects the unitary charge in $\mathbf{r}$ moving by unitary velocity. The symbol m0 denotes permeability of vacuum (m_0= 1.26 * 10^{-6} H * m^{-1}).

Boundary conditions

Boundary conditions are met by the electromagnetic field on the boundary of two different media. For the component of $\mathbf{E}$, which is tangential to the boundary, we write: $E_{tang\ 1} = E_{tang\ 2}$. Both the intensities are measured in points, which are located near the boundary on different sides. For normal components of electric displacement vector, we in analogy get: $D_{norm\ 1} = D_{norm\ 2}$. Dealing with magnetic field, tangential components of $\mathbf{H}$ and normal components of $\mathbf{B}$ have to meet similar conditions.

If one medium is perfectly conductive for electric currents or magnetic ones, boundary conditions can be divided to Dirichlet condition and to Neumann one.

Dirichlet condition expresses the fact that the component of electric (magnetic) field intensity, which is tangential to the perfect electric (magnetic) conductor, has to be zero

$$\mathbf{E}_t = \mathbf{0} \text{ on PEC}, \quad \mathbf{H}_t = \mathbf{0} \text{ on PMC}.$$

Neumann condition expresses the fact that derivative of the electric (magnetic) field intensity vector in the normal direction to the perfect magnetic (electric) conductor has to be zero

$$\mathbf{n} \cdot \operatorname{grad} E_j = \mathbf{0} \quad j = x,\ y,\ z \text{ on PMC}, \quad \mathbf{n} \cdot \operatorname{grad} H_j = \mathbf{0} \quad j = x,\ y,\ z \text{ on PEC}.$$

PEC denotes *Perfect Electric Conductor*, PMC denotes *Perfect Magnetic Conductor* and $\mathbf{n}$ is a normal to the conductive surface.

Cartesian coordinate system

It consists of three rectilinear mutually perpendicular basis vectors $\mathbf{x}_0$, $\mathbf{y}_0$, $\mathbf{z}_0$ (see figure). In this system, an arbitrary point is described by the vector

$$\mathbf{A} = A_x \mathbf{x}_0 + A_y \mathbf{y}_0 + A_z \mathbf{z}_0.$$

Coordinate system
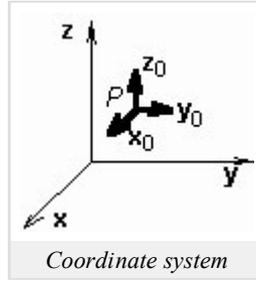
Differential operators are described by the following relations:

$$\mathrm{grad}\,\psi = \nabla\,\psi = \frac{\partial \psi}{\partial x}\,\mathbf{x}_0 + \frac{\partial \psi}{\partial y}\,\mathbf{y}_0 + \frac{\partial \psi}{\partial z}\,\mathbf{z}_0,$$

$$\mathrm{div}\,\mathbf{A} = \nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z},$$

$$\mathrm{rot}\,\mathbf{A} = \nabla \times \mathbf{A} = \begin{vmatrix} \mathbf{x}_0 & \mathbf{y}_0 & \mathbf{z}_0 \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix},$$

$$\nabla^2 \psi = \nabla \cdot \nabla \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2},$$

$$\nabla^2 \mathbf{A} = \nabla^2 A_x\,\mathbf{x}_0 + \nabla^2 A_y\,\mathbf{y}_0 + \nabla^2 A_z\,\mathbf{z}_0 =$$

$$= \left[\frac{\partial^2 A_x}{\partial x^2} + \frac{\partial^2 A_x}{\partial y^2} + \frac{\partial^2 A_x}{\partial z^2}\right]\mathbf{x}_0 + \left[\frac{\partial^2 A_y}{\partial x^2} + \frac{\partial^2 A_y}{\partial y^2} + \frac{\partial^2 A_y}{\partial z^2}\right]\mathbf{y}_0 + \left[\frac{\partial^2 A_z}{\partial x^2} + \frac{\partial^2 A_z}{\partial y^2} + \frac{\partial^2 A_z}{\partial z^2}\right]\mathbf{z}_0.$$

Caustics

Assume that a wave radiated by a point source propagates through the medium with the unitary dielectric constant first, and then continues through the layer which dielectric constant decreases down to a certain minimum, and then rises to 1 again. In a certain part of the layer, beams return by curls back to the source. Tops of the curls can be connected by an imaginary spline, which is not crossed by any beam, and in a close vicinity of which (below) two neighboring beams intersect. This spline (or a surface in space) is called caustics.

Cavity resonator

Cavity resonators belong among microwave circuits. As a cavity resonator, dielectrics of an arbitrary shape can be used, which is bounded by a conductive surface. Resonant frequency is given by dimensions and a shape of dielectrics, parameters of dielectrics, and by a mode of oscillations. Quality factor of cavity resonators varies from 10^3 to 10^5.

Characteristic impedance

Characteristic impedance Z_0 is an important parameter of the transmission line. If the transmission line is ended by the complex impedance Z_0, all the energy is transmitted by the line and there is no reflection. If the ending impedance is different, a part of the energy is reflected back to the source. A standing wave is formed then over the transmission line.

The characteristic impedance is equal to the ratio of the voltage over the current of the transit wave.

In an ideal transmission line Z_0 is real and there are no losses in the tranmission line. In a normal transmission line there is a small imaginary part. The value of the characteristic impedance depends on the radius, the shape and the dielectrical constant of the transmission line.

Charge cell

Charge cell is shifted for one half of the discretization element with respect to the current cell. The center of the charge cell appears at the edge of the current cell. This center is identical with the point, where the value of charge is computed.

Coherent radiation

Radiation exhibits time-coherence, if identical phase shift $\Delta\varphi$ can be measured whenever and everywhere for a given time interval Δt. Radiation exhibits space-coherence, if identical phase shift $\Delta\varphi$ can be measured whenever and everywhere for a given spatial distance Δl.

Collocation method

Method of weighted residua is called the collocation method when Dirac pulses play the role of weighting functions.

Continuity equation

Continuity equation describes the mutual relation between the time variation of a charge in a closed volume and currents, which flow out of this volume

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t}.$$

Outflow of current (divergence of current density) is equal to the decrease (therefore minus) of charge density per second.

Coulomb law

Assume two charges Q_1 and Q_2 in vacuum in the mutual distance d. Charges are of the opposite polarity and do not move. The charges mutually attracts by a force, which magnitude is

$$F = \frac{1}{4\pi\varepsilon_0} \frac{Q_1 Q_2}{d^2}.$$

The above-given relation expresses Coulomb law. The symbol π denotes Ludolf number and ε_0 is permittivity of vacuum ($\varepsilon_0 = 8.85 * 10^{-12}$ F * m^{-1}).

Critical frequency

Critical frequency of a mode in a waveguide specifies a frequency, when the mode starts to propagate in the waveguide. If frequency is lower than the critical one, then the mode does not propagate.

Cross polarization

Cross polarization equals to the ratio of an electric-field intensity component in a given plane (E or H) and of a component, which is perpendicular to the previous one.

Crystal lattice

Ideal crystal lattice is a model arrangement of atoms in the volume of material. This arrangement has a periodic character and can be described by the elementary cell, representing by a parallelepiped with the edge of length a in which individual particles are located. Edge length a is called the *lattice constant* and it is an important parameter to describe the crystal structure.

Current cell

Current cell is identical with the discretization element in the analysis of planar antennas by moment method. Over the surface of the current cell, a constant value of current is usually assumed.

Current distribution

Current, which flows into the dipole, reflects from the ends of the arms and returns back to the generator. The forward current and the backward one interfere. In the position, where the forward current and the backward one are of the same phase, the maximum amplitude of the current appears. In the position, where the forward current and the backward one are of the opposite phase, the minimum amplitude of the current appears. Spatial course of the current amplitude is called the current distribution.

Current induction

If electromagnetic wave impacts a conductor, free charge carriers are set into a motion (induced currents appear). Induced currents become a source of the secondary wave, which interferes with the primary one (impacting the conductor). The component of the total electric field intensity vector, which is tangential to the conductive surface, is very small (is zero if the conductor is perfectly conductive).

Cylindrical coordinate system

It consists of a single curvilinear basis vector $\boldsymbol{\varphi}_0$ and two rectilinear basis vectors $\mathbf{r}_0$, $\mathbf{z}_0$. Basis vectors are mutually perpendicular (see figure). In

this coordinate system, an arbitrary point is described by the vector

$$\mathbf{A} = \mathbf{A}_r\,\mathbf{r}_0 + \mathbf{A}_\varphi\,\boldsymbol{\varphi}_0 + \mathbf{A}_z\,\mathbf{z}_0.$$

Coordinate system
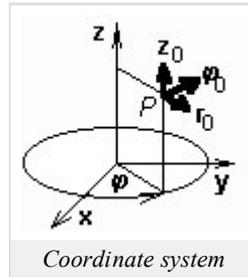
Differential operators are of the following form:

$$\operatorname{grad}\psi = \frac{\partial \psi}{\partial r}\,\mathbf{r}_0 + \frac{1}{r}\frac{\partial \psi}{\partial \varphi}\,\boldsymbol{\varphi}_0 + \frac{\partial \psi}{\partial z}\,\mathbf{z}_0,$$

$$\operatorname{div}\mathbf{A} = \frac{1}{r}\frac{\partial}{\partial r}\left(rA_r\right) + \frac{1}{r}\frac{\partial A_\varphi}{\partial \varphi} + \frac{\partial A_z}{\partial z},$$

$$\operatorname{rot}\mathbf{A} = \frac{1}{r}\begin{vmatrix} \mathbf{r}_0 & r\,\boldsymbol{\varphi}_0 & \mathbf{z}_0 \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \varphi} & \frac{\partial}{\partial z} \\ A_r & rA_\varphi & A_z \end{vmatrix},$$

$$\nabla^2 \psi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \psi}{\partial \varphi^2} + \frac{\partial^2 \psi}{\partial z^2},$$

$$\nabla^2 \mathbf{A} = \operatorname{grad\,div}\mathbf{A} - \operatorname{rot\,rot}\mathbf{A} =$$
$$= \left[\nabla^2 A_r - \frac{A_r}{r^2} - \frac{2}{r^2}\frac{\partial A_\varphi}{\partial \varphi}\right]\mathbf{r}_0 + \left[\nabla^2 A_\varphi - \frac{A_\varphi}{r^2} + \frac{2}{r^2}\frac{\partial A_r}{\partial \varphi}\right]\boldsymbol{\varphi}_0 + \nabla^2 A_z\,\mathbf{z}_0.$$

Cylindrical wave

Cylindrical wave is characteristic by the cylindrical shape of the equiphase surface. Infinitely long conductor flown by an alternate current can be the source of the cylindrical wave, e.g.

De-embedding

The above-given term denotes such a way of computing reflection (transmission) coefficient from voltage and current of a localized source so that they correspond to the measurement when long cables are used.

Diffraction factor

Diffraction factor is proportionality constant between the amplitude of diffraction waves and the amplitude of the incident wave.

Dirac pulse

Dirac pulse is an infinitely narrow pulse (is non-zero in a single point), and infinitely high pulse (the functional value tends to infinity in this single point). The filtering property of Dirac pulse

$$\int_a^b f(x)\delta(x - x_0)\,\mathrm{d}x = f(x_0)$$

is of highest importance. Since Dirac pulse $D\,(x\text{-}x_0)$ is non-zero in x_0 only, integration produces the functional value of f in x_0.

Directivity factor

Directivity factor $D(\varphi, \vartheta)$ quantifies antenna radiation in different directions. Considering isotropic radiator, directivity factor is unitary for all the angles φ a ϑ. Dealing with real antennas, directivity factor is higher than one for those directions, where radiations is concentrated by the antenna, and lower than one for those directions, where the radiation is suppressed.

Directivity pattern of the antenna

Directivity pattern is a graphic representation of directional properties of the antenna. Directional properties are usually expressed in the form of the absolute value of the radio of electric intensity of the radiated wave in a given direction and its maximum value (computation is performed for the far field of the antenna). Directivity pattern is usually drawn for a certain plane (e.g., plane perpendicular to the dipole, plane containing the dipole).

Duality principle

Investigating Maxwell equations

$$\mathrm{rot}\,\mathbf{H}_1 = \mathbf{J} + j\omega\varepsilon\,\mathbf{E}_1$$
$$\mathrm{rot}\,\mathbf{H}_2 = j\omega\varepsilon\,\mathbf{E}_2$$

(1)

$$\mathrm{rot}\,\mathbf{E}_1 = -j\omega\mu\,\mathbf{H}_1$$
$$\mathrm{rot}\,\mathbf{E}_2 = -M - j\omega\mu\,\mathbf{H}_2$$

(2)

we can simply understand, that exchanges

$$\mathbf{E}_1 \Leftrightarrow \mathbf{H}_2$$
$$\mathbf{H}_1 \Leftrightarrow -\mathbf{E}_2$$
$$\varepsilon \Leftrightarrow \mu$$
$$\mathbf{J} \Leftrightarrow \mathbf{J}_M$$

(3)

transform the left-hand side of equations to right-hand side and vice versa. Performing similar exchanges, the solution of right-hand set of equations can be obtained from the solution of left-hand side set under the same boundary conditions. E.g., two similar tasks are assumed, which differ by feeding (electric current in the first case, magnetic current in the second one) only. Solution of the first task is based on the left-hand couple of equations (1), (2), solution of the right-hand task on the right-hand couple. Solution of the first task is known, solution of the second task is rather problematic. According to the duality principle, solution of the second task is obtained by exchanging (3) in the result of the first task.

Conditions for the application of the duality principle are met at complementary objects. Two complementary objects are two planar objects. At the first one, a part of the plane is conductive (metal), the rest is not conductive (air). At the second object, the situation is reverse. E.g., a slot is complementary to a conductive strip, both the slot and the strip have to be of identical dimensions. In the strip, electric current can be excited, and the strip radiates as a dipole. In the slot, magnetic current is excited. Relation for slot radiation is the same as relation for dipole radiation if exchanges (3) are done.

Dyadic Green function

Dyadic Green function is a function of the proportionality between the current density vector $\mathbf{J}$ and vector potential $\mathbf{A}$. Dyadic Green function can be imagined as a matrix of the size 3 x 3, where each element is a function. E.g., the function in the first column and in the second row gives the contribution of the current component J_x to the potential component A_y, etc.

Edge finite elements

Edge finite elements are used to the approximation of a vector quantity in an analyzed structure. In the simplest case, the vector quantity is approximated over a finite element by a linear combination of vector functions, which are unitary vectors along a single edge of the element (different one for every function), and which are zero on the rest of edges. If partial vector functions are multiplied by the magnitude of the projection of the approximated quantity to the respective edge of the element and is the products are summed

$$\tilde{\mathbf{E}}^{(n)} = \sum_{m=1}^{3} E_m^{(n)}\,\mathbf{N}_m^{(n)},$$

we get approximation of vector quantity $\mathbf{E}$ over n^{th} finite element $\mathbf{E}^{\sim(n)}$. In the above-given relation, $E_m^{(n)}$ is the edge value (projection magnitude) of the vector $\mathbf{E}$ to m^{th} edge of nth element (a triangular one because the summation is performed over three vertexes) and $\mathbf{N}_m^{(n)}$ is a vector function, which is unitary on m^{th} edge of nth element and which is of the same orientation as the edge.

Effective permittivity

Analysis of structures, which consist of several dielectric layers (a substrate and air, e.g.), is rather complicated (boundary conditions have to be met on all the boundaries). Approximate solutions therefore replace inhomogeneous dielectrics by such homogeneous one, so that electromagnetic properties of the original structure are kept unchanged. Value of permittivity of such substitutive homogeneous dielectrics is

denoted as effective permittivity.

Eiconale

Eiconale is such a scalar function of coordinates which gradient (the direction of the steepest descent of the function) is identical with the direction of wave propagation in every point.

Electrically thin substrate

The above-given term denotes a dielectric substrate, which height is much smaller compared to wavelength in that substrate.

Elementary electric dipole

Assume an ideal dielectric (no free charges). Positive and negative charges form together neutral atoms (charges can move inside the atom only). If static external electric field takes the effect, then atoms keeps the neutrality in their whole, but the gravity center of positive and negative charges shifts. Atoms are polarized (atoms become electric dipoles).

Electric dipole is described by the moment of electric dipole **p** [C.m]

Elementary electric dipole
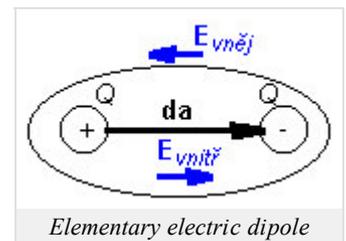
$$\mathrm{d}\mathbf{p} = Q\mathrm{d}\mathbf{a}$$

where d**a** is a positional vector from the negative pole of the atom to the positive one, and Q is charge on poles.

If external electric field varies in time, even the magnitude of a charge at the ends of the dipole changes a corresponding way. The changes of the magnitude of charge can be imagined being caused by a current flowing through the dipole (see continuity equation).

Elementary loop, magnetic dipole

In every atom, electrons move around the kernel. This movement of electrons along orbits can be imagined as a current loop, flowing by current I. Since electrons in motion create magnetic field, effects of the elementary current loop can be described by the moment of magnetic dipole

$$\mathrm{d}\mathbf{m} = I\,\mathrm{d}\mathbf{S}$$

where dS is the vector of the surface bounded by the current loop.

In ferromagnetic, external magnetic field causes such skewing of current loops so that moments of magnetic dipole approach the direction of the external magnetic induction. That way, the external field is amplified inside the ferromagnetic.

Elementary loop can be considered as the simplest loop antenna (elementary magnetic dipole).

Elementary loop
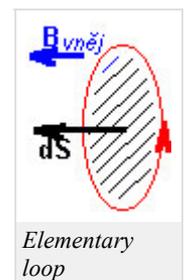
Equiphase surface

Equiphase surface is a surface of a constant phase of a wave. On the equiphase surface, amplitude can be constant (uniform wave) or can vary (non-uniform wave).

Evanescent wave

Evanescent wave is not a real wave in a fact. Phase does not change *in the propagation direction* because phase constant is zero. Amplitude exponentially decreases because attenuation constant is non-zero.

Evanescent wave can appear (under certain circumstances) on the boundary of two media when wave propagates *in parallel* with the boundary in one medium. Then, the wave does not enter the second medium, but electromagnetic field diffuses there (evanescent *wave*).

Far-field region

Far-field region is given by the distance

$$r = \frac{2d^2}{\lambda}$$

from the antenna to infinity (d is the largest transversal dimension of the antenna, λ denotes wavelength). For far-field region, trajectories of waves radiated from various parts of the antenna can be assumed to be in parallel.

Field intensity created by Huygens source

Huygens source (infinitely small facet dS) illuminated by the wave of the intensity $E^{(S)}$ from one side, which is a part of larger planar surface (also illuminated), creates a spherical wave on to non-illuminated side, which is of the following intensity in the distance r

$$\mathrm{d}E = \frac{j}{\lambda} E^{(S)} \cos(\vartheta) \frac{\exp(-jkr)}{r} \mathrm{d}S.$$

Angle ϑ is the angle between the normal to dS and the radius vector of the observation point.

Finite differences

Finite differences are used to replace the value of a derivative in such situation, when the derivative cannot be evaluated. The derivative of the function f in the point x determines the tangent direction of a function in a given point (black line in Figure). This direction can be approximated exploiting the red rectangle as
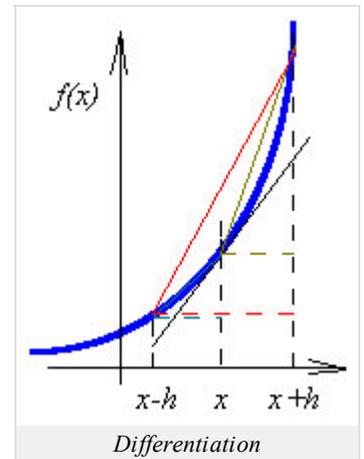
Differentiation

$$f_C'(x) \frac{f(x+h) - f(x-h)}{2h}.$$

If the point x is in the middle of the interval then we speak about the central differentiation. Red line and black one seem to be very similar.

Except of the central differentiation, the forward differentiation (dark-yellow rectangle)

$$f_F'(x) = \frac{f(x+h) - f(x)}{2h}$$

and the backward differentiation (cyan triangle)

$$f_B'(x) = \frac{f(x) - f(x-h)}{2h}$$

exist. Nevertheless, the difference between the dark-yellow line and black one (or a cyan line and black one) is more significant than in the case of the central differentiation.

Finite element

The term finite element denotes a spatial discretization element where the examined quantity is of a finite energy. A finite element can be even of an infinite extent if an examined quantity satisfactorily quickly decreases to zero over its area. E.g., energy of electromagnetic field is finite over an infinitely large element covering the area from a very long distance from the source to infinity.

In one-dimensional space, the finite element is of the form of an abscissa (the examined quantity is approximated over finite element by an abscissa, which is given by the values of the quantity in marginal points (nodes). In two-dimensional space, the finite element is of the form of a triangle (the quantity is approximated by a plane given by three points - nodal values in vertexes). In three-dimensional space, the finite element is of the form of a tetrahedron.

Finite-element method

Finite-element method is a general numerical method of solving partial differential equations. The method consists of the following steps:

1. An analyzed structure is subdivided into sub-regions (finite elements), which do not overlap on one hand, and which contain all the points of the analyzed structure on the other hand. In the region of a finite element, parameters of the analyzed structure (permittivity, permeability, conductivity) have to be constant. There are no restrictions with respect to the size and to the shape of finite elements.

2. If the analyzed structure is subdivided into finite elements, an unknown function (solution) can be approximated in formal way over the whole finite element. The unknown solution is usually approximated by linear combination of known, suitable elected basis functions and unknown approximation coefficients. That way, a single equation for M unknown approximation coefficients is obtained. If approximation coefficients are computed, an approximation of unknown quantity over the whole analyzed structure is obtained.

3. If the approximation is substituted to the solved equation, the equation is not met completely due to the approximate nature of the approximation. This fact is respected by introducing a residual function. If the approximation is accurate then the residual function is small. Therefore, the residual function is going to be minimized. The minimization is performed by the method of weighted residua.

4. Applying the method of weighted residua, the residual function is multiplied by a proper weighting function, the product is integrated over the whole analyzed space, and the result is put to equal zero. Exploiting as much properly elected weighting functions as much unknown approximation coefficients is contained in the formal approximation, a set of simultaneous linear equations is obtained, which solution uniquely determines approximation coefficients.

5. Computing approximation coefficients, a solution of the problem is obtained. Substituting approximation coefficients to the formal approximation, an approximation of the solution is obtained in every point of every finite element. Integrating approximations over finite elements, a global approximation of the solution is obtained for every point of the analyzed structure.

Focus

The term *focus* denotes a point where beams intersect.

Folded dipole

Folded dipole consists of a symmetric dipole and a near conductor, which is positioned in parallel to the dipole and which is of the same length as the dipole. Ends of the dipole are connected to the ends of the parallel conductor. The center of the parallel conductor (in front of the feeding gap) is of zero potential.

The folded dipole is of wider frequency band and of higher input impedance compared to the symmetric dipole.

Fourier series

Spectrum of a signal $s(t)$, which is periodic with the period T_1, is discrete, and its k-th component is given by the relation

$$c_k = \frac{1}{T_1} \int_{-T_1/2}^{+T_1/2} s(t) \exp(-jk\omega_1 t)\,\mathrm{d}t, \quad c_{-k} = c_k^*.$$

In the above-given relation (Fourier transform of a signal s), $\omega_1 = 2\pi/T_1$ denotes angular frequency and asterisk denotes complex conjunction.

If spectral components c_k are known, then time course of the signal can be reconstructed

$$s(t) = \sum_{k=-\infty}^{+\infty} c_k \exp(jk\omega_1 t).$$

The above-given relation is called Fourier series.

Frequency selective surface

Frequency selective surface can be imagined as a dielectric layer of very large extent, which contains planar conductive elements on its true side and which backside is free (without any metallic part). Planar conductive elements can be arbitrarily shaped, but within one selective surface, all the elements are usually identical. Conductive elements can be conductive (electrically conductive elements) or can be implemented as slots in

continuous metallic true side (magnetically conductive elements).

If the true side of the selective surface is illuminated by harmonic waves of various frequencies, some waves are transmitted with a minimum attenuation, some waves are totally reflected back to the half-space of the source, and some waves are partially transmitted and partially reflected. Hence, the surface performs a *frequency selection* of incident waves.

Fresnel diffraction

The term Fresnel diffraction describes the wave diffraction, which appears when wave transmits a cut in a thin planar obstacle. The obstacle neither transmits nor reflects the wave.

Fresnel integrals, klothoidy

Fresnel integrals are given by the relation

$$\int_0^x \exp\left(-j\frac{\pi}{2}v^2\right)dv = \int_0^x \cos\left(\frac{\pi}{2}v^2\right)dv - j\int_0^x \sin\left(\frac{\pi}{2}v^2\right)dv = C(x) - jS(x).$$

Their values can be computed by numerical integration in matlab, can be found in tables, can be read from klothoidy.

Klothoidy is a graphical representation of Fresnel integrals. On the horizontal axis, there is a real part of the integration result $C(x)$, on the vertical axis, there is an imaginary part of the result $S(x)$. In the above-described coordinate system, Fresnel integrals are of the form of a spiral curve, which is scaled by the parameter x. If the value of Fresnel integrals is computed for a given value of argument x_0, the value of argument is found on the scale of the spiral curve, and then, the value of integrals $C(x_0)$ and $S(x_0)$ is directly read on the horizontal axis and the vertical one.

Fresnel zones

In free space, assume a transmitter V and a receiver P as depicted in the figure. In an arbitrary planar cut, which is perpendicular to the line VP, such points A_1 are searched so that the pointed line VA_1P can be half wavelength longer than the direct distance VP. There is infinite number of such points and form together a circle of the center in O. Surface bounded by this circle is so called *first Fresnel zone*, and $OA_1= r_{01}$ is its radius. The same way, the other points $A_2, \ldots A_n$ can be found, which meet

First Fresnel ellipsoid
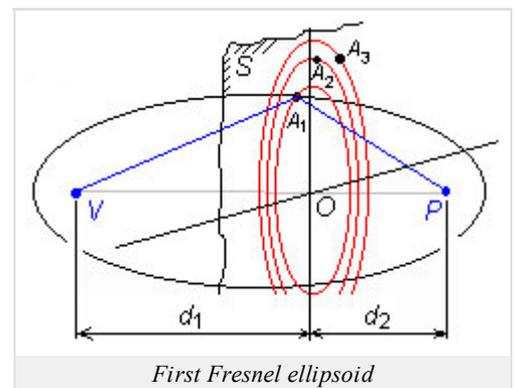
$$VA_nP = VP + n\frac{\lambda}{2}.$$

The annulus surface between A_2 and A_1 is the second Fresnel zone, etc. Radius of n_{th} zone can be computed using the approximate relation

$$r_{0n} = OA_n = \sqrt{n}\sqrt{\lambda\frac{d_1 d_2}{d_1+d_2}}.$$

Repeating the above-described consideration for points A_1 in different positions (cuts) along the line VP, different values of the radius of the first Fresnel zone are obtained. All the points A_1 are part of a rotational ellipsoid, which is called the first Fresnel ellipsoid.

Full-wave methods

Full-wave methods are based on the electromagnetic field description by complete Maxwell equations. Validity of full-wave methods is not frequency-limited. On the other hand, full-wave methods require adoption of numerical approaches (analytical solution of Maxwell equations is not known for planar structures), and hence, high computational power is needed.

Galerkin method

If weighting functions are identical with basis ones in the method of weighting residua, we speak about Galerkin method.

Gaussian beam, Gaussian function

Assume a laser beam of the basic mode TEM_{00} , which leaves the source as parallel (equiphase surface is planar and perpendicular to the direction of propagation, to the optical axis). Although the beam is not intentionally modulated, transversal field distribution is not constant. Considering the above-mentioned mode, the maximum intensity is in the axis of the beam, and the intensity decreases towards margin of the beam following Gaussian function:

$$E(x,\ y) = E_{\max} \exp\left(-\frac{\rho^2}{a_0^2} \right) = E_{\max} \exp\left(-\frac{x^2 + y^2}{a_0^2} \right).$$

Here, r is radial distance from the axis of the beam, and the function is depicted in the figure. Beam, which is characteristic by the above-described transversal Gaussian distribution of field intensity, is called Gaussian beam.

Gaussian beam, Gaussian function
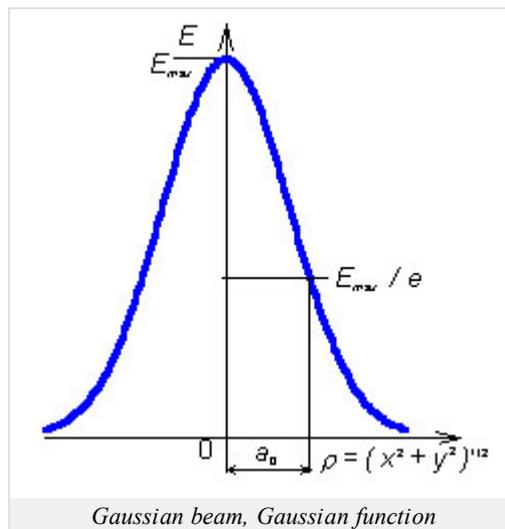
General theory of diffraction

General theory of diffraction deals with the wave diffraction on obstacles of an arbitrary shape. Since the general approach is rather complicated, results for relatively simple geometries are usually used only, and real obstacles are tried to be approximated by these objects (e.g., a conductor of finite length can be assumed as an extremely long ellipsoid).

Geometric theory of diffraction

Geometric theory of diffraction (GTD) is an extension of geometrical optics to the medium containing discontinuous inhomogeneities. We solve the interaction of beams and obstacles, and laws of geometrical optics are completed that way, so that obvious incorrectness (discontinuities on the boundary of reflection region and shade one) are eliminated on one hand, and the advantages of the geometrical optics (the idea of beams) are preserved on the other hand.

Geometrical optics

Geometrical optics is a method for the solution of wave phenomena in complex media. The method adopted the conception of wave propagation along *beams* from the classical geometrical optics. On the contrary, geometrical optics enables us to compute amplitude and phase variations of field intensity and wave polarization when wave propagates in the media of continuously changing parameters.

Global basis functions

Global basis functions are of non-zero value over the whole analyzed area. As global basis functions, harmonic functions or Bessel ones are usually used.

Green function

Green function is a function of the proportionality between the current density vector **J** and vector potential **A**.

Hankel functions

Bessel differential equation

$$y'' + \frac{1}{x} y' + \left(1 - \frac{v^2}{x^2}\right) y = 0$$

is met even by the linear combination of Hankel functions of the first kind $H_n^{(1)}(x)$ and of the second kind $H_n^{(1)}(x)$

$$y = C_3 H_n^{(1)}(x) + C_3' H_n^{(2)}(x).$$

Hankel functions meet

$$H_n^{(1)}(x) = J_n(x) + j N_n(x),$$

$$H_n^{(2)}(x) = J_n(x) - j N_n(x).$$

Hertz vector

Assume ideal dielectrics, which is polarized by the external electric field. The vector of enforced electric polarization $\mathbf{P}_{vn}$ describes polarization. Then, electric displacement vector can be expressed as

$$\mathbf{D} = \varepsilon_0 \, \mathbf{E} + \mathbf{P}_{vn}. \tag{1}$$

Substituting electric displacement vector **D** (1) to the first Maxwell equation, and assuming zero enforced current $\mathbf{J}_{vn} = \mathbf{0}$ (induced current is zero too, because the ideal dielectrics is of zero conductivity)

$$\text{rot}\,\mathbf{H} - j\omega\varepsilon_0 \, \mathbf{E} = j\omega \, \mathbf{P}_{vn}. \tag{2}$$

Then, the third Maxwell equation can be rewritten to

$$\text{div}\,\mathbf{E} = -\frac{\text{div}\,\mathbf{P}_{vn}}{\varepsilon_0}. \tag{3}$$

On the basis of (2) and (3), wave equation for vector potential can be obtained

$$\nabla^2 \mathbf{A} + k^2 \mathbf{A} = -j\omega\mu_0 \, \mathbf{P}_{vn}. \tag{4}$$

For the scalar potential, we get

$$\nabla^2 \varphi + k^2 \varphi = \frac{\text{div}\,\mathbf{P}_{vn}}{\varepsilon_0}. \tag{5}$$

We express potentials **A** and φ using an auxiliary vector $\varPi^e$

$$\mathbf{A} = j\omega\mu_0\varepsilon_0 \varPi^e. \tag{6}$$

$$\varphi = -\text{div}\,\varPi^e. \tag{7}$$

Substituting (6) to wave equation (4) and (7) to (5), we get

$$j\omega\mu_0\varepsilon_0 \left[\nabla^2 \varPi^e + k^2 \varPi^e + \frac{\mathbf{P}_{vn}}{\varepsilon_0} \right] = \mathbf{0}. \tag{8}$$

$$-\text{div}\left[\nabla^2 \varPi^e + k^2 \varPi^e + \frac{\mathbf{P}_{vn}}{\varepsilon_0} \right] = 0. \tag{9}$$

If vector π^e is determined to meet the equation

$$\nabla^2 \varPi^e + k^2 \varPi^e = -\frac{\mathbf{P}_{vn}}{\varepsilon_0}, \tag{10}$$

even (8) and (9) are met at the same time. That way, solution of Maxwell equations is transformed to the solution of a single inhomogeneous equation when $\mathbf{J}_{vn} = \mathbf{0}$. Hertz discovered this possibility, and the vector $\varPi^e$ is therefore called *electric **Hertz vector***.

In dual way, magnetic Hertz vector can be introduced.

Huygens principle

Huygens principle says that every point of a closed surface S (more accurately, every elementary facet dS), which is illuminated by a wave

arriving from outside, behaves as a source of an elementary spherical wave for the region inside the surface.

Huygens source, elementary source

The term Huygens source denotes infinitely small facet *dS*, which is illuminated by an incident wave from one side. Applying Huygens principle, this facet becomes a source of an electromagnetic wave. Real surface sources of waves (horn antennas, slot antennas, illuminated parabolic reflector) can be considered then as a set of Huygens sources.

Hybrid finite elements

If a shielded microstrip transmission line is analyzed using hybrid finite elements, all the components of electric field intensity or all the components of magnetic field intensity are included into the computation. The matter of hybrid finite elements consists in modeling a longitudinal intensity component by nodal approximation, and in modeling transversal intensity components by edge-vector approximation.

Impedance bandwidth

Impedance bandwidth can be defined as a frequency interval, where 90% of supplied energy enters the antenna (10% energy is reflected back to the generator). The described situation corresponds approximately to the standing wave ratio equal 2.

Impedance transformer

On high frequencies, impedance transformer plays the same role as the classical transformer on lower frequencies: voltage, current and impedance are transformed. Transformers differ in their construction.

As the simplest impedance transformer, one-quarter-wavelength-long transmission line can be used. The following relation describes such transmission line:

$$Z_{0tr}^2 = Z_{0in}\, Z_{0out}.$$

Here, $Z_{0tr}{}^2$ is squared value of characteristic impedance of transmission line. If the impedance Z_{0out} ends the transmission line, then the impedance Z_{0in} appears at its input.

Inhomogeneous wave equation

The term homogeneous wave equation denotes the equation of the form

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = \mathbf{0}. \tag{1}$$

This equation is valid for Cartesian coordinate system. If vector **E** meets the equation then this vector describes wave propagation. If right-hand side of the equation is zero, we are far away from the sources and wave is a source of a wave.

The right-hand side of the inhomogeneous wave equation is non-zero in contrast to (1)

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = j\omega\mu\, \mathbf{J}_e. \tag{2}$$

In the right-hand side of (2), time derivative (*jw*) of the source current $\mathbf{J}_e$ is present. This equation has to be applied if regions near sources (antennas) are investigated.

Input impedance of the antenna, radiation impedance

Input impedance of the antenna can be measured at the input port of the antenna. Input impedance consists of the radiation resistance R_Σ (when multiplied by the squared input current $I_{vst}{}^2$, real radiated power is obtained), of the loss-making resistance R_{ztr} ($R_{ztr}\, I_{vst}{}^2$ equals to the power, which is converted to the heat in the antenna and its surrounding) and of the radiation reactance X_Σ (describes the mutual energy exchange between the antenna and the electromagnetic field in its surrounding). Radiation resistance and radiation reactance form the radiation impedance together.

Ionospheric layer

The ultra-violet radiation of Sun ionizes atmospheric gases - the fission of molecules or atoms produce a positive ion and a negative electron. An ionized layer of atmosphere is called the ionosphere. The ionosphere begins in the height of 40 km from the earth surface. Concentration of electrons in the ionosphere varies with height, and significant local maxims can be observed. Intervals around local maxims are called ionospheric layers.

If electromagnetic wave impacts the ionospheric layer, then the wave trajectory is curved, and under proper conditions, the wave can return back to the earth.

Isotropic medium, anisotropic medium

Isotropic medium is of the same electric and magnetic properties in all the directions (of the same permittivity and permeability). Therefore, magnitude of the x-component of electric displacement density D_x depends on the magnitude of the x-component of the electric field intensity E_x only:

$$D_x = \varepsilon E_x, \quad D_y = \varepsilon E_y, \quad \text{etc.}$$

From the mathematical point of view, permittivity and permeability are scalar quantities.

Anisotropic medium is of different properties in different directions. Therefore, the x-component of the electric displacement density D_x depends (in general) on all the three components of electric field intensity:

$$D_x = \varepsilon_{xx} E_x + \varepsilon_{xy} E_y + \varepsilon_{xz} E_z.$$

From the mathematical point of view, permittivity and permeability are tensors, i.e. matrices of the size 3 x 3. Plasma, where steady magnetic field exists, or steady-magnetized ferrite, are examples of anisotropic media.

Lagrange interpolation polynomials

Lagrange polynomial of degree n can be in general expressed using the simplex coordinate x as

$$R_m\left(n, \xi\right) = \frac{1}{m!} \prod_{k=0}^{m-1} \left(n\xi - k\right) \quad m \geq 1 \quad R_0\left(n, \xi\right) = 1.$$

Here, n is degree of the approximation polynomial. Equation describes the whole family of polynomials; members of the family differ an index m, which can vary from zero to n.

Nulls of polynomials $R_m(n)$ are equidistantly distributed on coordinates $x=0$, $1/n$ to $(m-1)/n$, polynomial is of unitary value in $x = m/n$. Therefore, $R_m(n)$ contains m equidistantly distributed nulls on the left from $x = m/n$ and no null on the right.

Linear antenna

Linear antenna can be composed of a finite number of elementary dipoles. On the contrary, planar antennas can be composed of a finite number of Huygens sources.

Linear combination

The term linear combination of vectors (functions) denotes such vector (such function), which is obtained by adding primary vectors (primary functions) that are multiplied by arbitrary scalar non-zero coefficients.

Linear medium

Permittivity and permeability of a linear medium are constants, which are independent on intensities E and H. If electric field intensity is increased then electric displacement density proportionally rises. If magnetic field intensity is increased then magnetic flux density proportionally rises.

Local basis functions

Local basis functions are of non-zero value over a single (or a few) discretization element, and over the rest of elements, they are zero. Piecewise constant or piecewise linear function are typical representatives of local basis functions.

Longitudinally homogeneous waveguide

The term longitudinally homogeneous waveguide expresses the fact that waveguide parameters are constant in the longitudinal direction. If the waveguide is lossless, only phase of the wave changes in the longitudinal direction (amplitude stays constant).

Lorenz force

The force acting on a charged particle depends only on its position, speed and amount of charge. The force acting on the elementary charge can

be written as

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}),$$

where E and B is the electric field and magnetic induction at the position of the hub.

Matrix optics, matrix of beam, ABCD law

Matrix of beam is a matrix of size 2 x 2, which describes changes of Gaussian beam parameters when passing through an optical element. If matrix is multiplied (from the right-hand side) by a column vector, which contains beam parameters at the input of an optical element, we obtain a column vector of beam parameters at its output. The relation is called ABCD law. Due to the matrix nature of above-described operations, we speak about matrix optics.

Method of induced electromotoric forces, mutual impedance, self-impedance

The method helps us to compute voltages, currents and impedances in antenna systems. The method understands the voltage, which is observed at the input port of every single antenna element, being induced by the radiation of *all* the antenna elements (including the own element). The voltage can be composed from contributions of single elements. Each contribution is proportional to the current of the respective element. E.g., voltage U_1 at the input of the first antenna element equals to the summation

$$U_1 = Z_{11} I_1 + Z_{12} I_2 + Z_{13} I_3 + \ldots$$

where I_1, I_2, I_3 are currents at the input ports of single elements, Z_{11}, Z_{12}, Z_{13} are impedances. Z_{11} is self-impedance, Z_{1n} are mutual impedances between the first element and the other elements in the antenna system. These impedances depend on the mutual position and mutual distance of antenna elements.

Microstrip antenna, microstrip dipole, patch antenna

The term *microstrip antenna* denotes all the antennas, which are fabricated by the microstrip technology. In this textbook, a microstrip dipole (narrow microstrip conductor on the true side of the substrate, which is interrupted by a feeding gap) and a patch antenna (rectangular antenna element of the true side of the substrate, which can be fed by a microstrip, by a coaxial probe, or by an aperture in the ground plane) belong to this class of antennas.

Microstrip transmission line, ground plane

Microstrip transmission line consists of a very thin and relatively narrow strip line, which is placed to a true side of a dielectric substrate. The backside of the substrate is covered by continuous metallic layer and serves as a ground plate of zero potential. The whole transmission line is usually put into a rectangular waveguide, which plays to role of shielding. All the metallic parts of the structure (microstrip, ground plate, metallic waveguide) have to be fabricated from a well-conducting electric conductor in order to reach as low attenuation as possible. Considering the same reason, dielectric substrate has to be of as low losses as possible.

Microwave frequency band

Microwave frequency band denotes the interval of frequencies from 300 MHz to 300 GHz. Those frequencies correspond to the wavelengths from 1 m to 1 mm in a free space (in vacuum).

Mirror principle

The mirror principle enables simple modeling of the effect of well-conductive planes, which are near a radiating antenna. The radiation induces currents on the surface of the plane, which contribute to the antenna radiation as well. According to the mirror principle, contribution of induced currents equals to the contribution of a mirror images of real antenna elements. A vertical element mirrors with a current of the identical amplitude and phase. A horizontal element mirrors with a current of the identical amplitude and opposite phase.

Modal analysis of patch antennas

For modal analysis the planar antenna is modeled as a resonant cavity with perfect electric walls at the top and bottom. The main restriction is that the thickness of the dielectric substrate must be much smaller than the wavelength. Then the variance of the electric field in the plane perpendicular to the patch (axis z) is negligible. Because of the applied boundary conditions the electric field is non-zero only along the axis z, whereas the magnetic field has non-zero components only in the x-y plane. Along the boundaries of the metallic patch the Neumann condition for electric field is valid. It expresses the fact that the derivative of the electric field intensity vector normal to the perfect magnetic conductor has to be zero ($\partial \mathbf{E}/\partial \mathbf{n} = 0$). It implies the boundaries of the patch can be considered as perfect magnetic conductor.

Modes, mode numbers, dominant mode, band of unimodality

Waves propagating in a waveguide are characterized by a given transversal field distribution. The wave of a given transversal field distribution is denoted as a given mode of a wave. Every mode can be characterized by a given critical frequency, a given phase velocity, a given wavelength and a given group velocity. The mode of the lowest critical frequency is called the dominant mode. The frequency band, in which a single mode of a wave propagates in the waveguide, is called the band of unimodality.

Every mode is characterized by integer constants. Those constants are called mode numbers. Considering rectangular waveguide, mode numbers specify the number of half-waves of standing waves in the directions of transversal coordinates. E.g. TM_{11} (transversal magnetic wave) is of a single half-wave of the longitudinal component of electric field intensity in x and a single half-wave in y (if z is identical with waveguide axis). Considering TE_{10} (transversal electric wave, the dominant mode of a rectangular waveguide), a single half-wave of the longitudinal component of magnetic field intensity appears in x, and in y, this component is constant.

Moment method in spectral domain, spatial domain

Temporal-spatial distribution of electromagnetic field is described by the term $\exp[j(\omega t - kr)]$, where ω is angular frequency (a change of a phase over one second), t denotes time, k is wave-number (variation of phase over 1 meter in a given direction, i.e. spatial frequency), and r denotes length (in a given direction).

Space-domain analysis is based on computations for a single value of the spatial frequency k (the distance r is a variable). In spectral-domain analysis, the distance r is considered as a constant, and computations are performed for all the possible spatial frequencies k.

Moment method, moment solution

Moment method is a numeric method of solving integral equations (a sought function, e.g. current distribution on the antenna, is included in the integrated function). The method consists in the discretization of the analyzed structure, in formal approximation of the sought function (known basis functions and unknown approximation coefficients), in substituting the formal approximation to the solved equation (unknown approximation coefficients are removed from the integrated function and known basis functions are integrated), in minimizing the approximation error (the solved equation containing the formal approximation is sequentially multiplied by Dirac pulses in the center of discretization elements and the product is integrated over the whole analyzed structure - we obtain N equations for N unknown approximation coefficients) and in solving the final matrix equation (unknown approximation coefficients, and therefore, the approximation of the solution is obtained).

Moment table

The term moment table denotes a matrix, which contains mutual contributions among single charge cells and current ones.

Nodal finite elements

Nodal finite elements serve to piecewise continuous approximation of a scalar quantity in an analyzed structure. In the simplest case, the scalar quantity is approximated by linear function over a finite element. The linear approximation is composed of partial linear functions, which are unitary in a given vertex of the element (different one for every function) and which are zero in the rest of vertexes. If partial linear functions are multiplied by the value of the approximated quantity in such vertex of the element (in such node), where the function is of unitary value, and the products are summed

$$\tilde{E}^{(n)} = \sum_{m=1}^{3} E_m^{(n)} N_m^{(n)}$$

we get a linear approximation of E over n^{th} finite element $E^{\sim(n)}$. In the above-given relation, $E_m^{(n)}$ is a nodal value of the quantity E in m^{th} node (vertex) of nth finite element (a triangular one because the summation is performed over three vertexes) and $N_m^{(n)}$ is a linear function, which is unitary in m^{th} node of nth element.

Considering the above-described nodal approximation, the nodal values of the quantity E_m are spatial samples of the approximated quantity E.

Nodal value

The term nodal value denotes the value of a sought function in a given, exactly defined point (node). If piecewise constant approximation is used, then nodes are usually positioned to the center of discretization elements.

Parallel polarization of wave

In the case of the parallel polarization, the vector E of the wave is parallel with the plane of incidence.

Paraxial waves

Paraxial waves are waves, which directions of propagation are nearly the same (they mutually decline for a very small angle). We speak about paraxial waves if their sum is of a special meaning. E.g., a single non-uniform wave propagating in z-direction can be expressed as a sum of paraxial uniform waves, which propagate in directions those slightly decline z.

Perpendicular polarization of wave

In the case of perpendicular polarization, the vector E of the wave is perpendicular to the plane of incidence.

Perpendicular polarization, parallel polarization

Polarization describes orientation of the vector of electric field intensity of a wave.

If wave propagates near the boundary of two media, we distinguish the perpendicular polarization and the parallel polarization. First, the plane of incidence is defined as a plane, which is perpendicular to the plane of boundary and which is parallel to the direction of propagation. The vector **E** of the perpendicularly polarized wave is perpendicular to the plane of incidence. The vector **E** of the wave polarized in parallel is parallel to the plane of incidence.

If the plane of boundary is identical with the earth surface, then perpendicular polarization is called horizontal polarization, and the parallel polarization is called the vertical polarization.

Phase velocity, group velocity

Phase velocity determines the change of phase of wave per second. Phase velocity is not associated with the velocity of energy transmission, and therefore, it can be higher than velocity of light in vacuum $c = 3 * 10^{+8}$ m/s. Phase velocity tends to infinity in all directions lying on the equiphase surface (i.e. on the surface associating points of the same phase): the phase changes for 2π radians within infinitely short time period here.

On the other hand, group velocity determines the speed of energy transmission. Since in the directions, which lie on the wave-surface, energy is not transmitted (wave-surface is perpendicular to the propagation direction), group velocity is zero here.

In common media, phase velocity and group one are mutually related according to

$$v_f v_g = v^2.$$

Here v_f denotes phase velocity, v_g is group velocity, and v is free-space velocity of light.

Phonon

This is a quasi-paticle, which was introduced to represent the quantum vibrational energy of the crystal lattice. Using these particles we can describe the propagation of sound (mechanical) waves in crystals.

Planar microwave transmission line

The term planar microwave transmission line denotes a structure consisting of a metallic planar ground plate and of one or more plan-parallel planar conductors (metallic strips as electric conductors or slots in metallic plate as magnetic conductors). Various dielectric layers differing in their electric or magnetic parameters can complete the structure. Planar transmission line can be open (just described) or shielded (placed into a metallic waveguide.

Plane E, plane H

Magnetic lines of waves radiated by the antenna determine the plane H. Considering the dipole, plane H is perpendicular to the axis of the dipole.

Electric lines of waves radiated by the antenna determine the plane E. Considering the dipole, all planes containing the dipole are plane E.

Plane of boundary

Plane of boundary is a plane, which separates two media of different electric properties and magnetic ones.

Plane of incidence

Plane of incidence is perpendicular to the plane of boundary and is parallel with the propagation direction of the incident wave

Plane wave

Plane wave is characteristic by the planar equiphase surface (connecting points of the same phase, a plane is obtained). As a plane wave, we can consider a wave in a very long distance form the source when observed in a very small part of the space (in the frame of a large region, the wave is already of the spherical nature).

Point source

The term point source denotes such source of electromagnetic waves, which the only ability is the radiation ability. The source is of zero weight and zero size.

Polarization of dielectrics

Dielectrics do not contain free charge carriers. Dielectrics consist of atoms, where the positive charge of the kernel equals to the negative charge of the electrons. If the dielectrics are inserted into the external electric field, the kernel of positive charge is deflected in the contra-direction of the external electric-field intensity, and electrons of the negative charge are deflected in the direction of the external electric field. In dielectrics, an inner electrical field appears, which intensity is of the opposite orientation than the intensity of the external one. Therefore, the external electric field is weakened in the dielectrics.

Poynting vector

From the mathematical point of view, Poynting vector can be computed as vector product of electric field intensity vector and complex conjugated magnetic one. Direction of Poynting vector is identical with the direction of energy propagation. In most situations, magnitude of Poynting vector equals to the surface power density of a wave.

Propagation constant, attenuation constant, phase constant

Propagation constant γ describes variation of amplitude and variation of phase of a wave in the longitudinal direction in a waveguide. Propagation constant is given by the relation

$$\gamma = \beta + j\alpha.$$

Here, β is attenuation constant and α is phase constant). Attenuation constant describes the amplitude decrease on the trajectory 1 meter long; the amplitude decreases to exp$(-\beta)$ - multiple of its initial value. Phase constant describes the phase variation on the trajectory 1 meter long in the direction of wave.

Quality factor of a resonator

Quality factor of an arbitrary resonant circuit of angular frequency ω_0 is given by

$$Q = \frac{\omega_0 W}{P_Z}$$

where W is energy of electromagnetic field in a circuit, and P_Z is lost power in a circuit.

Quasi-static analysis, full-wave analysis

Quasi-static analysis of a microstrip transmission line consists in modeling a dominant mode (mode quasi-TEM) propagating along the line, by a transversal electromagnetic wave. Field distribution of such wave can be determined by a static analysis of the structure (microstrip is assumed to be of the potential 1 Volt, the shielding waveguide is grounded). Quasi-static analysis gives good results typically below 5 GHz.

If a microstrip transmission line is going to be analyzed on higher microwave frequencies, then *full-wave* methods have to be applied, which are based on a direct solution of Maxwell equations.

Quasi-static methods

Quasi-static methods of the analysis of planar microwave transmission lines are based on the assumption that the dominant wave propagating along the microstrip transmission line can be satisfactorily approximated by the *transversal electromagnetic wave*, *TEM*. Then, transversal components of the dynamic electromagnetic field are close to the components of the static field. Therefore, the approximate field distribution in the transmission line can be obtained by solving Laplace equation for scalar potential.

Radome

The term *radome* denotes an antenna cover, which should protect an antenna to the influence of the surrounding (rain, insects, birds, etc.). The antenna cover should even form a proper interface, which enables to reach desired properties of the antenna (minimal transmission losses, minimal influence to antenna parameters, etc.).

Rayleigh distance

From a certain distance (Rayleigh distance), wave does not propagate in directions, which correspond to the parallel beams, but it propagates in directions, which correspond to divergent beams.

Reflection coefficient, transmission coefficient

Reflection coefficient is defined as a ratio of the magnitude of a certain component of electric field intensity of a wave, which is reflected from an obstacle, and the same component of electric field intensity of incident wave. As an obstacle, we assume an object, which is large enough in the direction transversal to the direction of wave propagation (there is no diffraction there). The obstacle is of different electrical parameters than the surrounding.

Transmission coefficient is defined as a ratio of the magnitude of a certain component of electric field intensity of a wave, which transmitted through an obstacle, and the same component of electric field intensity of incident wave.

Reflector antennas

Reflector antennas consist of a primary radiator (a dipole, e.g.) and a reflector (parabolic one, e.g.). If plane wave is incident to such a reflector, the reflector concentrates all the energy in the focus. If a primary antenna is positioned to the focus, a strong power appears at its output.

Refractive index

Refractive index is given by the ratio of wave number in a given medium and wave number in the vacuum. In lossless medium, we get

$$n = \sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}} = \sqrt{\varepsilon_r\mu_r}.$$

Other words - refractive index says, how longer is wavelength in the vacuum comparing to the wavelength in a given medium.

Resonant resistance

The term denotes input resistance of a patch antenna in resonance (the patch length equals to the half of wavelength). Magnitude of resonant resistance can be influenced by the patch width and by the position of the antenna input (in patch is fed by a microstrip, then feeder is immersed into the patch).

Scalar Green function

Scalar Green function is a function of the proportionality between the charge density ρ and scalar potential V.

Scalar potential

Scalar potential simplifies the investigation of the influence of static charges (time derivative equals to zero). Field is then described by the first and the third Maxwell equation in the form

$$\begin{aligned}\operatorname{rot}\mathbf{E} &= \mathbf{0},\\ \operatorname{div}\mathbf{E} &= 0.\end{aligned} \tag{1}$$

Since curl of gradient identically equals to zero, (1) is met if

$$\mathbf{E} = -\operatorname{grad}\varphi, \tag{2}$$

(j is scalar potential). Hence, the single scalar equation (Laplace one) has to be solved

$$\nabla^2\varphi = 0. \tag{3}$$

Eqn. (3) is valid in every region, where potential is finite and continuous. Thanks to (3), the field is described by a single partial differential equation for a scalar function j.

Physical notion j can be built on the basis of (2)

$$\mathbf{E} \cdot \mathrm{d}\mathbf{r} = -\operatorname{grad}\varphi \cdot \mathrm{d}\mathbf{r} = -\mathrm{d}\varphi.$$

Since the product $\mathbf{E} \cdot d\,\mathbf{r}$ equals to the work, which is performed by the field when moving unitary charge along the elementary path $d\mathbf{r}$, elementary potential dj equals to the decrease (therefore minus) of the potential energy of this unitary charge in electrostatic field.

Shape function

The term shape function denotes a partial approximation function over a single finite element, which is unitary in a single node and which is zero in the rest of nodes. Among the most frequently used shape functions, a liner function belongs: a plane, which is unitary in a single node and which *slopes down* to zero in the rest of nodes.

All the shape functions, which are unitary in a common node, form a basis function together. In a two-dimensional case, linear basis function is a pyramid, which is unitary in a common node, and which slopes down to zero on edges of all the finite elements, which share the common node.

Sign convention

Temporal-spatial distribution of a harmonic wave is usually described by the term exp(+*jwt* - *jkr*), where angular frequency ω describes the phase change within one second and wave-number k the phase change within one meter. Moving in time, the phase increases. Moving in space in the direction of propagation, the phase decreases.

If signs in the exponential term are swapped to exp(-*jwt*+ *jkr*), then the matter of studied phenomena does not change (all the laws stay valid). Nevertheless, we have to say which sign convention is used and we have to keep the same convention during the whole computation. If some mathematical relation is going to be transformed from one convention to the second one, only signs at all the imaginary units have to be changed (+*j* « -*j*).

Simplex coordinates

Considering triangular finite element, simplex coordinate axes are of the same orientation as the triangle heights. Simplex coordinates are unitary in the vertex and zero on the opposite edge. Simplex coordinates depend neither on the shape nor on the size of the triangle. Therefore, computations can be performed for only one finite element, and results can be recomputed then to an arbitrary triangle.

Dealing with physical matter of simplex coordinates, a general point P inside the triangle subdivides its surface to three partial triangles. Ratio of the surface of a partial triangle, lying opposite to the first node, to the surface of the whole finite element, equals to the coordinate of P on the first simplex coordinate axis (similarly, the second coordinate and the third one can be obtained). Obviously, sum of all the three simplex coordinates equals to one.

Small numbers computing

Assume a very small scalar value $x \ll 1$. Then, following approximate relations are valid:

$$(1 \pm x)^n \cong 1 \pm nx,$$

e.g.

$$(1 \pm x)^n \cong 1 \pm 2x, \quad \sqrt{1 \pm x^2} \cong 1 \pm \tfrac{x}{2}, \quad \tfrac{1}{1 \pm x} \cong 1 \mp x.$$

Bessel functions:

$$J_0(x) \cong 1, \quad J_1(x) \cong \tfrac{x}{2}.$$

Neumann functions:

$$N_0(x) \cong \tfrac{2}{\pi}\left[\ln \tfrac{x}{2} + \gamma\right], \quad N_1(x) \cong -\tfrac{2}{x\pi} + N_0(x)J_1(x).$$

Snell law

Snell law describes the relation between the angle q (under this angle, wave impacts the planar boundary of two media), and the angle q_2 (angle between the direction of propagation and the normal to the boundary in the second medium). If refraction index of the first medium is n_1 (wave leaves this medium), and refraction index of the second medium is n_2 (wave enters this medium), then angles are related by the following equation:

$$n_1 \sin(\Theta_1) = n_2 \sin(\Theta_2).$$

Snell law
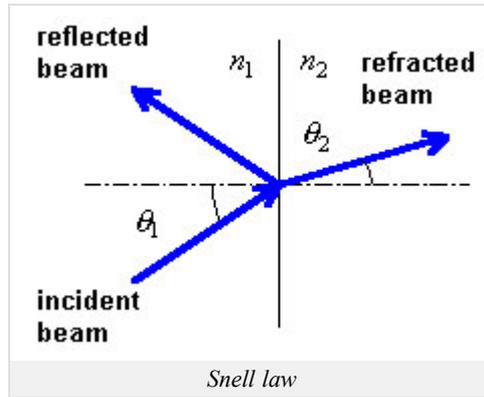
Spatial period, spatial frequency

In time domain, the term *period* denoted the shortest time distance (time interval, [s]) of two moments, when phase of harmonic oscillation differs for 2π. In spatial domain, the shortest distance [m] of two points, where phase differs for 2π is denoted as *spatial period*.

In Figure, plane wave propagation is depicted, and three planar equiphase surfaces, which are perpendicular to the wave propagation, are shown. The symbol T denotes the spatial period in the direction of propagation, and symbols T_x and T_y denote spatial periods in directions of coordinate axes x and y. Unit of the spatial period is [meter]. In the direction of propagation, spatial period equals to the wavelength λ.

Plane wave propagation

Inverted value of the spatial period is *spatial frequency*: $f = 1/T, f_x = 1/T_x$ a $f_y = 1/T_y$ (in the direction of propagation, in the direction x and in the direction y). The unit is 1/m.

Considering Figure, we can see $2\pi f = \omega = 2\pi /T = 2\pi/\lambda = k$ (wave-number in the direction of propagation), and similarly $2\pi f_y = w_y = k_y$ (wave-number in the direction y). Since $k_y = k \cos(j)$, spatial frequency depends od the direction of propagation.
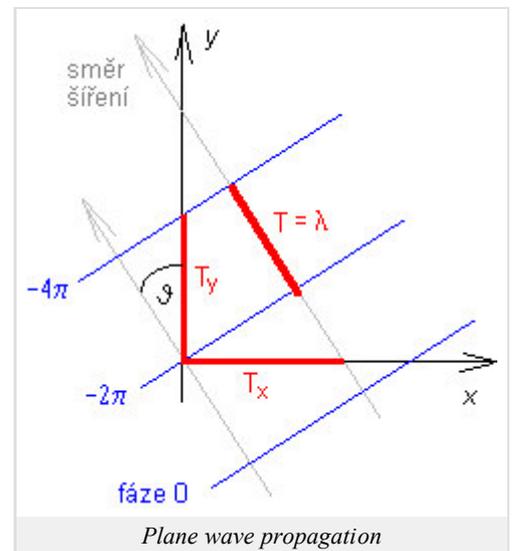
Spatial signal, optic signal, elementary spatial signal

Spatial signal (or optic signal) is transmitted by electromagnetic wave. Information is not stored in time course of field intensity but in the amplitude distribution of field intensity in the transversal plane (perpendicular to the direction of propagation): E = E(x, y), z denotes direction of propagation.

Amplitude distribution of elementary spatial signal is harmonic in one direction and is constant in another direction.

Spectral domain moment method

Assume a frequency selective surface of infinite extent and of periodic position of conductive elements. If such surface is illuminated by a plane wave, then even the distribution of induced currents is periodic. Since spectrum of periodic function is discrete, a continuous current distribution over conductive elements corresponds to the discrete spatial spectrum. Equation for discrete spectra can be simply rewritten to the matrix form, and matrix equations can be solved on the computer.

Spectral function, Fourier transform, inverse Fourier transform

Spectral function describes frequency dependency of complex amplitudes of spectral components of a signal. Spectral function S(ω) can be computed from a known time course of a signal s(*t*) using Fourier transform:

$$S(\omega) = \int_{-\infty}^{+\infty} U(t)\exp(-j\omega t)\mathrm{d}t.$$

On the contrary, time course of a signal can be computed from a known spectral function using inverse Fourier transform

$$U(t) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} S(\omega)\exp(j\omega t)\,\mathrm{d}\omega.$$

In the above-given relations, $S(\omega)$ denotes a spectral function, $U(t)$ is a time course of a signal, ω is angular frequency and t denotes time.

Spherical coordinate system

The system consists of two curvilinear basis vectors j_0, j_0 and a single direct basis vector r_0. Basis vectors are mutually perpendicular (see Figure). An arbitrary point in this coordinate system is described by the vector

$$\mathbf{A} = \mathbf{A}_r\,\mathbf{r}_0 + \mathbf{A}_\varphi\,\mathbf{\varphi}_0 + \mathbf{A}_\vartheta\,\mathbf{\vartheta}_0$$
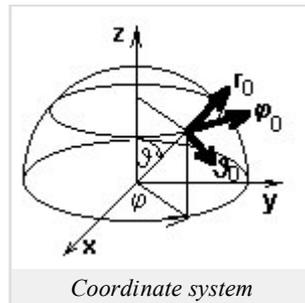
Coordinate system

Differential operators of the spherical coordinate system are described by the following relations:

$$grad\,\psi = \frac{\partial \psi}{\partial r}\,\mathbf{r}_0 + \frac{1}{r\sin\vartheta}\frac{\partial \psi}{\partial \varphi}\,\mathbf{\varphi}_0 + \frac{1}{r}\frac{\partial \psi}{\partial \vartheta}\,\mathbf{\vartheta}_0,$$

$$div\,\mathbf{A} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 A_r\right) + \frac{1}{r\sin\vartheta}\frac{\partial A_\varphi}{\partial \varphi} + \frac{1}{r\sin\vartheta}\frac{\partial}{\partial \vartheta}\left(\sin\vartheta A_\vartheta\right),$$

$$rot\,\mathbf{A} = \frac{1}{r^2\sin\vartheta}\begin{vmatrix} \mathbf{r}_0 & r\,\mathbf{\vartheta}_0 & r\sin\vartheta\,\mathbf{\varphi}_0 \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \vartheta} & \frac{\partial}{\partial \varphi} \\ A_r & rA_\vartheta & r\sin\vartheta A_\varphi \end{vmatrix}.$$

Spherical wave

Equiphase surfaces of a spherical wave are spherical surfaces. Their common center is called the phase center of a source. Spherical wave can be observed near a point radiator in free space (small antennas). Wave of real (large) sources is not spherical. However, assuming increasing distance from the source, the wave becomes spherical. In a free lossless medium, amplitude of the field intensity of the spherical wave is inversely proportional to the distance from the source.

Standing wave

If at least two coherent waves propagate in different directions, then they mutually interfere. In some regions, their phase is *permanently*

identical, intensities are summed and the total intensity is high. In other regions, their phase is not identical, and the sum of intensities is lower. The total field intensity differs from region to region. The variation is not monotonous, and even the phase does not change equably. In the space, the standing wave appeared. Regions of the lowest amplitude of field intensity are called *nodes*, and regions of the highest intensity are called *loops*.

Standing waves can be observed in media where wave reflection appears (inhomogeneous medium). The forward wave interferes with the backward one. Their propagation directions differ in 180 degrees.

Standing wave length

Standing wave length denotes the shortest distance between two neighboring points, where wave is of the maximally possible amplitude. Similar definition can be stated using minima of amplitudes (nodes).

Surface waves

Surface waves are guided by the boundary between two media. Surface waves propagate along this boundary. Surface waves can be observed

at the boundary of air and earth surface or at microstrip transmission lines, where the wave does not propagate along the microstrip conductor. Amplitude of surface waves exponentially decreases in the direction, which is perpendicular to the boundary.

Symmetric dipole

Symmetric dipole is an antenna consisting of two linear arms of the same length and the same axis. Arms are separated by a small gap, which is used for the excitation of the antenna.

Transmission line method

The above-given term denotes the method of analyzing a complex object (antenna, e.g.) by applying theory of transmission lines. That way, radiation of patch antennas can be computed, e.g. (radiating edges of patches are replaced by slot lines). The method excels in simplicity on one hand, and enables to reach relatively accurate results on the other hand.

Transversal electromagnetic wave

The term transversal electromagnetic wave (TEM) denotes the wave, which electric field intensity components and magnetic field ones are perpendicular to the propagation direction. Such wave propagates along the coaxial transmission line e.g.: components of electric-field intensity are oriented in radial direction, components of magnetic-field intensity form rings between the internal conductor and external one.

In general, TEM waves propagate along transmission lines consisting of two or more conductors in a homogeneous medium. Although the microstrip transmission line consists of two wires (strip and ground plate), TEM wave does not propagate along because the medium is not homogeneous (dielectric substrate is of different parameters than the vacuum above the strip). Since TEM wave does not enable to meet boundary conditions on the border dielectrics-vacuum, the wave has to contain even longitudinal components.

Traveling wave

If wave propagates in the infinite, totally free space, the wave is of traveling nature. In the propagation direction, the phase of oscillations is equably retarded for k' radians per meter. Wave amplitude can be constant (in lossless medium) or can monotonously decline (in lossy medium).

Uniform plane wave, non-uniform plane wave

The term uniform plane wave denotes a wave, which amplitude is constant on the equiphase surface. In case of non-uniform wave, amplitude vary on the waveform.

Vector potential

We introduce vector potential A in order to identically meet the fourth Maxwell equation

$$\operatorname{div} \mathbf{B} = 0.$$

Since divergence of curl operation equals to zero, vector potential is defined as

$$\mathbf{B} = \operatorname{rot} \mathbf{A}.$$

Distribution of potential $\mathbf{A}$ uniquely determines the field $\mathbf{B}$. On the contrary, a given field $\mathbf{B}$ is not uniquely related to a single potential $\mathbf{A}$. If (1) if valid for a vector $\mathbf{A}$, then the same relation is valid even for all vectors $\mathbf{A}' = \mathbf{A} + \operatorname{grad} y$, where y is an arbitrary scalar function (curl of gradient identically equals to zero). The described ambiguity can be eliminated if an additional condition is formulated for $\mathbf{A}$. Introduction of such condition is called calibration of a potential.

Physical notion of vector potential $\mathbf{A}$ is described by a relation defining the magnetic flux through the surface S

$$\Phi = \int_S \mathbf{B}\, d\mathbf{S} = \int_S \operatorname{rot} \mathbf{A}\, d\mathbf{S} = \oint_l \mathbf{A}\, d\mathbf{r}.$$

Obviously, the magnetic inductive flux through the surface S equals to the circulation of vector potential along the boundary curve of this surface.

Wave diffraction, diffraction task

If the wave illuminates an obstacle, which does not contain relatively large, planar and smooth surface, wave reflection does not appear in this situation. On the other hand, the obstacle significantly influences wave propagation. The respective electromagnetic phenomenon is called diffraction. We can imagine that wave trajectories are curved near the obstacle, and therefore, the wave appears even behind the obstacle (even if the obstacle does not transmit the wave).

As diffraction, even the other phenomena are explained although the trajectory curvature is not obvious at the first sight (e.g., some cases of guiding waves). In all the situations, the electromagnetic wave mutually interacts with some substantial object of different electrical parameters (permittivity, conductivity, permeability).

Wave dispersion

The term dispersion describes the phenomenon when signals on different frequencies propagate by different velocities. Different frequency components are of different delays at the output, and therefore, they represent a signal, which differs from the input one (signal is distorted).

Wave equation

Wave equation is second-order differential equation for the intensity vector of electric field or the magnetic one. If field intensities meet wave equation, the field is of the form of wave.

Considering intensity of harmonic electric field, wave equation is of the form

$$\nabla^2 \mathbf{E} + k^2 \mathbf{E} = j\omega\mu \mathbf{J}_{vn}.$$

Here, ∇ is differential operator (in Cartesian coordinates, $\nabla f = [\partial f/\partial x; \partial f/\partial y; \partial f/\partial z]$), $\mathbf{E}$ is the vector of electric field intensity, k denotes wave number, ω is angular frequency of harmonic wave, μ denotes permeability of the medium and $\mathbf{J}_{vn}$ is source current exciting wave.

Dealing with the intensity of harmonic magnetic field, wave equation is of the similar form.

Homogeneous wave equation is of the zero right-hand side (i.e., no source current appears here). Homogeneous wave equation describes the wave in a long distance from the source, where the influence of the source is negligible, and where wave becomes the source of another wave.

Wave polarization

Polarization describes the orientation of the electric field intensity vector of an electromagnetic wave.

If wave propagates in *free space*, *linear* or *elliptic* polarization can exist. The wave is linearly polarized if the vector $\mathbf{E}$ is oriented to the same direction during the whole time period. If an arrow represents instantaneous intensity vector, then its length changes from zero to maximum and back during the half-period. During the second half-period, the described phenomenon repeats with an opposite polarity. The ending point of the arrow moves along a line.

If the wave is of *elliptic polarization*, then the electric field intensity vector changes its direction during the time period. The arrow representing the instantaneous intensity vector rotates during the period, and at the same time, changes its length that way, that the ending point moves along the ellipse. The ellipse is run around within one period exactly. In a special case, the ending point can move along the circle *(circular polarization)*. Elliptic (circular) polarization can be of clockwise or anti-clockwise nature. The sense of rotation is determined from the direction of wave propagation. Elliptically polarized wave can be understood as a superposition of two coherent linearly polarized waves, which vectors $\mathbf{E}$ oscillate in different directions and a non-zero mutual shift.

If the wave propagates *near the plane of boundary* of two media, we distinguish between the *perpendicular* polarization and the *parallel* one. Let us define the *plane of incidence:* it is a plane, which is perpendicular to the plane of boundary, and at the same time, is parallel to the wave propagation. The vector $\mathbf{E}$ of the perpendicularly polarized wave is perpendicular to the plane of incidence. The vector $\mathbf{E}$ of the parallel-polarized wave is parallel to the plane of incidence.

If the plane of boundary is identical with the earth surface, then the perpendicularly polarized wave is denoted as *horizontally polarized,* and the parallel-polarized wave is called *vertically polarized one.*

Wave reflection

Reflection of a wave is caused by the incidence of a wave on satisfactorily *large* and perfectly *smooth* planar boundary between two media of different electrical parameters.

Wave scattering

If the boundary between two media is not perfectly smooth (is rough), then the incident wave "reflects" in different areas to different directions. This phenomenon is called scattering and the wave appearing on the rough surface is called the scattered wave. The term scattering denotes even the appearance of a new wave on a very high number of randomly situated particles or facets, which are illuminated by an incident wave in the same moment. As an example, scattering on a large set of raindrops can be given.

Wave vector, wave-number

In a lossless medium, wave number specifies the phase change [radians] of a propagating wave on the one-meter long trajectory in a given direction. Therefore, wave number is given in radians per meter. Wave number is associated with the wavelength $k = 2\pi/\lambda$ and with the phase velocity of propagation $k = \omega/v_f$. If wave-number k is associated with the direction of wave propagation, then both the wavelength and the phase velocity are associated with the direction of wave propagation. If another direction is considered, the above-given relations stay valid, but all the three quantities are associated with the given direction. In the direction of propagation, wave number is of the highest value, wavelength and phase velocity are of the lowest one.

In a lossy medium, wave number is complex. Its real part k' gives the phase change of wave on the one-meter long trajectory as already explained. Imaginary part (when sign is changed) k'' gives the specific wave attenuation. The wave amplitude decreases on the trajectory of the length r for

$$\left| \frac{E_2}{E_1} \right| = \exp\left(-k'' r\right)$$

where k'' is imaginary part of wave-number in the direction r.

Magnitude of the wave vector is identical with the magnitude of wave number. Direction of the wave vector is identical with the direction of wave propagation. If projections of wave vector to various directions are computed, then wavelength and phase velocity for those directions can be determined.

Waveguide

The term waveguide denotes a hollow metallic tube, which transversal dimensions are comparable to the wavelength. Wave propagation in a waveguide can be interpreted as sequential reflections from the walls of the waveguide. In the transversal directions, a standing wave appears (in different points of the cross-section, different amplitudes of field intensities can be measured).

Wavelength

Wavelength is the shortest distance between two points, where wave is of the same phase.

Weighting function, residuum, method of weighting residua

If approximation of the solution is substituted to the solved equation, the solved equation is not met exactly. This fact is respected by adding a residual function to the right-hand side of the equation. In order to minimize the residual function (i.e., in order to get as accurate approximation as possible), the residual function is multiplied by a weighting function, the product is integrated over the whole analyzed structure and the result is put to equal zero. This algorithm is called the method of weighted residua.

Wire dipole

The dipole is a direct segment of a conductor, which is interrupted by a feeding gap in the center. If a symmetric source of the harmonic voltage is connected to the gap, then in the first half-period, a current flows to one arm, and charges of the current concentrate at the end of the arm. In the second half-period, accumulated charges flow back to the source. In the second arm, the situation is similar (in the first half-period, accumulated charges flow back to the source, in the second half-period, the charges accumulate).

Dealing with the wire dipole, a circular profile and a small radius (with respect to the wavelength) of the antenna wire are assumed.

Yagi antenna, reflector, director

Yagi antenna consists of a single active element and several passive ones. The role of the active element is usually played by a symmetric (or folded) dipole operating in one-quarter-wavelength resonance. Passive elements are dipoles without the excitation gap. One passive element, which is longer than the active one, is positioned behind the active element and plays the role of the reflector. The other passive elements are shorter than the active one, are positioned in front of the active element, and are called directors.