

New directions in the wave propagation theory

ORSOLYA E. FERENCZ, CSABA FERENCZ

*Space Research Group, Institute of Geography and Earth Sciences, Eötvös University Budapest
spacerg@sas.elte.hu*

Reviewed

Keywords: wave propagation, modes, inhomogeneity

The paper presents a contradiction, originated from a root fallacy, which is commonly accepted and applied in the wave propagation calculations up to now, but yields wrong results. Further, a solving method will be briefly overviewed, by application of which it became possible to deduce new and exact solutions, to avoid the former errors, and to interpret successfully several registrations in space research.

1. Introduction

In the case of many important wave propagation problems we cannot avoid to create more and more accurate models of the physical phenomena and the structure of the propagating signals. One of the most sensitive topics is the exact description of the signals rising in inhomogeneous media, apart from the extremely strong inhomogeneities needing scattering calculations. The known and commonly applied models (e.g. W.K.B. description, Airy-functions, Stokes equation, eikonal-equation, generalized propagation vector, etc. [1]) involve fundamental misunderstanding regarding the structure of the signal. To enlighten this problem we demonstrate this contradiction in a simple example.

2. The structure of the signal

As a simple case, let a strictly monochromatic signal propagate in a linear, isotropic, time-invariant, lossless medium containing spatial inhomogeneity. In this case, a part of the signal reflects point by point during going through the medium, while the amplitude of the forward propagating signal-part will attenuate. From the simplicity of the model it is obvious that the permittivity can be defined as scalar $\varepsilon(\vec{r})$. Further simplification is assuming the permeability as μ_0 .

Consequently, the form of the signal is:

$$\vec{G}(\vec{r}, t) \triangleq \vec{G}_0(\vec{r}) e^{j[\omega t - \varphi(\vec{r})]} \quad (1)$$

where \vec{G} means $\vec{E}, \vec{B}, \vec{D}, \vec{H}$, functions, \vec{r} location vector, t is the time, ω is the angular frequency, φ is the phase.

In our case, the forms of Maxwell's equations are as follows:

$$\begin{aligned} \vec{\nabla} \times \vec{H} &= j\omega\varepsilon\vec{E}, \\ \vec{\nabla} \times \vec{E} &= -j\omega\mu_0\vec{H}, \\ \vec{\nabla} \cdot \vec{H} &= 0, \\ \vec{\nabla} \cdot (\varepsilon\vec{E}) &= 0 \end{aligned} \quad (2)$$

from which it can be obtained by the known way that the third and the fourth equations will be automatically fulfilled, if the first two equations are fulfilled, in so far as the medium is not characterized by distributions (the functions are derivable continuously).

So, the equations to be solved are the following:

$$\begin{aligned} (\vec{\nabla} \times \vec{H}_0) - j\vec{\nabla}\varphi \times \vec{H}_0 &= j\omega\varepsilon\vec{E}_0, \\ (\vec{\nabla} \times \vec{E}_0) - j\vec{\nabla}\varphi \times \vec{E}_0 &= -j\omega\mu_0\vec{H}_0. \end{aligned} \quad (3)$$

Introducing the $\vec{k} \triangleq \vec{\nabla}\varphi$ and $\vec{k} \times \vec{u} \triangleq \vec{k} \cdot \vec{u}$ notations (where \vec{u} is arbitrary vector), \vec{G}_0 and φ assuming that (as is usual in the simplest cases) are real functions, the equation-system to be solved will be disintegrated into two groups. The real part is $\vec{\nabla} \times \vec{H}_0 = 0$,

$$\vec{\nabla} \times \vec{E}_0 = 0; \quad (4)$$

while the imaginary part is $\vec{k} \times \vec{H}_0 = -\omega\varepsilon\vec{E}_0$,

$$\vec{k} \times \vec{E}_0 = \omega\mu_0\vec{H}_0. \quad (5)$$

(This separation is explained in the literature by the argumentation that weakly inhomogeneous medium is considered, in which the variation of the medium-parameters is very slow. But it is obvious that this assumption itself means a strong restriction in the validity limits of these models.)

As this separation automatically results that the law of the conservation of the energy cannot be fulfilled for the two parts separately, the W.K.B. philosophy eliminate this contradiction by introducing an additional condition regarding the constancy of the energy of the propagating signal.

(4) and (5) are investigated one by one. On the one hand, solution of (5) leads to the well-known dispersion-equation,

$$|\vec{k}\vec{k} + \omega^2\varepsilon\mu_0\vec{1}| = 0, \quad (6)$$

from that

$$k^2 = \omega^2\varepsilon\mu_0, \text{ and } k = \pm\omega\sqrt{\varepsilon\mu_0}, \quad (7)$$

can be obtained for the propagation vector, forecasting a forward and a backward propagating solution as a result. On the other hand (4) delivers a solution,

completely independent from (5), the solution of which is always

$$\vec{H}_0 = \text{constant} \quad \text{and} \quad \vec{E}_0 = \text{constant} \quad (8)$$

for the amplitudes. However, (9) is theoretically impossible in an inhomogeneous medium, and leads back to an obvious contradiction in comparison with (8).

What can be the reason of this contradiction? This evidently has to be hidden in the structure of the signal. In the traditional conceptions dealing with inhomogeneity the forward propagating and the reflected signals are taken into consideration during the derivation, as if they were the solutions of Maxwell's equations singly. As it is a well-known fact in the mathematics, the sum of several independent solutions of a linear differential equation-system is also a solution of that. But decomposing a known solution into additive parts, it cannot be assumed to be generally true, that these parts could be solutions of the original equation-system. The physical picture is clearer. In order to handle the forward propagating and the reflected signal independently, we must consider them to exist alone, as the solution of Maxwell's equations (and some coupling or relation between them can be created during the computation, by defining additional assumptions). However, the presence of the inhomogeneity automatically causes the reflection of the signal, so the propagating and the reflected signal-parts can appear only and exclusively together in inhomogeneous media, and not independently of each other.

To see the problem in more detail, let the application of the Stokes equation and Airy functions be examined [1,5].

As it can be found e.g. in Budden's book ([1] – chapters 9. and 15.), it is a routine procedure to lead back Maxwell's equations to the so called Stokes equation for inhomogeneous cases:

$$\frac{d^2 E_y}{dz^2} + k_0^2 q^2 E_y = 0 \quad (9)$$

where $q^2 = n^2$ (for longitudinal propagation) (10)

n is the refraction coefficient.

As this is well seen in Budden's deduction, he supposes the starting form of the signal as the sum of the propagating and the reflected parts:

$$E_y = A \cdot e^{-jk_z z} + B \cdot e^{jk_z z} \quad (11)$$

where $k_z = k_0 \cdot n = \frac{\omega}{c} \cdot n$

In the further deduction Budden states, that this signal form shown in (12) is used during solving the Stokes equation, the known solutions of which are the Airy-functions.

But in the followings Budden substitutes back the forward and backward propagating parts into Maxwell's equations separately, obtaining formally identical equations. After this he solves Maxwell's equations also separately, for the forward propagating and the reflected signals, and not for the sum of them.

However, as we mentioned above, the solution of Maxwell's equations can be only and exclusively the

resultant sum of the two signal-parts, because these signals cannot appear and fulfill the equations independently in the presence of spatial inhomogeneity. Let us control Budden's calculations for the resultant sum of the two signal-parts from (12), writing back their sum into the Stokes equation.

Considering Budden's assumption, whereas A and B are constants (although we must emphasize that this means strong restriction in the validity of the model) let (12) be rewritten into the Stokes equation. In this case the following will be yielded:

$$A = -B \cdot e^{j2k_z z} \quad (12)$$

This leads back to an obvious contradiction again, as from (13) A and B cannot be constants. Budden's solution therefore regards the forward and backward propagating signals independently, which cannot be assumed in inhomogeneous medium (and originally Budden has neither assumed).

If A and B are not constants and we write back (12) into the Stokes equation, the given forms will more widely differ from Budden's results, as no such differential equation will arise, the solution of which could be the Airy functions, but a more complicated relation can be written between A and B :

$$\begin{aligned} & \left[-2j \frac{dA}{dz} (\mp k_0 q) - jA \left(\mp k_0 \frac{dn}{dz} \right) + \frac{d^2 A}{dz^2} \right] \cdot e^{-jk_z z} + \\ & + \left[2j \frac{dB}{dz} (\mp k_0 q) + jB \left(\mp k_0 \frac{dn}{dz} \right) + \frac{d^2 B}{dz^2} \right] \cdot e^{+jk_z z} = 0 \end{aligned} \quad (13)$$

This relation on the one hand is not identical with the one deduced by Budden, and on the other hand, this results in an unsolvable underdetermined mathematical description.

By our investigation it turned out obviously, that the inhomogeneous computing methods using the Stokes equation involve implicitly the wrong and contradictive assumption, according to which the propagating and the reflected signals can exist and can be deduced from Maxwell's equations independently. This conclusion is valid independently from the nature of the signal (monochromatic or UWB transient).

3. Method of Inhomogeneous Basic Modes (MIBM)

As it was enlightened in details above, a wrong approach referring to the structure of the signal can cause fundamental inherent inconsistency in the solution. How could it be possible to avoid this? We have to assume such signal structure, which contains the resultant sum of all the possibly existing signals in each spatial and temporal point along the propagation path. We have to start from the point that only and exclusively this resultant sum can satisfy Maxwell's equations, but its parts (modes) independently cannot. This approach is the Method of Inhomogeneous Modes (MIBM, [2]).

To show the method, let us consider a linear, time-invariant, bi-anisotropic medium, where for the field-strengths one can write the followings:

$$\begin{aligned}\bar{D} &= \bar{\varepsilon}(\bar{r})\bar{E} + \bar{\kappa}(\bar{r})\bar{H}, \\ \bar{B} &= \bar{\nu}(\bar{r})\bar{E} + \bar{\mu}(\bar{r})\bar{H}.\end{aligned}\quad (14)$$

Supposing monochromatic functions, the general form of the signal is

$$\bar{G} = \sum_{i=1}^n a_i(\bar{r}) \cdot \bar{G}_{0i}(\bar{r}) \cdot \exp j(\omega t - \varphi_i(\bar{r})). \quad (15)$$

where $a_i(\bar{r})$ is a general envelope function depending on space, n is the number of the possible modes.

Substituting (16) into Maxwell's equations and applying some mathematical simplifications, the following equations to be solved are yielded (16):

$$\begin{aligned}\sum_{i=1}^n [\bar{\nabla}(\ln a_i - j\varphi_{ai}) \times \bar{H}_i + \bar{\nabla}_{TH0i} \bar{H}_i - j\bar{K}_i \times \bar{H}_i] &= \sum_{i=1}^n j\omega(\bar{\varepsilon}\bar{E}_i + \bar{\kappa}\bar{H}_i) \\ \sum_{i=1}^n [\bar{\nabla}(\ln a_i - j\varphi_{ai}) \times \bar{E}_i + \bar{\nabla}_{TE0i} \bar{E}_i - j\bar{K}_i \times \bar{E}_i] &= -\sum_{i=1}^n j\omega(\bar{\nu}\bar{E}_i + \bar{\mu}\bar{H}_i)\end{aligned}$$

where

$$\bar{\nabla}_{TG0i} = \begin{bmatrix} 0 & -\frac{\partial \ln G_{20i}}{\partial x_3} & \frac{\partial \ln G_{30i}}{\partial x_2} \\ \frac{\partial \ln G_{10i}}{\partial x_3} & 0 & -\frac{\partial \ln G_{30i}}{\partial x_1} \\ -\frac{\partial \ln G_{10i}}{\partial x_2} & \frac{\partial \ln G_{20i}}{\partial x_1} & 0 \end{bmatrix}; \quad (17)$$

$$\begin{aligned}\bar{K}_i &= \bar{\nabla} \varphi_i; \\ \bar{\nabla}_{\bar{\alpha}} &= \bar{\nabla} \cdot \bar{\alpha}; \\ (\bar{\nabla}_{\bar{\alpha}} \bar{G}_{0i})_{mn} &= \alpha_{mn} \frac{\partial \ln G_{0im}}{\partial x_m};\end{aligned}$$

Investigating (17) a very important feature can be recognized. This equation-system contains the whole solution arising in inhomogeneous medium, without any restriction. The final terms on the left side of the equations and the terms on the right side are completely identical with the ones valid for homogeneous case, while the first two terms on the left side are new; do not appear in homogeneous medium. As it seems to be reasonable to look for the solution in a form leads back to the known for homogeneous case, the further way of thinking is based upon this perception.

Let the inhomogeneous basic modes be defined in such a way that they deliver the solutions of the equation-parts remaining in homogeneous case, separately. But we must keep it in sight the fact, that these basic modes are not solutions of the full Maxwell's equation-system shown in (17), they fulfill just a part of it. But for homogeneous medium they trace back to the known solutions, a sin this case the first two terms disappear. Let the definition of the generalized propagation vector ($\bar{K}_i = \bar{\nabla} \varphi_i$) be the solution of the following dispersion relation, as follows

$$[(\bar{K}_i + \omega \bar{\kappa}) \bar{\mu}^{-1} (\bar{K}_i - \omega \bar{\nu}) + \omega^2 \bar{\varepsilon}] = 0. \quad (18)$$

So, the inhomogeneous basic modes belonging to \bar{K}_i are the solutions of the equations below:

$$\begin{aligned}(\bar{K}_i \times \bar{H}_i) &= -\omega(\bar{\varepsilon}\bar{E}_i + \bar{\kappa}\bar{H}_i) \\ (\bar{K}_i \times \bar{E}_i) &= \omega(\bar{\nu}\bar{E}_i + \bar{\mu}\bar{H}_i)\end{aligned}\quad (19)$$

Now, as an essentially new step differing significantly from the former methods, let the inhomogeneous modes given on the presented way be substituted into (17), into the full form of Maxwell's equations free from any eliminations. The envelope functions and the phase functions remain unknown variables. The parts remaining in homogeneous case now are cancelled out (as the inhomogeneous modes are solutions of these parts), the remnant equations are called as "coupling equations", as these will deliver the missing unknown parameters, they describe the relation among the modes and the excitation:

$$\begin{aligned}\sum_{i=1}^n [\bar{\nabla}(\ln a_i - j\varphi_{ai}) \times \bar{H}_i + \bar{\nabla}_{TH0i} \bar{H}_i] &= 0, \\ \sum_{i=1}^n [\bar{\nabla}(\ln a_i - j\varphi_{ai}) \times \bar{E}_i + \bar{\nabla}_{TE0i} \bar{E}_i] &= 0,\end{aligned}\quad (20)$$

By solving the coupling equations we can obtain the whole solution, all the simultaneously arising modes and the connection among them. This means in an inhomogeneous medium the resultant sum of the forward propagating and the reflected signal-parts, and their connection to the excitation as well.

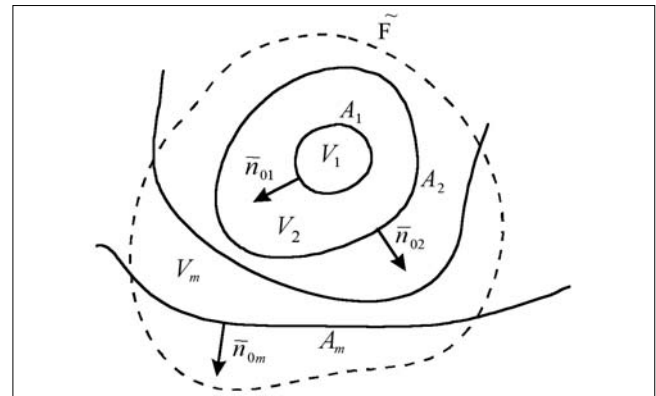
4. Solution of Maxwell's equations in the presence of distributions

Now, let the problem be examined in which the medium-parameters change suddenly at several opened or closed A_m surfaces not crossing each other (Fig. 1). Let the variation of the medium-parameters within the V_m volumes between the surfaces be continuous functions, which connect to each other by steps at the surfaces. This case is the variation of medium-parameters describable by distributions (functionals) [3].

Considering further strictly monochromatic electromagnetic signals, and supposing $\exp j(\omega t - \varphi)$ type solutions in volumes V_m , the forms valid for each volume are

$$\bar{G}_m = \left[\sum_i \bar{G}_i \right]_m = \left[\sum_i (a_i \cdot e^{-j\varphi_{ai}}) \bar{G}_{0i} \cdot e^{j(\omega t - \varphi_i)} \right]_m. \quad (21)$$

Figure 1. The structure of the medium



Furthermore, introducing the known $1(x)$ Heaviside (unit-step) and $\delta(x)$ Dirac-delta distributions, $1[\bar{F}(p_m, q_m)]$ notes the distribution the value of which changes from 0 to 1 at the surface $\bar{r} = \bar{r}(p_m, q_m)$. The $\bar{r}(p_m, q_m)$ vector is the parameter of surface A_m . Let gate-functions be created from $1[\bar{F}(p_m, q_m)]$ unit-steps belonging to surfaces A_m on the following way:

$$s_m(\bar{r}) = \{1[\bar{F}(p_{m-1}, q_{m-1})] - 1[\bar{F}(p_m, q_m)]\}, \quad (22)$$

the value of that is 1 between A_{m-1} and A_m and elsewhere 0.

By the application of the rules of derivation on these gate-functions, and keeping in mind that the generalized derivative of $1(x)$ is $\delta(x)$, one can get such a function the value of which differs from 0 only at the surfaces:

$$\bar{\nabla} \cdot s_m(\bar{r}) = \delta[\bar{r} - \bar{r}(p_{m-1}, q_{m-1})] \bar{n}_{0m-1} - \delta[\bar{r} - \bar{r}(p_m, q_m)] \bar{n}_{0m}, \quad (23)$$

where

\bar{n}_{0m} is the outward directed normal vector of A_m .

The whole solution is yielded again by the application of MIBM.

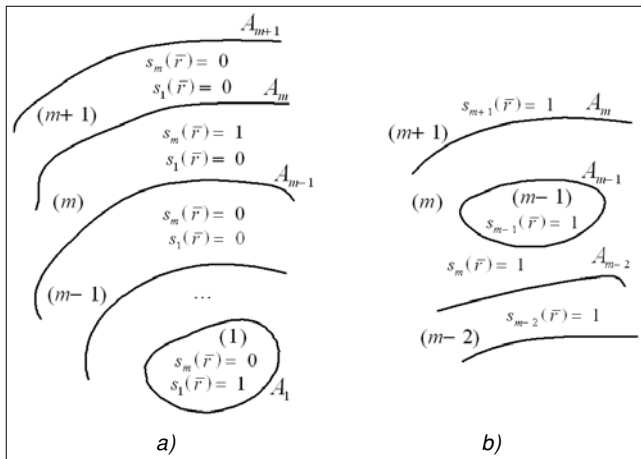


Figure 2. The distribution functions

Defining the gate-functions on a way shown on Fig. 2., one can write in each volume $s_m(\bar{r})=1$ the whole sum of all the possibly existing basic modes, and sum these in the complete examined as it follows:

$$\bar{G} = \sum_{m=1}^M s_m(\bar{r}) \left[\sum_{i=1}^n \bar{G}_i \right]_m, \quad (24)$$

where M is the number of the continuous V_m ranges.

The basic modes can be determined within each V_m on the way presented in Part 3., from the equations below:

$$\begin{aligned} \bar{K}_{im} \times \bar{H}_{im} &= -\omega(\bar{\epsilon}_m \bar{E}_{im} + \bar{\kappa}_m \bar{H}_{im}), \\ \bar{K}_{im} \times \bar{E}_{im} &= \omega(\bar{\nu}_m \bar{E}_{im} + \bar{\mu}_m \bar{H}_{im}), \end{aligned} \quad (25)$$

$$\bar{K}_{im} + \omega \bar{\kappa}_m \bar{\mu}_m^{-1} (\bar{K}_{im} - \omega \bar{\nu}_m) + \omega^2 \bar{\epsilon}_m = 0 \quad (26)$$

For the determination of the complete solution the obtained basic modes have to be substituted back into Maxwell's equations, and by solving the coupling equations the parameters still unknown can be delivered:

$$\begin{aligned} \sum_{m=1}^M \bar{\nabla} \cdot s_m(\bar{r}) \times \left[\sum_{i=1}^n \bar{H}_i \right]_m &= 0, \\ \sum_{m=1}^M \bar{\nabla} \cdot s_m(\bar{r}) \times \left[\sum_{i=1}^n \bar{E}_i \right]_m &= 0, \\ \sum_{m=1}^M \bar{\nabla} \cdot s_m(\bar{r}) \left\{ \bar{\epsilon}_m \left[\sum_{i=1}^n \bar{E}_i \right]_m + \bar{\kappa} \left[\sum_{i=1}^n \bar{H}_i \right]_m \right\} &= 0, \\ \sum_{m=1}^M \bar{\nabla} \cdot s_m(\bar{r}) \left\{ \bar{\nu}_m \left[\sum_{i=1}^n \bar{E}_i \right]_m + \bar{\mu} \left[\sum_{i=1}^n \bar{H}_i \right]_m \right\} &= 0 \end{aligned} \quad (27)$$

5. Results of the new model

Let us apply the presented method for monochromatic and transient (Ultra Wide Band, UWB) signals propagating in arbitrarily strongly inhomogeneous medium [4,5,6].

Let the medium be magnetized, anisotropic plasma (frequently occurring in the space research). Apart from the detailed overview of the deduction, here we show only several final solution formulas, illustrating that the new model modifies the structure of the signal essentially, in a large measure, in comparison with the former solutions.

In monochromatic cases (too), the solution given by MIBM is iterable by successive approximation. The zero-ordered solution of this is the well-known W.K.B. formula.

$$E_1(x) = C \sqrt{Z_0(x)} \quad (28)$$

where $C = \text{constant}$

$$E_2 = \frac{E_{10}}{2} \sqrt{Z_0(x)} \int_x^{x_M} \frac{d(\ln Z_0)}{du} e^{-j2 \int_0^u k(v) dv} du \quad (29)$$

The following, first order approximation gives more accurate formulas, and the coupling of the energy between the signal-parts can be well seen in the structure of the formulas:

$$E_1 = E_{10} \sqrt{Z_0(x)} \left\{ 1 - \frac{1}{4} \int_0^x \frac{d(\ln Z_0)}{du} e^{j2 \int_0^u k(v) dv} \left[\int_u^{x_M} \frac{d(\ln Z_0)}{dw} e^{-j2 \int_0^u k(v) dv} dw \right] du \right\} \quad (30)$$

Considering impulse propagation [7]

$$I_{x=0}(\omega) = \int_{-\infty}^0 \int_{-x_0}^0 J_0 \left(l, t + \frac{l}{c} \right) dl \left\{ e^{-j\omega t} dt \right\} \quad (31)$$

the solution for the reflected signal in the first step of the successive approximation is the following (32):

$$E_{z2}(x, t) = -\frac{j}{8\pi} \int_{-\infty}^{\infty} \left[\frac{C_0(\omega)}{\sqrt{k(x, \omega)}} \int_{\xi}^x \frac{1}{2k(u, \omega)} \frac{\partial k(u, \omega)}{\partial u} e^{-2j \int_0^u k(v, \omega) dv} du \right] e^{j \left(\omega t + \int_0^x k(h, \omega) dh \right)} d\omega$$

where

$$C_0(\omega) = I_{x=0}(\omega) \frac{k_0(\omega) \sqrt{k(x=0, \omega)}}{k_0(\omega) + k(x=0, \omega)} \quad (33)$$

$$k(x, \omega) = \frac{1}{c} \sqrt{\frac{\omega \omega_b(x) \omega_p^2(x) + \omega^2 [\omega_p^2(x) + \omega_b^2(x) - \omega^2]}{\omega_b^2(x) - \omega^2}} \quad (34)$$

It can be well seen point by point in the structure of the solution, by the integrals nested into each other, that the propagating and the reflected parts of the energy are in closed connection with each other, varying point by point.

6. Summary

In this paper a fundamental theoretical misunderstanding of the known and commonly used inhomogeneous wave propagation models was presented. This error is originated from the wrong assumption regarding the structure of the signal.

We briefly outlined the Method of Inhomogeneous Basic Modes (MIBM), by the application of that this contradiction and error cannot arise, and really accurate and right wave propagation description can be obtained.

The importance of the presented problem and solving method is great, as the wave propagation results of the last 100 years have to be revised, and opens the way toward such new, exact descriptions, by the application of which the interpretation of our knowledge and ideas regarding our global surrounding environment may go through serious development. This exact determination of the reflection will influence the research in many fields (space research, radar-technique, telecommunication etc.).

Acknowledgements

The presented results were born by the support of Hungarian Space Research Office (Ministry of Informatics and Telecommunication) and the MTA-ELTE Research Group for Geoinformatics and Space Sciences, the Hungarian Academy of Sciences, further the OTKA T037611 and F037603 contracts (already closed).

References

- [1] Budden K.G.:
Radio waves in the ionosphere;
Cambridge University Press, London 1966.
- [2] Ferencz Cs.:
Elektromágneses hullámterjedés;
Akadémiai Kiadó, Budapest 1996.
- [3] Idemen M.:
The Maxwell' equations in the Sense of Distributions;
IEEE Trans. on Ant. and Prop.; AP-21, 1973.
pp.736–738.
- [4] Cs. Ferencz:
Real solution of monochromatic wave propagation
in inhomogeneous media;
Pramana Journal of Physics, Vol.62, No.4, 2004.
pp.943–955.
- [5] Ferencz O.E.:
Full-wave solution of short impulses
in inhomogeneous plasma;
Pramana Journal of Physics, Vol.64, No.2, 2005.
pp.1–20.
- [6] Erhardtne Ferencz O. és Ferencz Cs.:
Elektromágneses impulzusok terjedésének vizsgálata
különböző közegekben;
Híradástechnika, Vol.LIX, 2004/5, pp.18–24.
- [7] Ferencz Cs., Ferencz O.E., Hamar D., Lichtenberger J.:
Whistler Phenomena, Short Impulse Propagation;
Kluwer Academic Publishers,
Astrophysics and Space Science Library,
Dordrecht, 2001.