МИНИСТЕРСТВО ОБЩЕГО И ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Известия высших учебных заведений

РАДИОФИЗИКА

ежемесячный научно–технический журнал

Издается с апреля 1958 г.

Том XLI N 8 **Нижний Новгород** **1998**

Содержание

КРАТКИЕ СООБЩЕНИЯ И ПИСЬМА В РЕДАКЦИЮ

УДК 523.985.77*3

FREQUENCY GAP BETWEEN FAST DRIFT AND TYPE III ASSOCIATED BURSTS

V. V. Zaitsev, M. Karlický

The frequency gap between the high-frequency fast drift bursts and metric type III associated bursts is interpreted using the collisional absorption of the radio emission in dense layers of the solar atmosphere. Studying the absorption of radio emission from the narrow, cold and dense filament which is embedded in hotter plasma and heated from the corona, it was found that the absorption causing the frequency gap has the maximum at about 1 GHz, in agreement with observations. Similar results were obtained also for the current sheet structure.

1. INTRODUCTION

It is well known that type III radio bursts do not extent much above 300 MHz [1]. On the other hand, fast-drift bursts (FDB) are observed in the range of 2–10 GHz [2, 3]. Some of them coincide with metric type III bursts. There are few examples when the FDB and the associated type III bursts were observed during the Hα surges [4]. Also the time coincidence between the type III bursts and active dark filaments is known [5].

It is interesting that between FDB and type III bursts, which are probably physically connected and generated by electron beams, there is a frequency gap where the burst radiation is reduced. As an example of this gap, the radio spectra of the June 14, 1989 event observed by the Ondřejov and Weissenau radio-spectrographs are shown in Fig. 1, where also the radio flux spectral profile made from data published in Solar Geophysical Data Journal is added. The gap is usually observed in the frequency range from 0.5 up to 1.5 GHz.

Electron beams and plasma mechanisms of the radio emission are considered for models of both FDB and type III radio bursts. Therefore, we assume one common electron beam generating both these bursts. The radiation frequency is related to electron density n as follows

$$n = \left\{ \begin{array}{c} 1.25 \cdot 10^{10} \\ 3.12 \cdot 10^{9} \end{array} \right\} \nu_{GHz}^2 \quad [\mathrm{cm}^{-3}], \tag{1}$$

where the upper case is for $\nu = \nu_p$, and the bottom one for $\nu = 2\,\nu_p$, ν and ν_p are the radiation and the plasma frequencies.

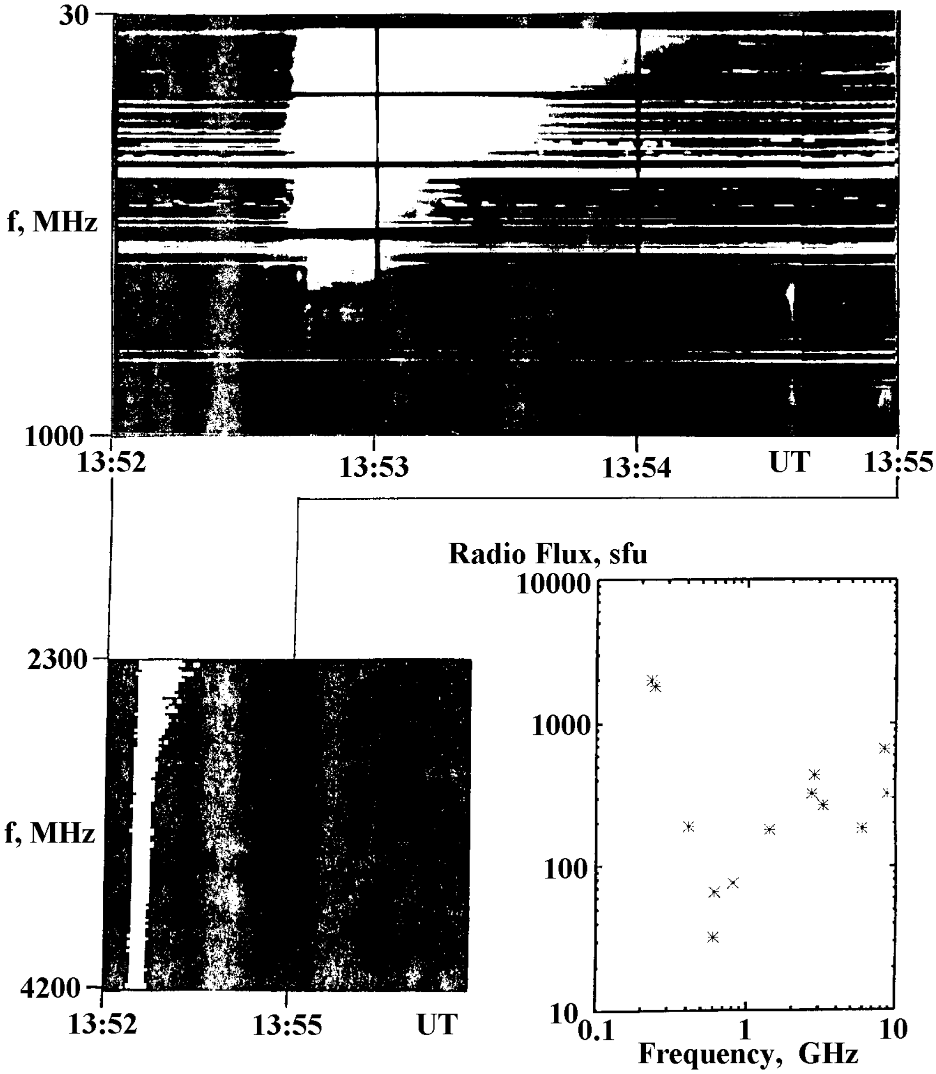


Fig. 1. Radio spectra of a group of type III and type V radio bursts (upper part) and of the fast drift burst (bottom part) observed at June 14, 1989 by the Weissenau and Ondřejov radiospectrographs. The radio flux spectral profile is added.

There is collisional absorption of the radio radiation, mainly for the radiation frequency close to the plasma frequency. This absorption can be expressed through the optical depth τ_0 [6]:

$$\tau_0 = \frac{\nu_{p0}^2}{\nu^2}\frac{\nu_{ei}}{c}\frac{H_0}{\cos\phi}, \tag{2}$$

where ν_{p0} is the plasma frequency of the plasma surrounding the radio source, ν_{ei} is the electron-ion collisional frequency, H_0 is the height scale, and ϕ is the angle between the propagation and radial directions. Thus simplifying (2) we get

$$\tau_0 = \left\{ \begin{array}{c} 3.64 \cdot 10^4 \\ 9.1 \cdot 10^3 \end{array} \right\} \left(\frac{n_0}{n}\right)^2 \frac{1}{T_0^{1/2}} \nu_{GHz}^2 \frac{1}{\cos\phi}, \tag{3}$$

where the upper case is for $\nu = \nu_p$, while the bottom one for $\nu = 2\nu_p$; n and n_0 are densities in and out of the radio source, respectively. For example, for densities $n = n_0$ and for the radiation frequency $\nu = 4$ GHz, the optical depth is

$$\tau_0 = \left\{ \begin{array}{c} 7.8 \cdot 10^3 \\ 1.8 \cdot 10^3 \end{array} \right\} \frac{1}{\cos\phi}\,, \tag{4}$$

where the temperature in relation (3) was determined from the VAL-C model of the solar atmosphere [7]. High values of the optical depth mean that no radiation can escape. Nevertheless, fast drift bursts are observed. The basic problem was solved by Benz et al. [3], who suggested that fast drift bursts are generated in narrow and dense filaments, where $n \gg n_0$. But, the problem with the frequency gap between FDB and metric type III bursts remained unsolved. Therefore, in the following we present a possible explanation of this phenomenon.

2. MODEL OF A FREQUENCY GAP

2.1. Model with dense vertical filament

Assuming a vertical dense plasma filament in deep layers of the solar atmosphere, its equilibrium can be expressed as

$$p_0 + \frac{B_0^2}{8\pi} = p + \frac{B^2}{8\pi}\,, \tag{5}$$

where p_0, p and B_0, B are the pressure and the magnetic field out and in the filament, respectively.

To simplify the calculations, let us assume $B_0^2/8\pi \ll p_0$ and $B^2/8\pi \ll p$ which gives the equilibrium condition in the form

$$\frac{n_0}{n} = \frac{T}{T_0}\,, \tag{6}$$

where T_0 and T are temperatures out and in the filament.

Moreover, we assume that

$$\nu = 2\nu_p\,, \tag{7}$$

i. e. the radiation frequency is generated on the harmonic plasma frequency.

In this case, for $\cos\phi \approx 1$, the optical depth (3) can be expressed as

$$\tau_0 = 9.1 \cdot 10^3 \left(\frac{n_0}{n}\right)^2 \frac{\nu_{GHz}^2}{T_0^{1/2}}\,. \tag{8}$$

In the transition region between the chromosphere and the corona all parameters change on scales much shorter than the atmospheric height scale $H_0 = 5000T_0$, therefore at these layers the pressure can be considered as $p_0 = \text{const}$. Furthermore, from optical observations and modelling of the transition region the relation [8]

$$n_0 T_0 = 10^{15}\ \text{cm}^{-3} \cdot \text{K} \tag{9}$$

was estimated. Now, let us assume that for the radiation frequency $\nu = 4$ GHz the optical depth of the surrounding plasma is $\tau_0(4\,\text{GHz}) = 1$. Then from this condition and relations (1), (6), and (9) the following parameters for the radiation on 4 GHz (*) can be determined:

$$n_* = 5 \cdot 10^{10}\ \text{cm}^{-3}, \quad T_* = 2 \cdot 10^4\ \text{K}, \tag{10}$$

$$n_{0*} = 3.13 \cdot 10^9\ \text{cm}^{-3}, \quad T_{0*} = 3.2 \cdot 10^5\ \text{K}. \tag{11}$$

In the VAL-C model of the solar atmosphere, these values of n_{0*} and T_{0*} correspond to the height of $h_* \approx 2300$ km.

For further analysis, let us determine the height dependence of temperature. Namely, the thermal conduction from the corona increases temperature at the transition region as well as the ratio n_0/n and the radiation absorption (see relation (8)). On the other hand, with further height increase, where $n \approx n_0$, the absorption decreases due to the radiation frequency decrease. Therefore, the optical depth $\tau_0(h)$ has a maximum at some height and thus at some radiation frequency the radiation frequency gap can be formed.

The conductive flux from the corona which heats the chromoshere can be estimated as $1.1{\cdot}10^{6}$ erg·cm^{-2}·s^{-1} [8]. This flux is conserved and therefore the height dependence of temperature can be determined from the following equation:

$$1.1 \cdot 10^{-6} T^{5/2} \frac{dT}{dh} = 1.1 \cdot 10^{6}, \tag{12}$$

the solutions in and out of the filament are

$$T(h) = \left[T_*^{7/2} + 3.5 \cdot 10^{12} (h - h_*)\right]^{2/7}, \tag{13}$$

$$T_0(h) = \left[T_{0*}^{7/2} + 3.5 \cdot 10^{12} (h - h_*)\right]^{2/7}, \tag{14}$$

where T_* and T_{0*} are given by relations (10), (11) at height of $h_* = 2300$ km.

Now, let us express the radiation frequency in dependence on temperature:

$$\nu_{GHz}^2 = \frac{n}{3.12 \cdot 10^9} = \frac{nT}{3.12 \cdot 10^9 T} = \frac{10^{15}}{3.12 \cdot 10^9 T} \approx \frac{3.2 \cdot 10^5}{T}. \tag{15}$$

Thus, using relations (6), (8), (15) the optical depth can be expessed as

$$\tau_0(h) = 2.92 \cdot 10^9 \frac{T(h)}{T_0^{5/2}(h)}, \tag{16}$$

where $T(h)$ and $T_0(h)$ are given by (13) and (14).

To express τ_0 as the function of ν the auxiliary relation

$$3.5 \cdot 10^{12} (h - h_*) = \frac{(3.2 \cdot 10^5)^{7/2}}{\nu_{GHz}^7} - T_*^{7/2} \tag{17}$$

can be derived.

Putting now this relation to the relation (16), where T and T_0 are expressed according to (13) and (14), the optical depth can be written as follows

$$\tau_0(\nu_{GHz}) = \frac{16}{\nu_{GHz}^2} \frac{1}{\left[1 - \left(\frac{T_*}{T_{0*}}\right)^{7/2} + \frac{1}{\nu_{GHz}^7}\right]^{5/7}}. \tag{18}$$

Neglecting the term $(T_*/T_{0*})^{7/2}$ in comparison with 1, we can write the final formula for the optical depth as

$$\tau_0(\nu_{GHz}) = \frac{16}{\nu_{GHz}^2} \frac{1}{\left[1 + \frac{1}{\nu_{GHz}^7}\right]^{5/7}}. \tag{19}$$

The corresponding profile in the frequency range from 0.1 to 10 GHz is shown in Fig. 2a, where it can be seen that the optical depth has the maximum at $\nu \approx 1.1$ GHz.

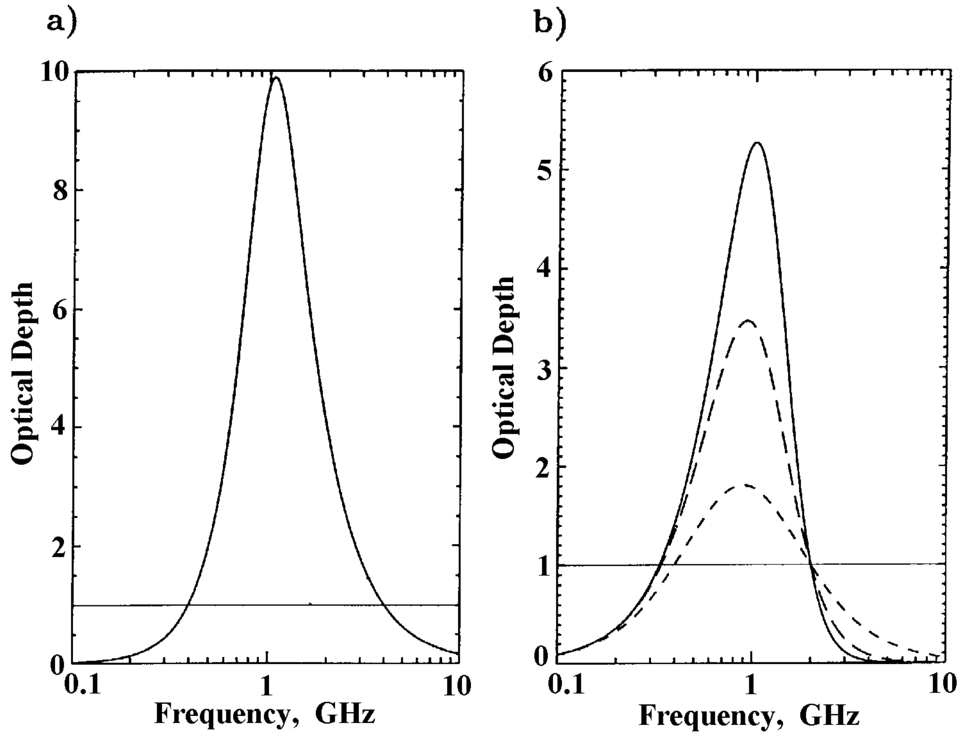


Fig. 2. a) The optical depth for the case with the dense and cold filament. b) The optical depth for the case with the current sheet. The short-dashed line is for $x = 2$, the long-dashed line is for $x = 3$, and the full line is for $x = 4$ in relations (25), (26). The horizontal thin lines express $\tau = 1$.

Finally, let us estimate the thickness of the dense filament. Namely, the optical depth of self-absorption in the dense filament, where the radio radiation is generated, must be less than 1, i. e.

$$\tau = \frac{\nu_p^2}{\nu^2}\frac{\nu_{ei}}{c} r \leq 1. \tag{20}$$

Using $\nu_p^2/\nu^2 = 1/4$, the previous condition can be rewritten to

$$r \leq \frac{4c}{70nT}T^{5/2} = \frac{4c}{7\cdot 10^{16}}\frac{(3.2\cdot 10^5)^{5/2}}{\nu_{GHz}^5} \approx \frac{10^8}{\nu_{GHz}^5}, \tag{21}$$

where r is the thickness of the filament in cm. It means that the filament thickness must be less than 1 km on 4 GHz, and 1000 km on 1 GHz, respectively.

2.2. Model with the current sheet

As mentioned in Introduction, there are few observations in which FDB associated with the type III bursts were observed during Hα surges. Therefore, we generalize the previous model to that suggested for surges and X-ray jets [9, 10]. In this case the dense vertical filament is replaced by the reconnecting current sheet.

Starting from a neutral vertical current sheet the equilibrium condition in such a sheet can be expressed as

$$\frac{B_0^2}{8\pi} + p_0 = p, \tag{22}$$

where B_0 and p_0 are the magnetic field and pressure outside the current sheet, and p is the pressure inside the sheet. Using the parameter $\beta = 16\pi nkT/B^2$, the previous equilibrium condition can be written as

$$\frac{n}{n_0} = \frac{T_0}{T}\left(\frac{1}{\beta_0} + 1\right). \tag{23}$$

Putting now relation (23) into relation (8), the optical depth can be written as

$$\tau_0 = 9.1 \cdot 10^3 \frac{\left(\frac{T}{T_0}\right)^2}{\left(\frac{1}{\beta_0} + 1\right)^2} \frac{\nu_{GHz}^2}{T^{1/2}}, \tag{24}$$

which for $T = T_0 = 10^6$ K can be simplified as follows

$$\tau_0 = 9.1 \frac{\nu_{GHz}^2}{\left(\frac{1}{\beta_0} + 1\right)^2}. \tag{25}$$

Now, putting $\tau_0(2\,\mathrm{GHz}) = 1$, we have β_0 at 2 GHz as $\beta_{00} = 0.1986$. To obtain the frequency dependence of the optical depth we need to define a reasonable form of the $\beta_0(\nu_{GHz})$ function. For this purpose we assume this function in the form

$$\beta_0 = \beta_{00}\left(\frac{\nu_0}{\nu_{GHz}}\right)^x, \tag{26}$$

where x is the parameter greater than 1, and ν_0 is equal to 2 GHz in our case.

Then for $x = 2,\ 3,\ 4$ we computed the optical depth τ_0 in dependence on radiation frequency. Results are shown in Fig. 2b, where it can be seen that the maximum of the optical depth is at about $\nu = 1$ GHz and it increases with increase of the parameter x.

3. DISCUSSION AND CONCLUSIONS

The most simple idea how to explain the frequency gap is that the electron beam crosses the density jump at the transition region. But clearly also the absorption, mainly in the chromosphere, plays an important role. Two similar models of the frequency gap between the FDB and type III associated bursts were suggested. In both models the FDB and the type III associated bursts are generated by one common electron beam. The most important aspect of both these models is that in deep layers, the radiation on frequencies 2–10 MHz is generated in a narrow and dense filament (or current sheet) with the density much greater than that of the surounding plasma, i. e. $n \gg n_0$. Furthermore, for higher heights, e. g. for lower radiation frequencies, the ratio n/n_0, which is essential for the radiation absorption, is decreasing: a) in the model with the filament due to the thermal conduction from the corona, b) in the model with the current sheet due to the change of magnetic field with the height. We think that in reality both reasons contribute simultaneously to the n/n_0 ratio.

It was shown that our model can explain the enhanced absorption of radio emission in the frequency range 0.5–1.5 GHz, as was observed. Clearly, changes of parameters of filaments or current sheets can change the resulting absorption and frequency gap profiles. It enables us, on the other hand, to derive some information about the density structure in deep layers of the solar atmosphere. Thus observations of the frequency gap between fast drift and associated type III bursts reveal a filamentary structure of the 2–10 GHz radio sources.

4. ACKNOWLEDGEMENTS

This work was supported through the key projects K1-003-601 and K1-043-601, and the grant A3003707 of the Academy of Sciences of the Czech Republic, and also by the grant 96-02-16045a of the Russian Foundation of Fundamental Researches.

We also thank Dr. H. Urbarz for the Weissenau spectral observations.

ЛИТЕРАТУРА

1. Suzuki S., Dulk G. A. — In: Solar Radiophysics /Eds. by D. J. McLean, N. R. Labrum. — Cambridge: Cambridge University Press, 1985. P. 289.
2. Tlamicha A. //Bull. Astron. Inst. Czechosl., 1991. V. 42. P. 257.
3. Benz A. O., Magun A., Stehling W., and Su H. //Solar Phys., 1992. V. 141. P. 335.
4. Kotrč P., Schmieder B., Karlický M., Heinzel P. //Solar Phys., 1996 (submitted).
5. Tlamicha A., Takakura T. //Nature, 1963. V. 200. P. 999.
6. Железняков В. В. Электромагнитные волны в космической плазме. — М.: Наука, 1977.
7. Vernazza J. E., Avrett E. H., Loeser R. //Astrophys. J. Suppl., 1981. V. 45. P. 635.
8. Mariska J. T. The Solar Transition Region, Cambridge Astrophysics Series 23. — Cambridge: Cambridge Univ. Press, 1992.
9. Yokoyama T., Shibata K. //Proceedings of IAU Colloq. № 153, Makuhari, Japan, 1995. P. 184.
10. Schmieder B., Shibata K., Van Driel-Gesztelyi L., Freeland S. //Solar Phys., 1995. V. 156. P. 245.

Institute of Applied Physics, Nizhny Novgorod, Russia;
Astronomical Institute, Academy of Sciences of the Czech Republic, Ondřejov, Czech Republic

Поступила в редакцию
17 февраля 1998 г.

УДК 537.874.4

ОБ ОТРАЖЕНИИ ПУЧКА РАДИОВОЛН ОТ ПЛАЗМЕННОГО СЛОЯ С КРУПНОМАСШТАБНЫМИ НЕОДНОРОДНОСТЯМИ

В. А. Алимов, А. В. Рахлин

С использованием метода параболического уравнения (МПУ) получены уравнения переноса для среднего поля, функции пространственной когерентности и лучевой интенсивности пучка радиоволн при отражении его от плазменного слоя со случайными неоднородностями. Найдены общие решения этих уравнений. Особо рассмотрен случай отражения пучка радиоволн от линейного плазменного слоя с крупномасштабными неоднородностями электронной концентрации. Показано, что при использовании слабонаправленной приемо–передающей КВ антенны только в условиях аномально сильных возмущений электронной концентрации в ионосфере рассеяние коротких радиоволн может приводить к заметному (порядка 3 дБ) уменьшению интенсивности отраженного от F2 слоя ионосферы сигнала вертикального зондирования.

1. МЕТОДЫ РЕФРАКЦИОННОГО РАССЕЯНИЯ РАДИОВОЛН И ПАРАБОЛИЧЕСКОГО УРАВНЕНИЯ В ДИФРАКЦИОННЫХ ЗАДАЧАХ

В начале 80-х годов для решения задач дифракции излучения в средах с крупномасштабными неоднородностями Букером был предложен метод рефракционного рассеяния радиоволн [1]. Впоследствии этот метод был развит в ряде работ (см., например, [2] и цитированную там литературу). Метод РРР применим в задачах дифракции излучения в средах с сильными (средний квадрат флуктуаций фазы волны $\overline{s^2}$ много больше единицы, $\overline{s^2} \gg 1$) крупномасштабными неоднородностями (внешний масштаб турбулентности среды L_0 больше соответствующей зоны Френеля $F_{\text{фр}}$ для исследуемой радиотрассы, $L_0 \gg F_{\text{фр}}$).

Именно такая ситуация реализуется в работах [3–6]. В [3–5] соответствующие численные расчеты ослабления когерентной составляющей интенсивности коротких радиоволн в неоднородном ионосферном слое были выполнены для турбулентности с внешним масштабом $L_0 = 10$ км (при $F_{\text{фр}} \simeq 1 \div 2$ км). При этом оказалось, что оптическая толщина среды распространения радиоволн τ много больше единицы ($\tau \gg 1$). Но, согласно [7], при дифракции излучения в среде с крупномасштабными неоднородностями $\tau = \overline{s^2}$. Поэтому не удивительно, что в [3–5] $\tau \gg 1$ при довольно слабых относительных флуктуациях электронной концентрации в неоднородностях с размерами $l \leq 1$ км $\left(\frac{\Delta N}{N} \simeq (1 \div 3) \cdot 10^{-3}\right)$. Большая оптическая толщина ионосферы обусловлена влиянием неоднородностей с размерами, близкими к внешнему масштабу турбулентности $L_0 \simeq 10$ км, а неоднородности с размерами $l \leq 1$ км дают в величину τ незначительный вклад (в этой связи см. [2] и цитированные там работы).

В [3–5] правильно отмечается, что при дифракции излучения в среде, обладающей большой оптической толщиной, необходим учет многократного рассеяния радиоволн. Согласно [1, 2], это можно сделать в рамках метода РРР.

Вместе с тем, в возмущенных геофизических условиях возможны довольно интенсивные флуктуации электронной концентрации ионосферы в неоднородностях с размерами $l \leq 1$ км $\left(\frac{\Delta N}{N} \simeq 3 \cdot 10^{-3}\right)$. При этом для отраженных коротких радиоволн возможны сильные флуктуации фазы, обусловленные этими неоднородностями ($\tau \simeq \overline{s^2}(l \leq 1\,\text{км}) \gg 1$). В этих условиях уже могут быть существенны дифракционные эффекты, вызванные неоднородностями с размерами $\lambda < l \leq F_{\text{фр}}$ (λ — длина волны излучения), которые не учитываются в рамках метода РРР. Но эти эффекты многократного рассеяния радиоволн могут быть учтены в рамках метода параболического уравнения и уравнения переноса излучения (УПИ). К анализу этих методов применительно к задаче отражения радиоволн от плоскослоистого слоя с крупномасштабными неоднородностями ($l > \lambda$) мы сейчас и перейдем.

Но вначале заметим, что впервые вопрос о влиянии области отражения на рассеяние плоских радиоволн крупномасштабными неоднородностями наиболее последовательно был решен в рамках метода возмущений Денисовым [20], а в рамках метода плавных возмущений — Кравцовым [21].

В дальнейшем эффекты многократного рассеяния радиоволн и, в частности, вопросы дифракции узких пучков радиоволн при малоугловом рассеянии их крупномасштабными неоднородностями рассматривались в целом ряде работ (см. [8, 9] и цитированную там литературу). Особо следует отметить пионерские работы Долина [10, 11], в которых впервые, более 30 лет назад, была продемонстрирована возможность эффективного ослабления интенсивности узкого пучка радиоволн после прохождения ими слоя с крупномасштабными неоднородностями показателя преломления. Этот известный эффект связан с пространственным расплыванием (диффузией) пучка радиоволн в случайно–неоднородной среде.* В последние годы метод параболического уравнения для решения задач дифракции получил дальнейшее развитие. Появились различные его модификации, из которых следует отметить, прежде всего, метод Маслова [12] и метод Орлова [13, 14], с помощью которых успешно рассматриваются различные аспекты распространения радиоволн в случайно–неоднородных средах. Вместе с тем, непосредственно сам МПУ, ввиду его вычислительных трудностей, применялся в задачах отражения радиоволн от плазменного слоя лишь для численного решения двумерных задач распространения плоских радиоволн в линейном слое случайно–неоднородной ионосферы [15]. Поэтому целый ряд вопросов, касающихся специфики МПУ и метода УПИ в задачах отражения пучков радиоволн от неоднородных плазменных образований, остается открытым. Некоторые из них мы сейчас и рассмотрим.

2. УРАВНЕНИЯ ПЕРЕНОСА ДЛЯ СРЕДНЕГО ПОЛЯ, ФУНКЦИИ ПРОСТРАНСТВЕННОЙ КОГЕРЕНТНОСТИ И ЛУЧЕВОЙ ИНТЕНСИВНОСТИ ПУЧКА РАДИОВОЛН

Рассмотрим следующую задачу. На слой плазмы со средней диэлектрической проницаемостью, изменяющейся по линейному закону

$$\bar{\varepsilon}(z) = 1 - \frac{z}{L} \tag{1}$$

(L — эффективная толщина отражающего слоя; для случая распространения высокочастотных радиоволн в плазме [17] $L = m\omega^2 \Big/ \Big(4\pi e^2 \dfrac{d\overline{N(z)}}{dz}\Big)$, где ω — круговая частота излучения, e, m — заряд и масса электрона, $\dfrac{d\overline{N(z)}}{dz}$ — линейный градиент электронной концентрации плазмы), вертикально падает пучок радиоволн. Он отражается и принимается на некотором расстоянии h в свободном пространстве за неоднородным слоем (см. рис. 1). Требуется определить статистические характеристики принимаемого излучения.

Поле сигнала удовлетворяет скалярному волновому уравнению Гельмгольца (в пренебрежении поляризационными эффектами для рассматриваемого нами случая распространения радиоволн в среде с крупномасштабными неоднородностями) [8, 17]

$$\Delta E(\vec{r}) + k_0^2[\bar{\varepsilon}(z) + \varepsilon_1(\vec{r})]E(\vec{r}) = 0. \tag{2}$$

*В [6] при анализе УПИ этим эффектом пренебрегается и ошибочно учитывается смещение пучка радиоволн, которое при корректном решении задачи о малоугловом рассеянии радиоволн крупномасштабными неоднородностями пренебрежимо мало [11, 16].

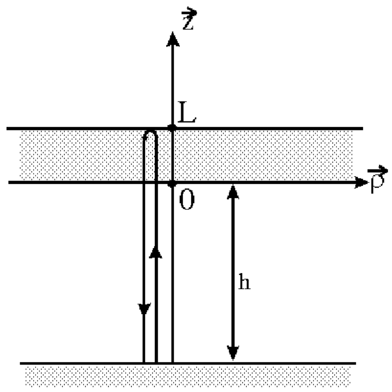


Рис. 1.

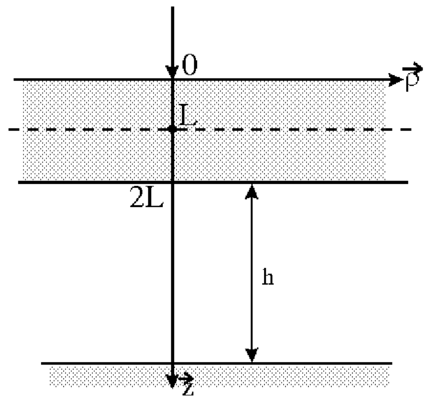


Рис. 2.

Здесь $k_0 = \dfrac{\omega}{c}$ — волновое число, $\varepsilon_1(\vec{r})$ — трехмерная пространственная флуктуация диэлектрической проницаемости среды $\Big($в нашем случае $\varepsilon_1(\vec{r}) = \dfrac{4\pi e^2}{m\omega^2}\Delta N(\vec{r})$, $\Delta N(\vec{r})$ — флуктуация электронной концентрации плазмы [17]$\Big)$.

Решение (2) будем искать в виде

$$E(z,\vec{\rho}) = u(z,\vec{\rho})\cdot E_0(z), \tag{3}$$

где $E_0(z)$ — решение уравнения (2) в отсутствие случайных неоднородностей электронной концентрации [15, 17]:

$$E_0(z) = C_s[A_i(s(z)) \mp iB_i(s(z))], \tag{4}$$

где C_s — константа, определяемая из граничных условий задачи; $A_i, B_i(s(z))$ — функции Эйри [18], аргумент $s(z)$ [15] равен

$$s(z) = -k_0^{2/3}L^{2/3}\left(1-\frac{z}{L}\right) = \frac{z-L}{\Lambda_{\text{эфф}}}, \tag{5}$$

где $\Lambda_{\text{эфф}} = \left(\dfrac{L}{k_0^2}\right)^{1/3}$ — эффективный размер волновой структуры поля вблизи каустики [15].

Для комплексной амплитуды $u(z,\vec{\rho})$ падающего излучения, рассеянного на крупномасштабных неоднородностях плазменного слоя, можно записать следующее параболическое уравнение [15]:

$$2g(z)\cdot\frac{\partial u(z,\vec{\rho})}{\partial z} + \nabla_\perp^2 u(z,\vec{\rho}) + k_0^2\varepsilon_1(\vec{r})u(z,\vec{\rho}) = 0, \tag{6}$$

где

$$g(z) = \frac{k_0^{2/3}[A_i'(s(z)) - iB_i'(s(z))]}{L^{1/3}[A_i(s(z)) - iB_i(s(z))]} = \frac{1}{E_0(z)}\,\frac{dE_0(z)}{dz}\,, \tag{7}$$

$\nabla_\perp^2$ — поперечный лапласиан (по координате $\vec{\rho}$).

Для комплексной амплитуды отраженного излучения справедливо соотношение (6) с заменой $g(z) = \operatorname{Re} g(z) + \operatorname{Im} g(z) \equiv g_1(z) + ig_2(z)$ на комплексно сопряженную величину $g^*(z)$.

Анализ соотношения (7) показывает, что в условиях плавнонеоднородной плазмы практически во всей области распространения радиоволн от начала слоя вплоть до уровня отражения $|g_2(z)| \gg |g_1(z)|$

(величины $|g_2(z)|$ и $|g_1(z)|$ сравниваются лишь в узкой области с размерами порядка $\Lambda_{эфф}$ вблизи первого максимума функции Эйри)*. С учетом этого обстоятельства, решение поставленной задачи фактически может быть сведено к решению параболического уравнения (ср. (6))

$$2ig_2(z)\frac{\partial u(z,\vec{\rho})}{\partial z}+\nabla^2_{\vec{\rho}}u(z,\vec{\rho})+k_0^2\varepsilon_1(\vec{r})u(z,\vec{\rho})=0 \tag{8}$$

для случая “просветного"распространения пучка радиоволн через неоднородный плазменный слой (см. рис. 2) с граничным значением падающего поля $u(0,\vec{\rho})=u_0(\vec{\rho})$. **

Используя соотношение (8) и следуя методике расчетов, изложенной в [8], можно получить следующие уравнения переноса для среднего поля сигнала $\overline{u(z,\vec{\rho})}$ и функции пространственной когерентности $\Gamma(z,\vec{\rho}_1,\vec{\rho}_2)=\overline{u(z,\vec{\rho}_1)u^*(z,\vec{\rho}_2)}$:

$$\frac{\partial\overline{u(z,\vec{\rho})}}{\partial z}-\frac{i}{2g_2(z)}\nabla^2_{\vec{\rho}}\overline{u(z,\vec{\rho})}+\frac{k_0^2}{8}\frac{\overline{(\Delta\varepsilon)^2(z)}}{g_2^2(z)}a(0)\overline{u(z,\vec{\rho})}=0, \tag{9}$$

$$\left[\frac{\partial}{\partial z}-\frac{i}{2g_2(z)}(\nabla^2_{\vec{\rho}_1}-\nabla^2_{\vec{\rho}_2})+\frac{\pi k_0^4\overline{(\Delta\varepsilon)^2(z)}}{4g_2^2(z)}H(\vec{\rho}_1-\vec{\rho}_2)\right]\Gamma(z,\vec{\rho}_1,\vec{\rho}_2)=0, \tag{10}$$

где

$$H(\vec{\rho}_1-\vec{\rho}_2)=\frac{1}{\pi}[a(0)-a(\vec{\rho})],$$

$$a(\vec{\rho})=\int\limits_{-\infty}^{\infty}\rho_\varepsilon(\vec{\rho},\zeta)d\zeta=2\pi\int\limits_{-\infty}^{\infty}\Phi_\varepsilon(\vec{\kappa}_\perp,0)e^{i\vec{\kappa}_\perp\vec{\rho}}d\vec{\kappa}_\perp.$$

При выводе этих уравнений было использовано известное марковское приближение [8], когда функция пространственной корреляции флуктуаций диэлектрической проницаемости среды аппроксимируется следующим соотношением:

$$\overline{\varepsilon_1(\vec{\rho}_1,z_1)\varepsilon_1(\vec{\rho}_2,z_2)}=\overline{(\Delta\varepsilon)^2(z)}\cdot a(\vec{\rho})\cdot\delta(\zeta),$$

где $\overline{(\Delta\varepsilon)^2(z=\frac{z_1+z_2}{2})}$ — средний профиль дисперсии флуктуаций диэлектрической проницаемости среды, $\delta(\zeta=z_1-z_2)$ — дельта-функция Дирака, $\rho_\varepsilon(\vec{\rho},\zeta)$ и $\Phi_\varepsilon(\vec{\kappa}_\perp,0)$ — нормированные функции пространственной корреляции и спектра флуктуаций диэлектрической проницаемости среды распространения [8].

Если с помощью соотношения [8]

$$\Gamma(\vec{\rho},\vec{\rho}_+,z)=k_0^2\iint\limits_{-\infty}^{\infty}e^{ik_0\vec{s}_\perp\vec{\rho}}\mathcal{I}(\vec{\rho}_+,\vec{s}_\perp,z)d\vec{s}_\perp \tag{11}$$

ввести понятие лучевой интенсивности $\mathcal{I}(\vec{\rho}_+,\vec{s}_\perp,z)$ ($\vec{\rho}_+=\frac{\vec{\rho}_2+\vec{\rho}_1}{2}$; $\vec{\rho}=\vec{\rho}_1-\vec{\rho}_2$; $\vec{s}_\perp$ — единичный вектор рассеяния в направлении, перпендикулярном распространению радиоволн в слое), то, используя

*Более того, практически во всей области распространения радиоволн для характерных параметров задачи вертикального зондирования ионосферы Земли ВЧ радиосигналами ($f=5$ МГц, $L=100$ км) вплоть до узкой области с размером порядка $\Lambda_{эфф}\simeq 200$ м (см. (5)) вблизи точки отражения функция $g_2(z)$ может быть аппроксимирована своим геометроптическим значением $g_2(z)\simeq k_0\sqrt{\varepsilon(z)}$ (ср. [17]).

**При такой постановке задачи эффектами двойного прохождения радиоволн неоднородностей плазменного слоя пренебрегается.

соотношение (10), можно получить уравнение переноса излучения (ср. [8, 9])

$$\frac{\partial \mathcal{I}(\vec{\rho}_+, \vec{s}_\perp, z)}{\partial z} + \frac{\vec{s}_\perp}{g_2(z)} \frac{\partial \mathcal{I}(\vec{\rho}_+, \vec{s}_\perp, z)}{\partial \vec{\rho}_+} + \frac{k_0^4 \overline{(\Delta\varepsilon)^2(z)}}{4 g_2^2(z)} a(0) \mathcal{I}(\vec{\rho}_+, \vec{s}_\perp, z) =$$

$$= \frac{\pi k_0^4 \overline{(\Delta\varepsilon)^2(z)}}{2 g_2^2(z)} \iint\limits_{-\infty}^{\infty} \Phi_\varepsilon(\vec{s}_\perp - \vec{s'}_\perp, 0) \mathcal{I}(\vec{\rho}_+, \vec{s'}_\perp, z) d\vec{s'}_\perp . \qquad (12)$$

Заметим, что для рассеянной составляющей интенсивности радиоволн $\mathcal{I}_{\text{диф}}(\vec{\rho}_+, \vec{s}_\perp, z)$ соотношение (12) может быть упрощено, используя приближение диффузии интенсивности излучения по угловым переменным [9, 11, 16]. Причем особенно простым уравнение переноса для функции $\mathcal{I}_{\text{диф}}(\vec{\rho}_+, \vec{s}_\perp, z)$ становится в случае сильного рассеяния радиоволн ($\tau \gg 1$), когда когерентной компонентой излучения можно пренебречь. В этом случае, следуя [9, 11, 16], из соотношения (12) находим

$$\frac{\partial \mathcal{I}_{\text{диф}}(\vec{\rho}_+, \vec{s}_\perp, z)}{\partial z} + \frac{\vec{s}_\perp}{g_2(z)} \frac{\partial \mathcal{I}_{\text{диф}}(\vec{\rho}_+, \vec{s}_\perp, z)}{\partial \vec{\rho}_+} - D(z) \nabla^2_{\vec{s}_\perp} \mathcal{I}_{\text{диф}}(\vec{\rho}_+, \vec{s}_\perp, z) = 0, \qquad (13)$$

где $D(z) = k_0^2 D_0 \dfrac{\overline{(\Delta\varepsilon)^2(z)}}{\overline{(\Delta\varepsilon)_0^2} g_2^2(z)}$; D_0 — коэффициент диффузии лучей в однородной среде ($\overline{\varepsilon}(z) \equiv 1$) с дисперсией флуктуаций диэлектрической проницаемости $\overline{(\Delta\varepsilon)_0^2}$ [11].

3. ДЕФОРМАЦИЯ ПУЧКА РАДИОВОЛН ПРИ ОТРАЖЕНИИ ОТ НЕОДНОРОДНОГО ПЛАЗМЕННОГО СЛОЯ

Уравнения (9), (10) и (12), (13) описывают перенос среднего поля, функции пространственной когерентности и лучевой интенсивности излучения в неоднородном слое с крупномасштабными неоднородностями. Решения их можно получить, следуя методике расчетов, изложенной в [8].

В результате несложных преобразований соотношения (9) для среднего поля излучения на выходе из слоя находим

$$\overline{u}(\vec{\rho}, z{=}2L) = \exp\left(-\frac{\tau}{2}\right) \cdot u_0(\vec{\rho}), \qquad (14)$$

где оптическая толщина τ [8, 9] равна

$$\tau = \frac{k_0^2 a(0)}{4} \int\limits_0^{2L} \frac{\overline{(\Delta\varepsilon)^2(z)}}{g_2^2(z)} dz. \qquad (15)$$

Для функции пространственной когерентности, решая уравнение (10), получаем следующее выражение (ср. [8]):

$$\Gamma(\vec{\rho}, \vec{\rho}_+, z) = \frac{1}{(2\pi)^2 \left[\int_0^z \frac{dz'}{g_2(z')}\right]^2} \iint\limits_{-\infty}^{\infty} d\vec{\rho'} d\vec{\rho'}_+ \Gamma_0(\vec{\rho'}, \vec{\rho'}_+) \cdot$$

$$\cdot \exp\left\{ i \frac{1}{\left[\int_0^z \frac{dz'}{g_2(z')}\right]} (\vec{\rho} - \vec{\rho'})(\vec{\rho}_+ - \vec{\rho'}_+) - \frac{\pi k_0^4}{4} \int\limits_0^z d\zeta \frac{\overline{(\Delta\varepsilon)^2(\zeta)}}{g_2^2(\zeta)} \cdot \right.$$

$$\left. \cdot H\left[\vec{\rho} + \frac{1}{\left[\int_0^z \frac{dz'}{g_2(z')}\right]} (\vec{\rho'} - \vec{\rho}) \int\limits_\zeta^z \frac{dz'}{g_2(z')}\right] \right\}. \qquad (16)$$

Здесь интегрирование по координате z ведется до $z = 2L$, $\Gamma_0(\vec{\rho}_+, \vec{\rho})$ — функция пространственной когерентности для падающего пучка радиоволн ($\Gamma_0(\vec{\rho}_+, \vec{\rho}) = u_0(\vec{\rho}_1) u_0^*(\vec{\rho}_2)$).

Для средней интенсивности принимаемого излучения $\overline{I}(\vec{\rho}_+, z) = = \Gamma(0, \vec{\rho}_+, z)$ из соотношения (16) имеем

$$\overline{I}(\vec{\rho}_+, z) = \frac{1}{(2\pi)^2} \iint\limits_{-\infty}^{\infty} d\vec{\kappa}_\perp F_0\left(\vec{\kappa}_\perp, \vec{\kappa}_\perp \int_0^{2L} \frac{dz'}{g_2(z')}\right) \cdot$$
$$\cdot \exp\left[-\frac{\pi k_0^4}{4} \int_0^{2L} dz\, \frac{\overline{(\Delta\varepsilon)^2(z)}}{g_2^2(z)} H\left(\vec{\kappa}_\perp \int_\zeta^{2L} \frac{dz'}{g_2(z')}\right)\right] \cdot \exp(-i\vec{\kappa}_\perp \vec{\rho}_+), \tag{17}$$

где $F_0(\vec{\kappa}_\perp)$ — фурье-спектр функции $\Gamma(\vec{\rho'}, \vec{\rho'}_+)$ по переменной $\vec{\rho'}$, $\vec{\kappa}_\perp \equiv \vec{\rho'}$.

Для случая сильных флуктуаций ($\tau \gg 1$) соотношение (17) упрощается:

$$\overline{I}(\vec{\rho'}_+, z) = \frac{1}{(2\pi)^2} \iint\limits_{-\infty}^{\infty} d\vec{\kappa}_\perp F_0\left(\vec{\kappa}_\perp, \vec{\kappa}_\perp \int_0^{2L} \frac{dz'}{g_2(z')}\right) \cdot$$
$$\cdot \exp\left[-\frac{\pi k_0^4}{8} H''(0) \int_0^{2L} dz \frac{\overline{(\Delta\varepsilon)^2(z)}}{g_2^2(z)} \left[\int_z^{2L} \frac{dz'}{g_2(z')}\right]^2 \cdot \kappa_\perp^2\right], \tag{18}$$

где $H''(0) = \left.\dfrac{d^2 H(\rho)}{d\rho^2}\right|_{\rho=0} = -\dfrac{a''(0)}{\pi}$ (см. (10)).

Несложно получить решение уравнения переноса излучения (13) в интересующем нас случае сильных флуктуаций ($\tau \gg 1$). В этом случае

$$\mathcal{I}_{\text{диф}}(z, \vec{\rho}_+, \vec{s}_\perp) = \frac{1}{(2\pi)^4} \iiiint\limits_{-\infty}^{\infty} d\vec{\kappa}_\perp d\vec{q}_\perp F\left(z, \vec{\kappa}_\perp, \vec{q}_\perp + k_0 \vec{\kappa}_\perp \int\limits_0^z \frac{dz'}{g_2(z')}\right) \cdot$$
$$\cdot \exp(-i\vec{s}_\perp \vec{q}_\perp - i\vec{\kappa}_\perp \vec{\rho}_+), \tag{19}$$

$$F\left(z, \vec{\kappa}_\perp, \vec{q}_\perp + k_0 \vec{\kappa}_\perp \int_0^z \frac{dz'}{g_2(z')}\right) = F_0\left(\vec{\kappa}_\perp, \vec{q}_\perp + k_0 \vec{\kappa}_\perp \int_0^z \frac{dz'}{g_2(z')}\right) \cdot$$
$$\cdot \exp\left[-\int\limits_0^z D(z') \left|\vec{q}_\perp + k_0 \vec{\kappa}_\perp \left[\int_{z'}^z \frac{dz''}{g_2(z'')}\right]\right|^2 dz'\right]. \tag{20}$$

Интегрирование в соотношениях (19), (20) по переменной z ведется до значений $z = 2L$.

Особенно простой вид решение (19) принимает в случае, когда прием флуктуирующего излучения в точке наблюдения осуществляется на слабонаправленную антенну (точка наблюдения расположена в ближней зоне относительно рассеивающей области на выходе неоднородного слоя). При этом получаем следующее выражение (см. (19)):

$$\mathcal{I}_{\text{диф}}(z, \vec{\rho}_+) = \iint\limits_{-\infty}^{\infty} \mathcal{I}_{\text{диф}}(z, \vec{\rho}_+, \vec{s}_\perp) d\vec{s}_\perp =$$
$$= \frac{1}{(2\pi)^2} \iint\limits_{-\infty}^{\infty} d\vec{\kappa}_\perp F(z, \vec{\kappa}_\perp) \exp(-i\vec{\kappa}_\perp \vec{\rho}_+), \tag{21}$$

где функция $F(z,\vec{\kappa}_\perp)$ определяется соотношением (20) при $\vec{q}_\perp \equiv 0$ (ср. (18)).

Используя полученные выше общие соотношения для среднего поля, функции пространственной корреляции и лучевой интенсивности излучения, несложно оценить влияние неоднородностей электронной концентрации отражающего плазменного слоя на эти статистические характеристики принимаемого излучения. При этом они легко могут быть рассчитаны в приближении френелевской дифракции флуктуирующего излучения в свободном пространстве за неоднородным слоем (см. [8]). Причем входными функциями в этих расчетах будут служить найденные выше статистические характеристики радиоволн на выходе плазменного слоя с крупномасштабными неоднородностями.*

Наибольший интерес представляет вопрос об искажении пространственного распределения интенсивности отраженного от неоднородностей плазменного слоя пучка радиоволн в случае сильных возмущений электронной концентрации.

Допустим, что на слой ионосферной плазмы вертикально падает пучок радиоволн с гауссовским распределением интенсивности $u_0(\vec{\rho}) = \exp\left(-\dfrac{\rho^2}{2a^2}\right)$. Тогда, полагая, что точка наблюдения расположена в ближней зоне относительно рассеивающей области на выходе неоднородного слоя (случай слабонаправленной приемо—передающей антенны) и проводя необходимые преобразования в формуле (18), получаем

$$\overline{I}(\vec{\rho}_+, z_{\text{н}}) \simeq \frac{1}{A^2}\exp\left[-\frac{1}{A^2}\cdot\frac{\rho_+^2}{a^2}\right], \tag{22}$$

где

$$A^2 = 1 + \left[\frac{1}{a^2}\int_0^{2L}\frac{dz'}{g_2(z')}\right]^2 - \frac{k_0^4 a''(0)}{2a^2}\int_0^{2L}\frac{\overline{(\Delta\varepsilon)^2(z)}}{g_2^2(z)}\left[\int_z^{2L}\frac{dz'}{g_2(z')}\right]^2 dz. \tag{23}$$

Расчет величины A^2 несложно выполнить для различных статистических моделей среды распространения радиоволн. При этом следует заметить, что второе слагаемое в (23) описывает регулярное (дифракционное) расплывание пучка радиоволн в неоднородном плазменном слое, а третье слагаемое — среднестатистическое уширение пучка за счет рассеяния радиоволн на крупномасштабных плазменных неоднородностях слоя. Для характерных параметров задачи вертикального зондирования F-слоя ионосферы (частота зондирования $f_0 = 5$ МГц, толщина отражающего слоя $L = 100$ км, высота ионосферного F2-слоя $h \simeq 200$ км [17]) при использовании слабонаправленных приемо—передающих антенн (с углом раскрыва главного лепестка диаграммы направленности $\vartheta \geq 60^\circ$) вторым слагаемым можно пренебречь, по сравнению с первым. Вклад третьего слагаемого в (23) легко оценить для простейших статистических моделей ионосферного слоя. Так, полагая распределение флуктуаций электронной концентрации в слое однородным с дисперсией $\overline{(\Delta\varepsilon)^2(z)} = \overline{(\Delta\varepsilon)_0^2} = \overline{\left(\dfrac{\Delta N}{\overline{N}_{\text{отр}}}\right)^2}$ ($\overline{N}_{\text{отр}}$ — электронная концентрация плазмы в области отражения радиоволн), спектр флуктуаций электронной концентрации степенным с показателем $p = 2$ или 4 (с учетом [19] $a''(0) = -\dfrac{2}{L_0}\ln\left(\dfrac{2L_0}{CL_i}\right)$ или $a''(0) = -\dfrac{2}{L_0}$ соответственно для случаев $p = 2$ и $p = 4$; L_0, L_i — внешний и внутренний масштабы ионосферной турбулентности, $L_0 \gg L_i$ [19], $C \simeq 0{,}577$ — постоянная Эйлера), третье слагаемое в (23) оказывается сравнимым с единицей при относительных флуктуациях электронной концентрации в слое $\sqrt{\overline{\left(\dfrac{\Delta N}{\overline{N}_{\text{отр}}}\right)^2}} \geq 10^{-2}$ и внешнем масштабе турбулентности $L_0 \simeq 1\div 3$ км ($f_0 = 5$ МГц, $L = 100$ км,

*В конкретных расчетах можно применять геометрооптическую аппроксимацию функции $g_2(z) = k_0\sqrt{\overline{\varepsilon}(z)}$ везде, за исключением узкой области порядка эффективного масштаба $\Lambda_{\text{эфф}}$ вблизи точки отражения (см. выше), где следует использовать численные значения функции $g_2(z)$ [18].

$h=200$ км, $\vartheta_0=60°$). При $\sqrt{\overline{\left(\frac{\Delta N}{\overline{N}_{\text{отр}}}\right)^2}} \leq 3\cdot 10^{-3}$ третье слагаемое пренебрежимо мало.

Итак, в условиях аномально сильных возмущений электронной концентрации в ионосфере (относительные флуктуации электронной концентрации для неоднородностей с внешним масштабом турбулентности L_0 в единицы километров во всем отражающем слое толщиной $L\simeq 100$ км составляют единицы процентов) возможно заметное (порядка 3 дБ) уменьшение интенсивности отраженного от F2-слоя ионосферы сигнала даже при использовании слабонаправленной приемо–передающей КВ антенны вертикального зондирования ($\vartheta_0 \simeq 60°$). * В спокойных геофизических условиях влияние крупномасштабных неоднородностей ионосферной плазмы на интенсивность отраженного КВ сигнала вертикального зондирования пренебрежимо мало (ср. [6]!).

Работа выполнена в рамках проекта 96–02–18632 РФФИ.

ЛИТЕРАТУРА

1. Booker H. G. //Indian Radio Space Phys., 1986. V. 15. № 5–6. P. 197.
2. Алимов В. А., Рахлин А. В., Выборнов Ф. И. //Изв. ВУЗов. Радиофизика, 1997. Т. 40. № 11. С. 1323.
3. Бронин А. Г., Денисенко П. Ф., Заботин Н. А. //Геомагнетизм и аэрономия, 1993. Т. 33. № 1. С. 169.
4. Бронин А. Г., Денисенко П. Ф., Заботин Н. А., Ямпольский Ю. М. //Геомагнетизм и аэрономия, 1991. Т. 31. № 5. С. 945.
5. Бронин А. Г., Заботин Н. А., Черкашин Ю. Н. //Геомагнетизм и аэрономия, 1996. Т. 36. № 4. С. 107.
6. Заботин Н. А. //Изв. ВУЗов. Радиофизика, 1993. Т. 36. № 12. С. 1075.
7. Алимов В. А. //Изв. ВУЗов. Радиофизика, 1995. Т. 38. № 11. С. 1118.
8. Рытов С. М., Кравцов Ю. А., Татарский В. И. Введение в статистическую радиофизику. — М.: Наука, 1978. Ч. 2.
9. Исимару А. Распространение и рассеяние волн в случайно–неоднородных средах. — М.: Мир, 1981. Ч. 2.
10. Долин Л. С. //Изв. ВУЗов. Радиофизика, 1964. Т. 7. № 4. С. 380.
11. Долин Л. С. //Изв. ВУЗов. Радиофизика, 1966. Т. 9. № 1. С. 61.
12. Кравцов Ю. А. //Акустический журнал, 1968. Т. 14. № 1. С. 1.
13. Авдеев В. Б., Демин А. В., Кравцов Ю. А., Тинин М. В., Ярыгин А. П. //Изв. ВУЗов. Радиофизика, 1988. Т. 31. № 11. С. 1279.
14. Tinin M. V. et al. //Radio Sci., 1992. V. 27. № 2. P. 245.
15. Kiang Y. W., Lie C. H. //Radio Sci., 1985. V. 20. № 3. P. 652.
16. Апресян Л. А., Кравцов Ю. А. Теория переноса излучения. — М.: Наука, 1983.
17. Гинзбург В. Л. Распространение электромагнитных волн в плазме. — М.: Наука, 1967.
18. Справочник по специальным функциям /под ред. М. Абрамовича и М. Стиган — М.: Наука, 1979.
19. Алимов В. А., Рахлин А. В. //Изв. ВУЗов. Радиофизика, 1996. Т. 39. № 9. С. 1114.
20. Денисов Н. Г. //Изв. ВУЗов. Радиофизика, 1960. Т. 3. № 2. С. 208.
21. Кравцов Ю. А. //Изв. ВУЗов. Радиофизика, 1965. Т. 8. № 5. С. 876.

*При этом влияние регулярной неоднородности слоя является существенным фактором: ослабление интенсивности отраженного от статистически однородного плазменного слоя ($\overline{\varepsilon}(z)\equiv 1$) сигнала при тех же условиях было бы на порядок меньше.

Научно–исследовательский
радиофизический институт,
г. Н. Новгород, Россия

Поступила в редакцию
25 февраля 1998 г.

RADIO WAVE BEAM REFLECTION FROM PLASMA LAYER WITH LARGE–SCALE INHOMOGENEITIES

V. A. Alimov, A. V. Rakhlin

The mean field transport equations, spatial coherence and radio wave beam intensity functions have been obtained for a radio wave beam reflected from the plasma layer with random inhomogeneities using the parabolic equation method. Special attention has been paid to the case of the radio wave beam reflection from a linear plasma layer with large–scale inhomogeneities of electron density. A marked intensity decrease ($\sim$ 3 dB) of a vertical sounding signal reflected from the ionospheric F2 layer is shown to take place only under conditions of abnormally strong disturbances of the ionospheric electron density when using a low-gain transmitting-and-receiving SW antenna.

УДК 533.951, 537.868

О ЗАВИСИМОСТИ АНОМАЛЬНОГО ОСЛАБЛЕНИЯ ПРОБНЫХ ВОЛН ОТ ЧАСТОТЫ ПРИ ВОЗДЕЙСТВИИ МОЩНЫМ РАДИОИЗЛУЧЕНИЕМ НА ИОНОСФЕРУ

С. М. Грач, Г. П. Комраков, М. М. Шварц, М. А. Юрищев

Экспериментально исследована зависимость аномального ослабления пробных волн от частоты f при воздействии мощным радиоизлучением на F-слой ионосферы. Показано, что при $|f - f_0| < 20$ кГц, $f < f_0$ (f_0 — частота волны накачки) аномальное ослабление увеличивается, а при $f \geq f_0$ — уменьшается по сравнению с частотами, более далёкими от f_0. Эффект возникает вследствие вытеснения плазмы и образования ступенеобразного профиля электронной концентрации в области верхнего гибридного резонанса волны накачки.

1. При воздействии на F-область ионосферы мощным КВ радиоизлучением O-поляризации наблюдается вытеснение плазмы из области верхнего гибридного резонанса (ВГР) мощной волны (волны накачки, ВН). На эксперименте этот эффект исследуется с помощью измерений изменения частоты f нескольких отражённых от ионосферы пробных волн (многочастотное допплеровское зондирование) и проявляется как уменьшение f после включения ВН для $f \approx f_{\rm pe}(z_{\rm u}) = \sqrt{f_0^2 - f_{\rm ce}^2}$ [1–3], где f_0 — частота ВН, $f_{\rm ce}$ — электронная циклотронная частота, $f_{\rm pe}(z_{\rm u})$ — электронная плазменная частота на высоте ВГР ВН $z = z_{\rm u}$. Вытеснение плазмы из области ВГР является следствием возбуждения здесь плазменных волн в результате развития тепловой параметрической неустойчивости [4, 5] и, соответственно, увеличения как газокинетического давления плазмы NT из-за её нагрева, так и усреднённого высокочастотного (стрикционного, W) давления плазменных волн [6, 7]. Здесь N и T — соответственно концентрация и температура электронов.

Стрикционное вытеснение плазмы должно приводить к образованию “ступенеобразного"профиля электронной концентрации в ионосфере в области ВГР, когда на некоторой высоте $z_j \approx z_{\rm u}$ формируется скачок плотности плазмы, ниже которого имеет место “плато-— почти однородный участок плазмы размером порядка 1 км [6, 7]. Этот факт, как было показано в [6, 8], должен приводить к характерной зависимости величины аномального ослабления пробных волн $K_{\rm a}$ от их частоты f при $f \approx f_0$ (более подробно см. раздел 3). Напомним, что в возмущённой мощным радиоизлучением ионосфере аномальное ослабление (АО) ВН и пробных волн связано с их рассеянием на мелкомасштабных ($l < c/f_0$) сильно вытянутых вдоль магнитного поля неоднородностях плотности плазмы в плазменные (верхнегибридные) волны [5].

В настоящей работе приводятся результаты экспериментальных исследований влияния ВН на амплитуду (АО) и частоту пробных волн в зависимости от частоты последних. Ниже, во втором разделе, излагаются методика и результаты наблюдений, в третьем разделе проводится сопоставление данных эксперимента с теоретическими расчётами. Краткое обсуждение результатов дано в четвёртом разделе.

2. Эксперименты проводились в 1989–1990 гг. в вечерние и ночные часы, когда влияние линейного поглощения в нижних слоях ионосферы и эффекты дефокусировки в E-области ионосферы мало. Импульсы ВН с частотой $f_0 = 5828$ кГц и длительностью 1 мин. излучались с периодом 10 мин. Достаточно длинные паузы в работе мощного передатчика необходимы для того, чтобы избежать т. н. “перегретого режима"воздействия на ионосферу, при котором сказывается последействие предыдущего включения ВН [9]. В процессе эксперимента исследовались характеристики восьми пробных волн с частотами $|f - f_0| < 50$ кГц. В ряде сеансов сетка частот пробных волн изменялась в паузах между импульсами ВН и попеременно исследовались области частот с $f \sim f_0$ и $f \sim f_{\rm pe}(z_{\rm u})$, в условиях эксперимента $f_0 - f_{\rm pe}(z_{\rm u}) \approx 160$ кГц. Состояние ионосферы в процессе измерений контролировалось с помощью ионограмм, которые регистрировались каждые 15 мин.

Используемая диагностическая аппаратура — установка многочастотного доплеровского зондирования — позволяет измерять как доплеровские сдвиги частоты пробных волн f_{Di}, $i = 1 \div 8$, так и их аномальное ослабление $K_{\rm a}$. Подробное описание установки приведено в [2, 3]. Измерения величин f_D и на сетке частот пробных волн позволяют исследовать изменения профиля электронной концентрации в ионосфере под воздействием волны накачки. Действительно, доплеровский сдвиг частоты и набег фазы φ_D для i-й пробной волны определяются выражениями

$$f_D(f_i,t) = \frac{d}{dt}\left\{\frac{2f_i}{c}\int_{z_e}^{z_{R_i}} n(N(z),f_i)\,dz\right\}, \tag{1}$$

$$\varphi_D(f_i,t) = 2\pi\int_0^t f_D(f_i,\tau)\,d\tau, \tag{2}$$

где z_e — высота “входа"пробных волн в ионосферу, z_{R_i} — высота отражения i-й пробной волны, n — её показатель преломления. Начало интегрирования $t = 0$ в формуле (2) соответствует включению импульса ВН, конец — достижению стационарного состояния. Поскольку $n(z_{R_i}) = 0$, основной вклад в (1) вносят изменения N на пути распространения пробных волн. В силу $\partial n/\partial N < 0$, отрицательные значения f_D и φ_D соответствуют уменьшению N $(dN/dt < 0)$, а положительные — увеличению $(dN/dt > 0)$. При достаточно малом разносе частот пробных волн полученные значения $\varphi_D(t, f_i)$ могут использоваться для восстановления модифицированного профиля электронной концентрации в ионосфере с помощью решения обратной задачи [3].

Величина аномального ослабления $K_{\rm a}$ (в дБ) определяется как отношение стационарных (после окончания переходных процессов) амплитуд отражённой от ионосферы пробной волны во время паузы в излучении ВН ($A_{\rm b}$) и во время её излучения ($A_{\rm p}$):

$$K_{\rm a} = 20\lg(A_{\rm b}/A_{\rm p})\,. \tag{3}$$

На рис. 1 приведена зависимость величины аномального ослабления пробных волн $K_{\rm a}$ от их частоты f, полученная 28.06.89 в 00:00 – 02:00 LT при эффективной излучаемой мощности ВН $P_{\rm ef} = 66$ МВт в условиях стабильной ионосферы. Каждая точка на рис. 1 является результатом усреднения по 10 импульсам ВН. Сплошная линия на рисунке, как и на рис. 2, 3, соответствует модельным расчётам (см. раздел 3). Из рис. 1 видно, что максимальное значение АО $K_{\rm a}^{\max}$ (приблизительно на 3–4 дБ выше среднего по частотам) наблюдается для частоты, несколько меньшей f_0, $\Delta f = f - f_0 = -8$ кГц, в то время как величина $K_{\rm a}(f_0)$ оказывается минимальной — на ≈ 3 дБ ниже среднего. Для бо́льших отстроек Δf какой-либо регулярной зависимости $K_{\rm a}$ от f не наблюдается.

На рис. 2 приведена зависимость $K_{\rm a}$ от f для более широкого диапазона частот пробных волн, полученная 30.06.89 в 23:30 – 00:00 LT при $P_{\rm ef} = 66$ МВт. Для получения такой зависимости сетку частот пробных волн приходилось изменять от одного импульса ВН к другому, и здесь каждая точка является результатом усреднения по 3–7 импульсам ВН. Из рисунка видно, что наряду с повышенными значениями $K_{\rm a}$ при $\Delta f = -8{,}5$ и $-15{,}5$ кГц наблюдается заметное уменьшение $K_{\rm a}$ (на ≈ 2–$2{,}5$ дБ) при $\Delta f \approx -160$ кГц $\approx f_0 - f_{\rm pe}(z_{\rm u})$. На рис. 3 приведена зависимость доплеровского набега фазы φ_D от f при $f \sim f_{\rm p}(z_{\rm u})$, полученная в том же эксперименте [2]. Величина φ_D здесь оценивалась с помощью формулы $\varphi_D = -f_{D_m}\Delta t/2$, где f_{D_m} — максимальное по величине доплеровское отрицательное смещение частоты пробной волны, Δt — характерный интервал времени, в течение которого оно наблюдалось. На рисунке отчётливо наблюдается уменьшение набега фазы пробных волн, инициированное воздействием ВН, $\varphi_D < 0$, причём наибольшие значения $|\varphi_D|$ наблюдаются при

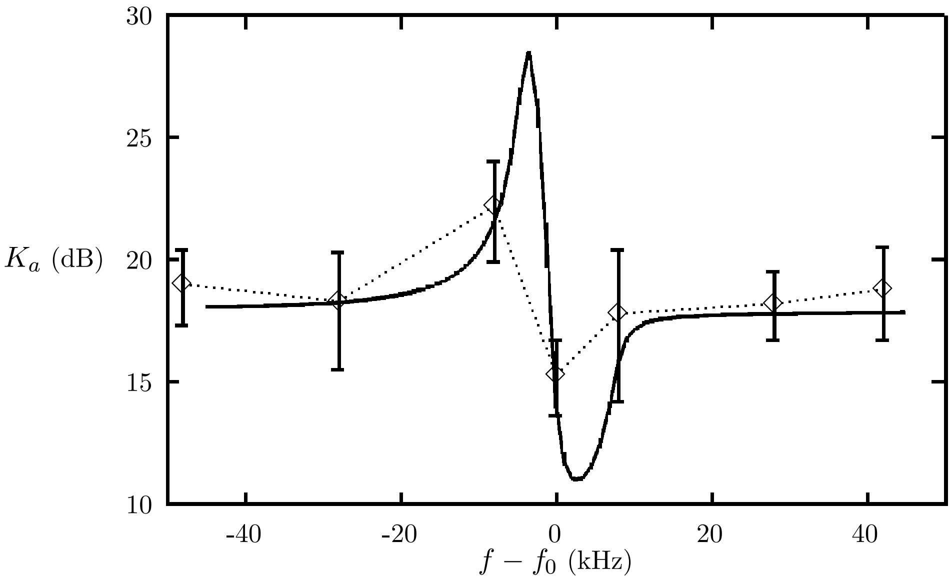


Рис. 1. Зависимость величины аномального ослабления пробных волн $K_{\rm a}$ от их частоты f, полученная 28.06.89 в 00:00 − 02:00 LT при $f_0 = 5828$ кГц, $P_{\rm ef} = 66$ МВт (ромбы, пунктир) и рассчитанная теоретически для модельного профиля (7) (сплошная кривая). Параметры расчёта: $L_0 = = 80$ км, $L_1/L_0 = 3$, $\sqrt{\Delta n^2/N^2} \approx 0{,}99 \cdot 10^{-2}$, $1 - u - v_{\rm p} = 3{,}2 \cdot 10^{-3}$, $z_{\rm p}/L_0 = -3{,}3 \cdot 10^{-3}$, $z_j - z_{\rm p} = 4 \cdot 10^{-3}$.

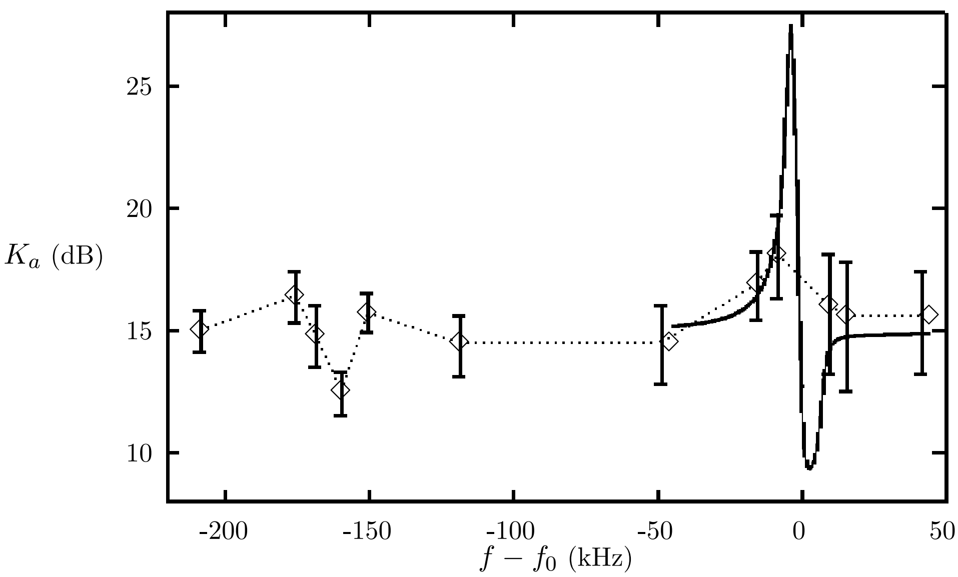


Рис. 2. Зависимость величины аномального ослабления пробных волн $K_{\rm a}$ от f, полученная 30.06.89 в 23:30 − 00:00 LT при $f_0 = = 5828$ кГц, $P_{\rm ef} = 66$ МВт (ромбы, пунктир) и рассчитанная теоретически для модельного профиля (7) (сплошная кривая). Параметры расчета: $L_0 = 80$ км, $L_1/L_0 = 4$, $\sqrt{\Delta n^2/N^2} \approx 0{,}8 \cdot 10^{-2}$, $1 - u - v_{\rm p} = 3{,}2 \cdot 10^{-3}$, $z_{\rm p}/L_0 = -3{,}3 \cdot 10^{-3}$, $z_j - z_{\rm p} = 1{,}2 \cdot 10^{-2}$.

$f \simeq f_{\mathrm{p}}(z_{\mathrm{u}}) \pm 30$ кГц. Отметим, во-первых, что разброс данных на рис. 1—3 определяется вариациями измеряемых величин от одного импульса ВН к другому, характер же зависимостей $K_{\mathrm{a}}(f)$ и $\varphi_D(f)$ сохранялся практически для всех импульсов ВН. Во-вторых, аналогичные зависимости $K_{\mathrm{a}}(f)$ были получены ещё в трёх циклах измерений, проведённых 27.06.89—01.07.89 при $20 < P_{\mathrm{ef}} < 80$ МВт, в измерениях 22.09.91, проведённых в дневные часы при $f_0 = 7815$ кГц и $P_{\mathrm{ef}} = 230$ МВт, в экспериментах 29.04.89 $f_0 = 5828$ кГц, $P_{\mathrm{ef}} = 52$ МВт. В последнем случае частоты пробных волн выбирались в диапазоне $|\Delta f| \leq 5$ кГц. При этом максимальные значения $K_{\mathrm{a}} \approx 23$ дБ наблюдались при $\Delta f = -1{,}5$ кГц, минимальные $K_{\mathrm{a}} \approx 15$ дБ — при $\Delta f = +5$ кГц.

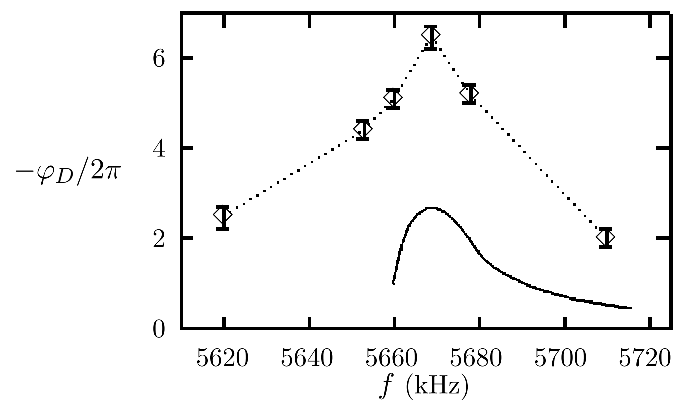


Рис. 3. Зависимость допплеровского набега фазы φ_D от f при $f \sim f_{\mathrm{p}}(z_{\mathrm{u}})$, полученная 30.06.89 в 23:30 – 00:00 LT (ромбы, пунктир) и рассчитанная теоретически для модельного профиля (7) (сплошная кривая). Параметры расчёта те же, что на рис. 2.

Такое поведение $K_{\mathrm{a}}(f)$ наблюдается, однако, не всегда. На рис. 4 приведена зависимость аномального ослабления от частоты при $-200 < \Delta f < 50$ кГц для экспериментов, проведённых 28—29.08.90 в аналогичных ионосферных условиях (высота, масштаб и критическая частота F-слоя, вечерние часы) при $P_{\mathrm{ef}} = 90$ МВт.

Усреднение здесь проводилось по 5—20 импульсам ВН. Зависимость $K_{\mathrm{a}}(f)$ имеет здесь другой характер: увеличения K_{a} при $f < f_0$ не отмечается, в то же время наблюдается некоторое увеличение (на 2—3 дБ) величины K_{a} при $|\Delta f| \leq 5$ кГц, а уменьшение АО при $\Delta f \approx -160$ кГц едва заметно и лежит в пределах доверительного интервала. Величина $\varphi_D/2\pi$ в этих сеансах оставалась положительной и варьировалась в пределах $0 - +10$, причём не отмечено какой-либо регулярной зависимости φ_D от f.

3. Аномальное ослабление радиоволны в возмущённой мощным радиоизлучением области ионосферы определяется, в основном, рассеянием ВН в плазменные (верхнегибридные) волны на мелкомасштабных неоднородностях концентрации плазмы, сильно вытянутых вдоль геомагнитного поля $\vec{B}$. Величина “текущего"(вдоль траектории волны) АО даётся выражением [4, 7]

$$K(\kappa(z)) = 20 \lg \frac{A_{\mathrm{in}}}{A(z)} = 10 \lg e \times \int\limits_{\kappa(z=z_m(\omega,\kappa))}^{\infty} \frac{\pi \omega L(\omega,\kappa) n_{\vec{\kappa}}^2}{2 v_{\mathrm{t}} N^2} d\vec{\kappa}\,, \tag{4}$$

где A_{in} — амплитуда падающей электромагнитной волны, $A(z)$ — её амплитуда в области взаимодействия с плазменными волнами и неоднородностями, $\omega = 2\pi f$ — частота волны, v_{t} — её групповая скорость, $n_{\vec{\kappa}}^2$ — двумерная спектральная интенсивность вытянутых мелкомасштабных неоднородностей, $L(\omega,\kappa) = (dN/N dz)^{-1}$ — локальный масштаб неоднородности ионосферной плазмы (обратный градиент) в точке взаимодействия $z_m(\omega,\kappa)$ электромагнитной волны с плазменными волнами и вытя-

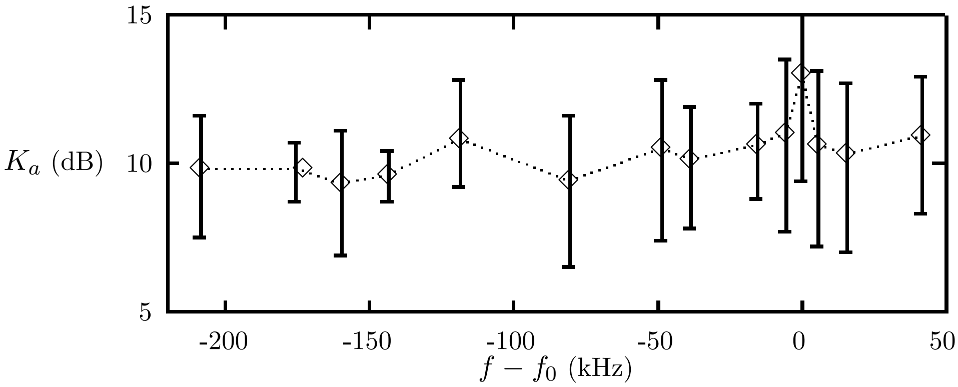


Рис. 4. Зависимость аномального ослабления от частоты при $-200 < \Delta f < 50$ кГц (28–29.08.90) при $f_0 = 5828$ кГц, $P_{\text{ef}} = 90$ МВт.

нутыми неоднородностями с волновым вектором $\vec{\kappa}$ ($\vec{\kappa} \perp \vec{B}$). Полное (измеряемое на Земле) значение АО определяется подстановкой в (4) значения амплитуды отражённой волны A_r (с учётом двукратного прохождения волной области турбулентности):

$$K_{\text{a}} = 20 \lg \frac{A_{\text{in}}}{A_r} = 2K(\kappa_{\min} \simeq \omega/c). \tag{5}$$

Поскольку величина K_{a} из (5) представляет собой дополнительное (аномальное) ослабление электромагнитной волны вследствие её рассеяния на мелкомасштабных неоднородностях, величины АО, определяемые (3) и (5) должны совпадать, когда мало влияние крупномасштабных неоднородностей, возникающих в F-области вследствие самофокусировки ВН, а также нелинейных эффектов в нижних слоях ионосферы (см., например, [5]). Положение точки взаимодействия $z_m(\omega, \kappa)$ определяется дисперсионным уравнением плазменных волн $D(\omega, \vec{k}_{\text{p}}, N(z), B) = 0$ при выполнении условий пространственного синхронизма между электромагнитной волной, плазменными волнами и вытянутыми неоднородностями $k_{\text{t}} = k_{p\parallel}$, $\vec{k}_{p\perp} = \pm\vec{\kappa}$, где $\vec{k}_{\text{t}}, \vec{k}_{\text{p}}$ — волновые векторы соответственно электромагнитной и плазменной волн.

Стрикционное давление плазменных волн приводит, как уже указывалось, к вытеснению плазмы из области рассеяния и образованию “ступенеобразного"профиля электронной концентрации в области ВГР, когда на некоторой высоте $z_j \approx z_{\text{u}}$ наблюдается скачок плотности плазмы, где $L \to 0$. Ниже скачка имеет место “плато-— участок плазмы размером порядка 1 км, где величина L может в несколько раз превышать масштаб регулярной неоднородности в невозмущённой ионосфере. Этот факт, как было показано в [6, 8], должен приводить к характерной зависимости величины аномального ослабления пробных волн K_{a} от их частоты f при $f \approx f_0$. Действительно, в выражении (5) величина $L(\omega, \kappa)$ играет роль весовой функции, увеличивая вклад в АО тех неоднородностей $n_{\vec{\kappa}}$, взаимодействие с которыми имеет место в области плато, и, наоборот, значительно уменьшая вклад тех неоднородностей, взаимодействие с которыми должно происходить в области скачка плотности. При изменении частоты пробной волны изменяется $z_m(\omega, \kappa)$, и рассеяние пробной волны на неоднородностях с тем же κ будет происходить при другом L. Расчёты [8] показали, что для степенного спектра неоднородностей

$$n_{\vec{\kappa}}^2 = \frac{\Delta n^2 l_{\text{T}\perp}}{\pi^2 \kappa (1 + \kappa^2 l_{\text{T}\perp}^2)}, \tag{6}$$

где $l_{\text{T}\perp}$ — поперечная магнитному полю длина теплопроводности в ионосфере, K_{a} увеличивается при $f < f_0$ и уменьшается при $f < f_0$. При частотах f, достаточно далёких от f_0, взаимодействие электромагнитных волн с неоднородностями и плазменными волнами происходит вне ступеньки и её влияние на аномальное ослабление несущественно.

В работах [7, 8] модифицированный профиль электронной концентрации и АО находились как решение самосогласованной стационарной задачи о стрикционном вытеснении плазмы из области возбуждения плазменных волн и неоднородностей, причём пространственный спектр последних считался заданным в форме (6). Отметим, однако, что в рамках стационарной задачи положение скачка на профиле определить не удаётся,* поскольку существует целая область возможных положений скачка на профиле, причём в случае “верхнего"скачка размер плато и перепад концентрации в скачке максимальны, в случае “нижнего-— минимальны (см. рис. 5). В настоящей работе для сопоставления с данными эксперимента нами используется модельное кусочно—линейное представление модифицированного профиля:

$$N = \begin{cases} N_{\rm p}[1 + (z - z_{\rm p})/L_0] & \text{при } z < z_{\rm p},\ z > z_j, \\ N_{\rm p}[1 + (z - z_{\rm p})/L_1] & \text{при } z_{\rm p} < z < z_j, \end{cases} \tag{7}$$

при этом масштаб регулярной неоднородности невозмущённой ионосферы L_0 брался из анализа ионограмм, а размер плато $\Delta z = z_j - z_{\rm p}$, позиция скачка z_j и масштаб регулярной неоднородности в области плато L_1 подбирались таким образом, чтобы, с одной стороны, достигнуть наилучшего согласия с данными эксперимента, а с другой — чтобы профиль (7) был близок к допустимым решениям стационарной самосогласованной задачи с параметрами, близкими к условиям эксперимента. Величина $N_{\rm p}$ в (7) соответствует электронной концентрации на нижнем крае плато при $z = z_{\rm p}$.

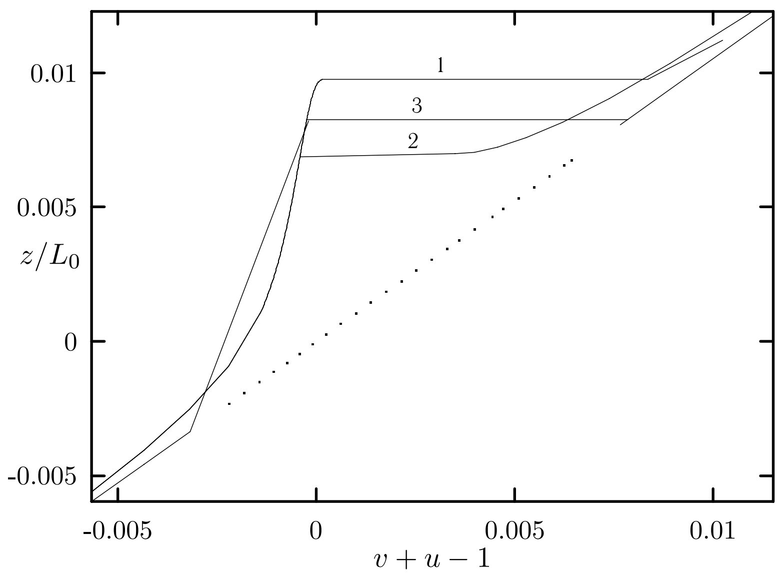


Рис. 5. Результаты расчётов самосогласованного профиля в случае верхнего (кривая *1*) и нижнего скачков (кривая *2*) и модельный профиль (7) (кривая *3*) при $L_0 = 80$ км, $L_1/L_0 = 4$, $P_{\rm ef} = 50$ МВт, $h = 300$ км, $\gamma = 300$ с$^{-1}$, $\sqrt{\Delta n^2/N^2} \approx 0{,}8 \cdot 10^{-2}$, $u = f_{\rm ce}^2/f_0^2 = 0{,}05$, $v_{\rm p}$, $z_{\rm p}$ и z_j — те же, что для рис. 2.

Результаты расчёта аномального ослабления пробных волн по формулам (4), (5) для спектра неод-

*В работе [6] показано, что в случае, когда групповая скорость плазменных волн направлена вверх по z ($v_\parallel > 0$), реализуется верхний скачок. В то же время, из работы, например, [4], следует, что возникающие вследствие развития тепловой параметрической неустойчивости в области ВГР плазменные волны заполняют весь возможный фазовый объём как с $v_\parallel > 0$, так и с $v_\parallel < 0$.

нородностей (6) в профиле (7) представлены на рис. 1, 2 сплошными линиями, для следующих параметров: $u == \dfrac{f_{\rm ce}^2}{f_0^2} = 0{,}05$, L_0 =80 км, для рис. 1 (28.06.89) $\dfrac{L_1}{L_0} = 3$, $\sqrt{\dfrac{\Delta n^2}{N^2}} \approx 0{,}99 \cdot 10^{-2}$, $1 - u - v_{\rm p} = 3{,}2 \cdot 10^{-3}$, $\dfrac{z_{\rm p}}{L_0} = -3{,}3 \cdot 10^{-3}$, $\dfrac{z_j - z_{\rm p}}{L_0} = 4 \cdot 10^{-3}$; для рис. 2 (30.06.89) $\dfrac{L_1}{L_0} = 4$, $\sqrt{\dfrac{\Delta n^2}{N^2}} \approx 0{,}8 \cdot 10^{-2}$, $1 - u - v_{\rm p} = 3{,}2 \cdot 10^{-3}$, $\dfrac{z_{\rm p}}{L_0} = -3{,}3 \cdot 10^{-3}$, $\dfrac{z_j - z_{\rm p}}{L_0} = 1{,}2 \cdot 10^{-2}$. Здесь $v_{\rm p} = 2e^2 N_{\rm p}/(\pi m f_0^2)$. На рис. 5 приведены результаты расчётов самосогласованного профиля в случае “верхнего"и “нижнего"скачков (сплошные линии) и модельный профиль (7) для параметров эксперимента 30.06.89 (рис. 2, 3). Для расчёта самосогласованного профиля использовались следующие величины: $L_0 = 80$ км, $P_{\rm ef} = 50$ МВт, высота ступеньки над поверхностью Земли $h = 300$ км, декремент затухания плазменных волн $\gamma = 300$ с $^{-1}$, $\sqrt{\Delta n^2/N^2} \approx 0{,}8 \cdot 10^{-2}$, $l_{\rm T\perp}^2 = 2{,}5$ м. Из рис. 1, 2 видно, что, во-первых, зависимость АО от частоты пробных волн, полученная в [8], имеет место на эксперименте, и, во-вторых, соответствующим подбором параметров удаётся достичь неплохого количественного согласия теории с экспериментом.

На рис. 3 представлена сплошной кривой зависимость набега фазы пробной волны, дополнительного по сравнению с невозмущённым профилем, от её частоты в профиле (7) с параметрами, соответствующими эксперименту 30.06.89. Вычисления проводились по формуле

$$\frac{\varphi}{2\pi} = -\frac{2f}{c} \int\limits_{z_{\rm p}}^{z_R} n\, dz\,, \tag{8}$$

где n — показатель преломления “обыкновенной"пробной волны, z_R — её точка отражения, в которой $f_{\rm pe} = f$. Видно, что расчётный набег фазы оказывается приблизительно на $\approx 4\pi$ меньше экспериментально измеренного, хотя зависимость $\varphi(f)$ подобна экспериментально измеренной. Причины такого расхождения обсуждаются в следующем разделе.

4. Таким образом, в серии экспериментов по воздействию на ионосферу мощной радиоволной обыкновенной поляризации, проведённых в июне—июле 1989 г., была отмечена характерная зависимость аномального ослабления $K_{\rm a}$ пробных волн от частоты f при $\Delta f < 20$ кГц (f_0 — частота накачки): при $f < f_0$ наблюдались повышенные значения $K_{\rm a}$ по сравнению со средним значением, а при $f \geq f_0$ — пониженные. Кроме того, в экспериментах 30.06.89, когда использовались пробные волны в более широком диапазоне частот, были отмечены отрицательные сдвиги частоты (и фазы) пробных волн. Обнаружено также уменьшение их аномального ослабления при $f \approx f_{\rm pe}(z_{\rm u})$. Подобные явления отмечались также в ряде других подобных экспериментов.

Все три эффекта имеют одну и ту же причину — вытеснение плазмы из области верхнего гибридного резонанса волны накачки. Такое вытеснение связано с увеличением газокинетического давления плазмы, вследствие её нагрева плазменными волнами, и со стрикционным давлением плазменных волн. Как показано выше, характер зависимости АО от частоты определяется образованием ступенеобразного профиля электронной концентрации, связанного со стрикционным вытеснением. Такого стрикционного вытеснения недостаточно, однако, чтобы объяснить наблюдаемые величины φ_D, здесь существенный вклад вносит также тепловое вытеснение, которое проявляется на масштабах, существенно бóльших размера плато Δz. Действительно, если расчётную величину $|\varphi|$ на рис. 3 увеличить на $\Delta\varphi \approx 4\pi$, то совпадение расчётных значений с данными эксперимента станет удовлетворительным. Дополнительный набег фазы $\Delta\varphi$ связан, скорее всего, с тепловым выдавливанием.

Вытеснение плазмы наиболее эффективно в центре диаграммы направленности возмущающего передатчика, где интенсивность ВН максимальна, и плавно убывает к её краям. В результате для пробных волн, отражающихся в области пониженной плотности, формируется вогнутое зеркало (плазменный рефлектор, [1, 6]), которое должно приводить к фокусировке отражённого сигнала, т. е. увеличению его амплитуды. С таким рефлектором связано, по-видимому, увеличение амплитуды пробных волн (уменьшение $K_{\rm a}$ на 3–4 дБ) при $f \approx f_{\rm pe}(z_{\rm u})$. Отметим, что близкие значения увеличения амплитуды наблюдались в эксперименте [1].

В то же время, в экспериментах 28–29 августа 1990 г., проведённых в аналогичных условиях (высота, масштаб и критическая частота F-слоя ионосферы, частота и мощность волны накачки) наблюдается существенно иная зависимость $K_{\rm a}(f)$ (см. рис. 4). При этом сама величина $K_{\rm a}$ оказывается несколько меньшей, а допплеровские сдвиги частоты f_D и фазы φ_D оказываются положительными. Последнее соответствует увеличению концентрации плазмы вследствие её ионизации электронами, ускоренными плазменной турбулентностью [10–13]. Увеличение N происходило приблизительно равномерно по высоте (φ_D не зависит от f), поэтому в (4), (5) L и, следовательно, $K_{\rm a}$ не зависят от ω. В настоящее время трудно определённо указать причину преобладания дополнительной ионизации ускоренными электронами над вытеснением плазмы из области ВГР ВН в последнем эксперименте при внешне аналогичных условиях. Одной из причин может служить наличие в ионосфере повышенного фона надтепловых электронов с энергиями 1–5 эВ, которые существенно более эффективно ускоряются плазменной турбулентностью [13]. Однако какие-либо регулярные способы мониторинга таких низкоэнергичных частиц при проведении ионосферных экспериментов отсутствуют. Косвенным свидетельством в пользу предложенного объяснения может служить некоторое (на 1–2 порядка) повышение потока протонов с энергиями 4–8 МэВ на орбите Земли в период с 25 по 27 августа 1990 г., зарегистрированное на спутнике “Goals", максимум возмущения пришёлся на 6:00 UT 26.08.90. В это же время наблюдалось некоторое повышение Kp-индекса магнитной активности [14]. В то же время, возможная связь отмеченных явлений с повышением фона надтепловых электронов в F-области ионосферы через 1,5–2 суток требует дальнейшего изучения.

Авторы благодарят В. В. Васькова за обсуждение постановки эксперимента, О. А. Шейнер за предоставление и обсуждение солнечных данных. Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант РФФИ № 96–02–18499) и гранта INTAS-RFBR 95–0434.

ЛИТЕРАТУРА

1. Васьков В. В., Голян С. Ф., Гуревич А. В. и др. //Письма в ЖЭТФ, 1986. Т. 43. С. 512.
2. Березин И. В., Белянский В. Б. и др. //Геомагнетизм и аэрономия, 1991. Т. 31. С. 874.
3. Lobachevsky L. A., Gruzdev Yu. V., et al. //J. Atm. Terr. Phys., 1992. V. 54. P. 75.
4. Грач С. М., Митяков Н. А., Рапопорт В. О., Трахтенгерц В. Ю. — В кн.: Тепловые нелинейные явления в плазме. — Горький: ИПФ АН СССР, 1979. С. 46.
5. Митяков Н. А., Грач С. М., Митяков С. Н. Возмущение ионосферы мощными радиоволнами. — М.: ВИНИТИ, 1989.
6. Васьков В. В., Димант Я. С. //Геомагнетизм и аэрономия, 1989. Т. 29. № 3. С. 417.
7. Грач С. М., Митяков Н. А., Шварц М. М. //Геомагнетизм и аэрономия, 1989. Т. 29. № 4. С. 590.
8. Грач С. М., Шварц М. М. //Геомагнетизм и аэрономия, 1990. Т. 30. № 6. С. 1008.
9. Ерухимов Л. М., Фролов В. Л. Динамические и спктральные характеристики искусственного радиоизлучения ионосферы //Препринт НИРФИ № 185. — Горький, 1984.
10. Белякова В. Н., Березин И. В. и др. //Геомагнетизм и аэрономия, 1991. Т. 31. № 3. С. 466.
11. Grach S. M., Komrakov G. P. et al. //Phys. Rev. Lett., 1997. V. 78. № 5. P. 883.
12. Gurevich A. V., Dimant Ya. S. et al. //J. Atm. Terr. Phys., 1985. V. 47. P. 1057.
13. Грач С. М., Митяков Н. А., Трахтенгерц В. Ю. //Физика плазмы, 1986. Т. 12. С. 693.
14. Solar-Geophysical Data.

Научно–исследовательский
радиофизический институт,
г. Н. Новгород

Поступила в редакцию
21 августа 1997 г.

ON FREQUENCY DEPENDENCE OF PROBE WAVE ABNORMAL ATTENUATION ROHILE MODIFYING THE IONOSPHERE BY POWERFUL RADIO EMISSION

S. M. Grach, G. P. Komrakov, M. M. Shwarts, M. A. Yurishchev

The dependence of probe wave abnormal attenuation on frequency f has been experimentally studied when modifying the ionospheric F-layer by powerful radio emission. It has been shown that at $|f - f_0| <$ 20 kHz, $f < f_0$ (f_0 is the pump frequency) the abnormal attenuation increases and at $f \geq f_0$ it decreases as compured with frequencies far from f_0. The effect arises due to plasma displacement and the formation of electron density stepprofile near the upper hybrid resonance of the pump wave.

УДК 533.951

МЕХАНИЗМ ГЕНЕРАЦИИ НАДТЕПЛОВЫХ ЧАСТИЦ УДАРНОЙ ВОЛНОЙ В ПЛАЗМЕ

И. Р. Смирновский

Исследуются эффекты диссипации во фронте плазменного “предвестника"ударной волны в слабоионизированной плазме без учета влияния магнитного поля. Рассматриваются возможные механизмы рассеяния энергии отраженных от потенциального барьера ионов. “Скачок"потенциала на фронте предвестника обусловлен дисперсией ионного звука, имеет пространственный масштаб порядка дебаевского радиуса D. Оценивается длина релаксации импульса отраженных ионов в случае их рассеяния на электронах, которая оказалась много больше D и много меньше длины свободного пробега электронов. Высказывается предположение, что ионизация и свечение газа может быть результатом существенно неравновесных эффектов образования надтепловых электронов и их быстрой релаксации. Аналогичный процесс в распадающейся плазме вызывает усиленное индуцированное излучение.

1. В работах [1–4] эффекты предыонизации и свечения газа перед ударной волной (УВ) в плазме связывают с появлением там надтепловых или “горячих"электронов. При этом конкретный механизм их генерации часто не указывается [1, 3, 4]. Допускается даже известное смешение представлений о таких различных явлениях как возникновение надтепловых частиц и квазиравновесное увеличение электронной температуры T_e [2, 3] за счет, например, теплопроводности или плазменной турбулентности. То, что эффекты ионизации и свечения впереди УВ вызываются энергичными электронами, доказывается прямыми экспериментами [3, 4] и косвенными фактами — спектральным составом свечения [1].

Обычно появление надтепловых частиц связывают с плазменной турбулентностью в ударном слое или в области за фронтом УВ [5]. Обоснование наличия или отсутствия турбулентности требует тщательного исследования. Пример такого анализа данных эксперимента содержится в [2], где показано, что в некоторых случаях при выбросе с борта космоплана воды турбулентность все-таки генерируется. Тем не менее в [2] нет удовлетворительного объяснения, почему вспышки света [6] и увеличение отражения радиоволн от окружающей ионосферы [7] возникают именно в моменты включения и выключения систем выброса. Более того, установлено [2], что эти эффекты часто не сопровождались увеличением интенсивности плазменной турбулентности в данные моменты времени.

Ниже предлагается интерпретация подобных явлений в случае слабоионизированной изотропной плазмы.

2. Полет метеорита в E-области ионосферы сопровождается сильной УВ нейтрального компонента. Идеализируя условия задачи, исследуем воздействие плоской УВ, распространяющейся с постоянной скоростью c_p в слабоионизированной изотропной плазме, на плазменный компонент. Перед ее фронтом за счет электронной теплопроводности возникает, так называемая, прогревная область с пространственным масштабом $\Delta = l_e(v_{Te}/c_p)$ [8], где l_e — длина свободного пробега, v_{Te} — тепловая скорость электронов (величины, относящиеся к различным компонентам плазмы здесь и ниже помечены индексами “e”, “i” или “n”). Поскольку в этой области температура ионов T_i практически не изменяется, имеем $T_e \gg T_i$. Следовательно, в прогревной зоне возникает ионно–звуковое плазменное возмущение [9] — “предвестник". Если поля, описывающие плазменное возмущение, зависят только от комбинации $c_p t - r$, можно ввести число Маха предвестника M_s как отношение скорости c_p к скорости ионного звука $V_s = \sqrt{T_e/m_i}$. Температура электронов T_e примерно равна температуре газа в равновесной области за фронтом УВ. Тогда M_s при $c_p \gg v_{Ti}$ оценивается следующим образом:

$$M_s \simeq \frac{\gamma + 1}{\sqrt{2(\gamma - 1)}}\,, \tag{1}$$

где γ — показатель адиабаты для нейтрального компонента. Таким образом, $M_s > 1$. Это соответствует условию “опрокидывания"гиперболических волн [10—12]. Такая ситуация, в принципе, отлична от рассмотренной в [9], где предполагалось $M_s \ll 1$.

Сглаживание разрыва, образовавшегося в ионно—звуковом предвестнике, можно объяснить ионной вязкостью [13]. Однако в случае сильной нелинейности ($c_p \gg v_{Ti}$) пригодность газодинамического описания ударного профиля достаточно проблематична, поскольку ширина фронта оказывается порядка длины свободного пробега l_i. Наконец, эффекта опрокидывания не возникает в силу дисперсии ионного звука.

Структура фронта нелинейной волны, определяемая дисперсией и тепловым разбросом скоростей ионов, строится [14, 15] на основе кинетического уравнения Власова для электронов и ионов совместно с уравнением Пуассона для потенциала самосогласованного электрического поля. В масштабе дебаевского радиуса D, много меньшего длины свободного пробега ионов, ударное возмущение состоит из четырех характерных пространственных областей: невозмущенная плазма, подножие (квазигоризонтальный участок профиля), образованное двумя потоками ионов, собственно “ударный"переход толщиной порядка D и слабозатухающая колебательная структура полей за фронтом волны.

Стационарный профиль ионно—звуковой волны формируется при условии диссипации импульса отраженного потока ионов. При этом возможно многократное отражение иона от потенциального барьера. Поля на подножии выражаются через невозмущенные значения с помощью законов сохранения числа частиц, импульса и энергии. Для случая сильно ионизированной равновесной плазмы это показано в [16], где предполагалось, что энергия потока отраженных ионов полностью переходит в тепловую энергию электронов. Такого рода стационарное ионно—звуковое возмущение представляет собой своеобразную УВ.

3. Здесь мы не будем останавливаться на математических доказательствах существования стационара, описанного в п. 2. Представляет интерес обсудить, каким образом происходит релаксация неравновесного состояния плазмы на подножии.

Один из возможных механизмов связан с развитием плазменной нестабильности и возникновением турбулентности. Например, в работе Сагдеева [17] исследуется эффект рассеяния потока ионов на флуктуациях магнитного поля, вызванных анизотропной неустойчивостью. При этом результат воздействия поля на ионы подобен действию столкновений: энергия направленного движения ионов отраженного потока преобразуется в их тепловую энергию. Характерная длина релаксации отраженных ионов в этом случае равна [17] $\dfrac{m_i}{m_e}\,\dfrac{c}{\omega_{pi}}$ ($\omega_{pi} = \sqrt{4\pi e^2\, n/m_i}$ — ионная плазменная частота, n — концентрация плазмы, c — скорость света). Этот масштаб становится меньше Δ только при сильной неизотермичности плазмы

$$\frac{T_e}{T_i} > \left(\frac{c}{c_p}\right)^{2/3} \frac{m_i}{m_e}\,.$$

Турбулентность рассматривается [18, 19] как наиболее вероятный механизм нагрева плазмы в околоземной УВ. В этих условиях развивается ион—ионная неустойчивость коротковолновых ионно—ленгмюровских и ионно—звуковых колебаний [18]. В слабоионизированной плазме подобные неустойчивости должны быть сильно демпфированы за счет столкновений ионов с нейтральными частицами.

Уровень длинноволновых ионно—звуковых флуктуаций на подножии не превышает тепловой, поскольку двухпотоковая неустойчивость относительно ионно—звуковых колебаний не имеет места. В самом деле, условие возникновения неустойчивости [20]

$$\frac{V_s}{u_f - u_r} > 1$$

не выполняется, т. к. в нашем случае $u_f - u_r \simeq 2c_p$ и $M_s > 1$, согласно (1).

Другой сценарий релаксации не предполагает возникновения турбулентности. Надтепловые ионы отраженного потока испытывают торможение в наведенном ими продольном электрическом поле поляризации. Это поле волны тормозит быстрые ионы и ускоряет фоновые частицы (главным образом электроны). Так "работает"механизм трения за счет кулоновских столкновений. Длина столкновительной релаксации потока отраженных ионов на электронах оценивается на основе [20] при $T_e \gg T_i$ и $(u_f - u_r) \geq V_s$ (u — скорость в системе волны, индекс "r" относится к отраженному, а "f" — к падающему потокам) следующим образом: $l_d = l_e(2M_s)^4(T_i/T_e)$. Существенно, что l_d отлично от длины свободного пробега ионов, которая оценивает длину релаксации импульса частиц при малых отклонениях от равновесия. С другой стороны, l_d много меньше характерного масштаба выравнивания электронной и ионной температур, равного $l_e\sqrt{m_i/m_e}$. Это обусловлено неизотермичностью плазмы в прогревной зоне.

Ниже пренебрежем влиянием на структуру УВ эффектов неустойчивости в силу наличия эффективного механизма торможения ионов полем поляризации.

4. За счет возникновения продольного электрического поля поляризации импульс отраженных ионов передается электронам плазмы. Энергия, теряемая "быстрым"ионом в результате дальних столкновений с электронами, оценивается при $T_e > T_i$ и $v \geq V_s$ ($v = u_f - u_r$ — скорость иона в лабораторной системе отсчета) следующим образом [20]:

$$\left(\frac{d\mathcal{E}}{dt}\right)^{\mathrm{e}} = -\frac{mv^2}{2} \cdot \frac{3}{2}\sqrt{\frac{\pi}{2}}\,\frac{V_s}{l_e}\,\frac{T_e}{T_i}\,\left(\frac{V_s}{v}\right)^3.$$

Легко получить оценку для скорости электрона, приобретенной им в поле поляризации E_p, на основе одномерного уравнения движения в системе координат, связанной с волной,

$$m_e\ddot{x} = eE_p\,,$$

$$E_p = \left(\frac{d\mathcal{E}}{dt}\right)^{\mathrm{e}} \frac{1}{2Zc_p}\,,$$

где Z — зарядовое число иона. Предполагая начальную скорость электрона $\dot{x}(0) = 0$, в конце участка взаимодействия в момент времени Δt $(x(\Delta t) = l_d)$ для v^* имеем

$$v^* = c_p\sqrt{1 + 16\pi Z m_i/m_e}\,.$$

Таким образом, для отношения v^* к тепловой скорости электронов в горячей электронной зоне получаем искомую оценку

$$\frac{v^*}{v_{Te}} \simeq \frac{5c_p}{V_s} \sim 10.$$

Следовательно, в результате рассеяния надтепловых ионов $(2c \gg v_{Ti})$ на электронах возникают надтепловые электроны. Концентрация отраженных ионов составляет порядка 5% от концентрации падающих частиц [16]. Это значение можно считать оценкой "сверху"и для концентрации надтепловых электронов.

Спектр быстрых электронов сильно размыт вследствие их эффективного рассеяния на флуктуациях электрического поля и процессов деградации. В результате последних увеличивается ионизация плазмы. При некоторых условиях не исключено возникновение токовой неустойчивости, приводящей к лавинообразному росту ионизации и увеличению T_e (здесь такая ситуация не анализируется). Другое следствие процессов формирования спектра надтепловых электронов — образование перед фронтом УВ неравновесной относительно внутренних степеней свободы частиц области, в которой возможно усиление света. Механизм возникновения инверсной заселенности уровней аналогичен таковому в плазменном лазере [21], в котором неравновесное состояние обусловлено движением вниз по дискретным уровням рекомбинационного потока электронов.

Интерпретация возникновения надтепловых электронов на основе представлений о неустойчивости плазменных волн обусловлена существованием фазовых скоростей, больших тепловой скорости электронов (в отличии от “разогрева"теплового ядра функции распределения турбулентными волнами с фазовыми скоростями порядка тепловой скорости). В случае раскачки магнитозвуковых колебаний оказалось [22], что наблюдаемые параметры спектра надтепловых электронов достигаются уже на квазилинейной стадии развития неустойчивости. Таким образом, турбулентность не является необходимой для генерации надтепловых частиц. Кроме того, в слабоионизированной изотропной плазме затруднено или невозможно развитие неустойчивостей, возникающих в околоземной УВ.

В силу дальнодействующего характера кулоновских столкновений, рассматриваемый механизм достаточно эффективен в слабоионизированной плазме. Потери энергии отраженных ионов в результате упругих столкновений с нейтралами равны

$$\left(\frac{d\mathcal{E}}{dt}\right)^{\mathrm{n}} = -\frac{mv^2}{2}\cdot\frac{3}{2}\sqrt{\frac{\pi}{2}}\,\nu_{in}\frac{v}{v_{Ti}}\,.$$

При выполнении неравенства

$$\frac{D}{l_{in}} < g\left(\frac{c_p}{v_{Ti}}\right)^3\left(\frac{c_p}{v}\right)^3$$

($g = (nD^3)^{-1}$ — плазменный параметр), что имеет место на высотах Е-области ионосферы, наличие нейтрального компонента пренебрежимо мало сказывается на скорости диссипации отраженных ионов.

Таким образом, в слабоионизированной изотропной плазме возможен механизм образования надтепловых электронов за счет энергии отраженных от УВ ионов, не связанный с появлением неустойчивости. Усиленное свечение и повышение ионизации в окружающей плазме на больших расстояниях перед фронтом УВ при этом объясняется наличием рекомбинационного потока электронов.

Автор благодарен В. А. Павлову за внимание к работе. Исследование выполнено при поддержке РФФИ № 96–05–64723.

ЛИТЕРАТУРА

1. Ольховатов А. Ю. //Геомагнетизм и аэрономия, 1990. Т. 30. № 1. С. 161.
2. Ольховатов А. Ю. //Геомагнетизм и аэрономия, 1992. Т. 32. № 4. С. 64.
3. Пункевич Б. С., Степанов Б. М. — В сб.: Физические процессы при горении и взрыве. — М.: Атомиздат, 1980. С. 151.
4. Пункевич Б. С., Степанов Б. М. — В сб.: Физические процессы при горении и взрыве. — М.: Атомиздат, 1980. С. 166.

5. Tidmann D. A., Krall N. A. Shock waves in collisionless plasmas. — New York, 1971.
6. Covault C. //Aviation Week and Space Techn., 1991. V. 134. № 18. P. 18.
7. Bernhard P. A., Swartz W. E., Kelly M. C. et. al. //Astroph. Lett. Commun., 1988. V. 27. № 3. P. 183.
8. Великович А. Л., Либерман М. А. Физика ударных волн в газах и плазме. — М.: Наука, 1987. С. 74.
9. Авраменко Р. Ф., Рухадзе А. А., Теселкин С. Ф. //Письма в ЖЭТФ, 1981. Т. 34. Вып. 9. С. 485.
10. Уизем Дж. Линейные и нелинейные волны. — М.: Мир, 1977.
11. Павлов В. А., Смирновский И. Р. //ПМТФ, 1993. Т. 34. № 5. С. 3.
12. Pavlov V. A., Smirnovski I. R. — In: Abstracts XXIY Gener. Assemb. of URSI, Kyoto, Japan. 1993.
13. Павлов В. А. //Физика плазмы, 1996. Т. 22. № 2. С. 182.
14. Алиханов С. Г., Белан В. Г., Кичигин Г. Н., Чеботаев П. З. //ЖЭТФ, 1971. Т. 60. Вып. 3. С. 982.
15. Бардаков В. М., Морозов А. Г., Шухман И. Г. //Физика плазмы, 1975. Т. 1. Вып. 6. С. 955.
16. Леденев В. Г. //ПМТФ, 1990. Т. 30. № 2. С. 17.
17. Сагдеев Р. З. — В сб.: Вопросы теории плазмы /Под ред. М. А. Леонтовича. — М.: Атомиздат, 1964.
18. Строкин Н. А. //ЖЭТФ, 1985. Т. 88. С. 2005.
19. Балихин М. А., Бородкова Н. Л., Вайсберг О. Л., Галеев А. А. и др. //Физика плазмы, 1988. Т. 14. С. 1326.
20. Электродинамика плазмы /Под ред. А. И. Ахиезера. — М.: Наука, 1974.
21. Гордиец Б. Ф., Осипов А. И., Шелепин Л. А. Кинетические процессы в газах и молекулярные лазеры. — М.: Наука, 1980. С. 445.
22. Вайсберг О. Л., Галеев А. А. //ЖЭТФ, 1985. Т. 85. С. 1232.

НИИ Радиофизики
Санкт-Петербургского
государственного университета,
Россия

Поступила в редакцию
30 июля 1997 г.

MECHANISM OF EPITHERMAL PARTICLE GENERATION BY A SHOCK WAVE IN PLASMA

I. R. Smirnovskij

The dissipation effects are studied in the front of the shock wave plasma “precursor"in a weakly ionized plasma without regard to the magnetic field. Possible energy dissipation mechanisms are considered for the ions reflected from the potential barrier. The potential “jump"at the precursor front is due to the ion sound dispersion, it has spatial scale of about Debye radius D. The relaxation pulse length of reflected ions scattered by electrons is estimated which turned out to be much more than D and much less than electron free path. The suggestion has been made that gas ionization and fluorescence may be a result of essentially nonequilibrium effects of epithermal electron formation and fast relaxation. A similar process in decomposing plasma produces strong induced emission.

УДК 533.951

ГЕНЕРАЦИЯ ВТОРОЙ ГАРМОНИКИ АЛЬФВЕНОВСКОЙ ВОЛНЫ В РЕЗУЛЬТАТЕ РАЗВИТИЯ ВЗРЫВНОЙ НЕУСТОЙЧИВОСТИ В СИСТЕМЕ ПОТОК — ПЛАЗМА

А. Е. Белянцев, В. П. Дворяковский, С. М. Файнштейн, Е. А. Чернова

В статье изучена возможность ап-конверсии и взрывной неустойчивости в системе поток — магнитоактивная плазма. Выяснены условия синхронизма, энергии мод, выведены укороченные уравнения для медленно меняющихся амплитуд мод. Показано, что “взрыв"стабилизируется за счет генерации слабо затухающих гармоник альфвеновской волны.

ВВЕДЕНИЕ

Известно (см., например, [1—3]), что в неравновесных средах, какой является система поток—плазма, возможна взрывная или ВЧ-неустойчивость. Взрывная нестабильность реализуется, если волна высшей частоты (или две низшие) [4] имеет отрицательную энергию, а две низшие моды резонансного триплета имеют положительную энергию. ВЧ-неустойчивость (или ап-конверсия) имеет место, если волна промежуточной частоты имеет отрицательную, а две другие — положительную энергии [5, 6]. Взрывная неустойчивость стабилизируется либо нелинейным сдвигом частоты [1], либо нелинейным поглощением [7, 8]. В [7, 8] впервые решена задача о стабилизации “взрыва"за счет генерации затухающих гармоник волны положительной энергии, что эквивалентно нелинейному поглощению. В данной работе развивается эта идея на примере генерации второй слабо затухающей гармоники альфвеновской волны в системе поток нагретого ионизированного газа, пронизывающего “холодную"плазму. В МГД-приближении выведены и решены укороченные уравнения для комплексных амплитуд мод. Найдены амплитуды гармоник альфвеновской волны. Выяснены условия возникновения взрывной и ВЧ-неустойчивости. Полученные результаты представляют несомненный интерес как для лабораторной, твердотельной плазмы [9], так и для интерпретации наблюдаемого космического излучения в атмосфере Солнца, звезд и т. д. [10]. Кроме того, решенная задача полезна для создания генераторов (усилителей) МГД-излучения и инерционного УТС [11].

1. ПОСТАНОВКА ЗАДАЧИ, ВЫВОД УКОРОЧЕННЫХ УРАВНЕНИЙ

Рассмотрим нагретый до температуры T полностью ионизованный газ, пронизывающий холодную плазму. Материнскую плазму считаем имеющей конечную проводимость σ (считается, что проводимость порядка квадратичной нелинейности), а проводимость потока $\sigma_s \gg \sigma$ ($\sigma_s \to \infty$). Причем исследуем одномерный случай, т. е. волны распространяются вдоль внешнего магнитного поля $\vec{H}_0$. Исходная система уравнений имеет вид [12]

$$\operatorname{rot}\vec{E} = -\frac{1}{c}\frac{\partial \vec{H}}{\partial t}, \quad \operatorname{rot}\vec{H} = \frac{4\pi}{c}(\vec{j}+\vec{j}_s), \quad \vec{j} = \sigma\left\{\vec{E} + \frac{1}{c}[\vec{v},\ \vec{H}+\vec{H}_0]\right\},$$

$$\vec{j}_s = \sigma_s\left\{\vec{E} + \frac{1}{c}[\vec{v}_s+\vec{V}_0,\ \vec{H}+\vec{H}_0]\right\}, \quad \frac{\partial \vec{v}}{\partial t} + (\vec{v}\nabla)\vec{v} = \frac{1}{c(\rho+\rho_0)}[\vec{j},\ \vec{H}+\vec{H}_0], \tag{1}$$

$$\frac{\partial \vec{v}_s}{\partial t}+(\vec{v}_s+\vec{V}_0,\nabla)\vec{v}_s=\frac{1}{c(\rho_{0s}+\rho_s)}\left[\vec{j}_s,\,\vec{H}+\vec{H}_0\right]-\frac{c_{\text{т}}^2\nabla\rho_s}{\rho_{0s}+\rho_s}\,,$$
$$\frac{\partial \rho}{\partial t}+\operatorname{div}(\rho+\rho_0)\vec{v}=0,\qquad \frac{\partial \rho_s}{\partial t}+\operatorname{div}(\rho_{0s}+\rho_s)(\vec{V}_0+\vec{v}_s)=0,$$

где $\vec{E}$, $\vec{H}$, $\vec{v}$, ρ, $\vec{v}_s$, ρ_s — соответственно отклонения электрического, магнитного полей, скорости и плотности плазмы, скорости и плотности частиц пучка от равновесных значений 0, $\vec{H}_0$, 0, ρ_0, $\vec{V}_0$, ρ_{0s}; $c_{\text{т}}$ — скорость звука, $\vec{j}$, $\vec{j}_s$ —плотности тока частиц плазмы и пучка. Линеаризуем (1) вплоть до квадратичной нелинейности и перейдем к безразмерным переменным:

$$t_\delta=t\Omega_{H_i},\quad x_\delta=x\Omega_{H_i}C_{\text{A}}^{-1},\quad \rho_\delta=\rho\,\rho_0^{-1},\quad \rho_{\delta_s}=\rho_{0s}^{-1}\rho_s,$$

$$\left\{\begin{array}{c}E_\delta\\ H_\delta\end{array}\right\}=\left\{\begin{array}{c}E\\ H\end{array}\right\}\cdot\frac{1}{H_0},\quad v_\delta=vC_{\text{A}}^{-1},\quad v_{\delta_s}=v_sC_{\text{A}}^{-1},\quad \sigma_\delta=\sigma\Omega_{H_i}^{-1},$$

$$\sigma_{\delta_s}=\sigma_s\Omega_{H_i}^{-1},\quad \Omega_{H_i}+eH_0(mc)^{-1},\quad M=V_0C_{\text{A}}^{-1},\quad \gamma=C_{\text{A}}c^{-1},$$

$$\beta=C_sC_{\text{A}}^{-1},\quad n_s=\rho_0\rho_{0s}^{-1},\quad C_{\text{A}}^2=H_0^2(4\pi\rho_0)^{-1}$$

(индекс “δ” в дальнейшем опускаем). В линейном приближении получим два дисперсионных уравнения:

$$(\omega-kM)^2-k^2\beta^2=0 \tag{2}$$
$$\{(\omega-kV_0)^2=k^2C_s^2\ \text{— в размерных переменных}\}$$
$$(\omega-kM)^2-k^2n_s=0 \tag{3}$$
$$\{(\omega-kV_0)^2=k^2C_{\text{A}}^2\rho_0/\rho_{0s}\ \text{— в размерных переменных}\},$$

где уравнение (2) описывает звуковые волны, а уравнение (3) — альфвеновские моды.

Исследуя (2) и (3), легко обнаружить, что для волн выполняются условия синхронизма

$$\omega_1+\omega_2=\omega_3,\quad k_1+k_2=k_3\,. \tag{4}$$

Из (2) следует, что при $\omega k^{-1}<M$ одна звуковая волна имеет отрицательную энергию, а альфвеновские моды — положительную. Отметим, что общее дисперсионное уравнение для альфвеновских волн комплексно, но при условии

$$(\omega-kM)\ll 4\pi^2\gamma_s^2\sigma_s\left(n_s-\frac{(\omega-kM)^2}{k^2}\right) \tag{5}$$

оно имеет лишь действительные корни. Кроме того, учитываем каскадный процесс $\omega_1+\omega_1=2\omega_1$, $k_1+k_1=2k_1$, причем ограничиваемся лишь двумя гармониками альфвеновской моды, т. к. высшие гармоники сильно затухают благодаря циклотронному затуханию и затуханию Ландау [12]. Используя асимптотический метод [13–15], получим укороченные уравнения для медленно меняющихся комплексных амплитуд волн:

$$\frac{\partial a_{1,2}}{\partial t}+\frac{\partial a_{1,2}}{\partial x}=\sigma_{\text{A}}a_{2,1}^*b-\nu_{\text{m}}a_{1,2},\quad \frac{\partial b}{\partial t}+\beta\frac{\partial b}{\partial x}=\sigma_{\text{зв}}a_1a_2+\delta b^2,$$
$$\frac{\partial \bar{a}}{\partial t}+\beta\frac{\partial a}{\partial x}=2\sigma_{\text{A}}a_1^2-4\nu_{\text{m}}\bar{a}, \tag{6}$$

где $a_{1,2}$, $\bar{a}$ — комплексные амплитуды волн Альфвена ($a_j \sim E_j$), b — амплитуда звука, а $\sigma_{\text{А}}$, $\sigma_{\text{зв}}$, δ соответственно равны

$$\delta = \frac{\sigma_{\text{зв}}}{\omega}\,, \quad \sigma_{\text{зв}} = -ik_{1,3}B_{1,3}\,,$$
$$B_j = k_j\{(4\pi)^2\gamma^3 n_s\sigma_s[(\omega_j - k_jM) + \gamma\omega_j A_j(\omega_j k_j)] - \beta^2 k_j\}(\omega_j - k_jM)^2 - 1+$$
$$+ 4i\pi\sigma_s\{\gamma A_j[k_j\omega_j(\omega_j - k_jM)^{-1}\gamma + n_s\omega_j^{-1}] - \gamma k_j - (\omega_j - k_jM)\omega_j^{-2}n_s\},$$
$$\sigma_{\text{А}} = -k_2(4\pi\gamma^2\sigma_s\omega_2)^{-1}B_2\,, \tag{7}$$
$$A = -(\omega_j - k_jM)\cdot 4\pi\gamma n_s\sigma_s\omega_j^{-1}[4\pi\gamma^2 n_s\sigma_s + i(\omega_j - k_jM)]^{-1},$$

$\sigma_{\text{зв},\,A}$ в общем случае комплексны, их типичные значения приведены в таблице.

Таблица

β	M	i	k_j	ω_j	$\mathrm{Re}\,\sigma_j$	$\mathrm{Im}\,\sigma_j$
0,2	10	1	0,085	0,87	0,82	45,1
		2	0,0045	0,005	4,59	1,8
		3	0,0895	0,875	1,1	46,9

2. АНАЛИЗ УКОРОЧЕННЫХ УРАВНЕНИЙ

Рассмотрим стационарный случай, когда $\dfrac{\partial}{\partial t} \equiv 0$, причем волны попутные. Тогда при условии $|\sigma_{\text{А}}||a_{1,2}|^2\nu_{\text{m}}^{-1} \ll 1$ амплитуды альфвеновских волн "следят"за изменениями звука:

$$a_1^{(0)} = |\sigma_{\text{А}}|a_2 b\eta_{\text{эф}}^{-1}, \quad a_2^{(0)} = -3\delta a_1^{(0)2}\left(6\nu_{\text{m}} + \delta^2\left|a_1^{(0)}\right|^2\right)^{-1}, \tag{8}$$

где верхний индекс "(0)"означает равновесное значение первой и второй гармоник альфвеновской волны (ν_{m} — магнитная вязкость),

$$\eta_{\text{эф}} = \left(\nu_{\text{m}} + \frac{3}{2}\delta^2\left|a_1^{(0)}\right|^2\right)\left(\nu_{\text{m}} + \delta^2\left|a_1^{(0)}\right|^2\right)^{-1}. \tag{9}$$

Отметим, что взрывная неустойчивость реализуется при $M > \omega k^{-1}$ и $\eta_{\text{эф}}$ характеризует нелинейное затухание волны Альфвена, а также звука за счет перекачки энергии во вторую гармонику. Качественно $\eta_{\text{эф}}$ растет от 1 при малых $a_1^{(0)}$ и достигает насыщения $\sim 3/2$. Для определения характера изменения $b(x)$ необходимо подставить $a_1^{(0)}$ во второе уравнение (6). Можно также качественно определить фазовые соотношения между модами, используя (8), (9). Для ап-конверсии "распадается"волна низшей частоты. Инкремент параметрической неустойчивости ($|a_1 \gg |b|, |a_2|$) имеет вид $\Gamma = |a_1|$ $\left(\text{при } \dfrac{\partial}{\partial x} \equiv 0\right)$.

ЗАКЛЮЧЕНИЕ

1. Полученные результаты представляют интерес для создания усилителей (генераторов) магнитного поля в лабораторной, твердотельной плазме. Оценим амплитуды гармоник альфвеновской волны для полупроводниковой плазмы полупроводника InSb при $T = 77$ К. Параметры плазмы таковы [9]: пусть $H_0 \sim 10^3$ гс, $m_g m_e^{-1} \sim 10^2$ (m_g, m_e — массы дырок и электронов, соответственно), тогда $\omega_{H_i} \sim 10^{12}$ с$^{-1}$, а отношение частоты соударений к циклотронной частоте электронов $\sim 10^{-2}$; при энергии ~ 1 эВ скорость дрейфа пучка частиц 10^8 см·с$^{-1}$; концентрация частиц $\sim 10^8$см$^{-3}$; $\rho_{0s}\rho_0^{-1} \sim 10^{-4}$; при накачке $\sim 10^{-2}$ мгс генерируется вторая гармоника с амплитудой $\sim 5 \cdot 10^{-2}$ мгс, при этом преобразование частоты излучения вверх по спектру ~ 10, т. е. в заданном поле НЧ волны генерируется НЧ излучение с частотой на порядок выше частоты накачки — ап-конверсия.

2. Кроме того, теоретические выводы могут быть полезны для интерпретации НЧ излучения в космической плазме, например, в атмосфере Солнца, при развитии вспышечных явлений, когда в плазму инжектируются потоки заряженных частиц. Поскольку альфвеновские волны поперечны, то НЧ пульсации могут быть приняты как наземными, так и спутниковыми антеннами.

Авторы признательны П. А. Беспалову за обсуждение работы.
Работа выполнена при поддержке гранта РФФИ № 98–02–16973.

ЛИТЕРАТУРА

1. Вильхельмссон Х., Вейланд Я. Когерентное нелинейное взаимодействие волн в плазме. — М.: Энергоиздат, 1981.
2. Кадомцев Б. Б. Коллективные явления в плазме. — М.: Наука, 1988.
3. Незлин М. В. Динамика пучков в плазме. — М.: Энергоиздат, 1982.
4. Файнштейн С. М., Чернова Е. А. //ЖЭТФ, 1996. Т. 109. № 4. С. 1.
5. Рабинович М. И., Файнштейн С. М. //ЖЭТФ, 1972. Т. 63. № 5(11). С. 1672.
6. Тамойкин В. В., Файнштейн С. М. //Изв. ВУЗовю Радиофизика, 1988. Т. 31. № 9. С. 1036.
7. Файнштейн С. М. //ЖЭТФ, 1976. Т. 71. № 9. С. 1021.
8. Белянцев А. Е., Файнштейн С. М., Чернова Е. А. //Изв. ВУЗов Радиофизика, 1996. Т. 39. № 8. С. 1001.
9. Стил М., Вюраль Б. Взаимодействие волн в плазме твердого тела. — М.: Атомиздат, 1973.
10. Железняков В. В. Радиоизлучение Солнца и планет. — М.: Наука, 1964.
11. Высокочастотный нагрев плазмы: Сб. ст. /Под ред. А. Г. Литвака. — Горький: Изд-во ИПФ АН СССР, 1983.
12. Гинзбург В. Л. Электромагнитные волны в плазме. — М.: Наука, 1964.
13. Гапонов А. В., Рабинович М. И., Островский Л. А. //Изв. ВУЗов. Радиофизика, 1970. Т. 13. № 2. С. 164.
14. Цытович. Нелинейные эффекты в плазме. — М.: Наука, 1964.
15. Пустовалов В. В., Силин В. П. //Физика плазмы, 1973. Т. 61. № 42. С. 287.

Нижегородский государственный технический университет, г. Н. Новгород, Россия

Поступила в редакцию 27 марта 1997 г.

THE SECOND HARMONIC OF ALFVEN WAVE GENERATION AS A RESULT OF AN EXPLOSIVE INSTABILITY DEVELOPMENT IN A BEAM–PLASMA SYSTEM

A. E. Belyantsev, V. P. Dvoryakovsky, S. M. Fainshtein, E. A. Chernova

The possibility of an explosive and high frequency instabilities in a beam–magnetoactive plasma system is analysed. The synchronism conditions are explained. By the asymptotic method the shortened equations have been derived and solved for complex amplitudes of Alfven and sound modes. The "explosion"stabilization due to effective Alfven waves absorption in the highest harmonics is shown.

УДК 621.391:53.08

ДИФФЕРЕНЦИАЛЬНЫЙ МЕТОД МАКСИМАЛЬНОЙ ЭНТРОПИИ

А. Т. Байкова

Рассматривается дифференциальный метод максимальной энтропии, основанный на свойстве линейности преобразования Фурье и заключающийся в восстановлении изображений по разностям функции видности. Показана эффективность метода применительно к восстановлению изображений источников с яркими компонентами на фоне достаточно слабой протяженной подложки. Приводятся результаты моделирования, а также карты компактного внегалактического радиоисточника 0059+581, полученные обычным и дифференциальным методом максимальной энтропии для трех дат наблюдений, показывающие, что применение принципа дифференциального картографирования позволяет значительно повысить динамический диапазон изображений.

1. ПРИНЦИП ДИФФЕРЕНЦИАЛЬНОГО КАРТОГРАФИРОВАНИЯ

Идея дифференциального картографирования не нова. В настоящее время в наиболее полном виде она реализована в пакете программ “DifMap-[1] Калифорнийского технологического института, в котором в качестве операции деконволюции применяется процедура “чистки"(CLEAN).

Метод дифференциального картографирования основывается на фундаментальном свойстве линейности преобразования Фурье и заключается в том, что восстановленные на первом этапе яркие компоненты источника вычитаются из исходной функции видности, дальнейшее восстановление производится по остаточной функции видности, а на конечном этапе результаты восстановления суммируются.

В случае CLEAN такой подход к картографированию влияет, в основном, на скорость сходимости. В случае же использования в качестве операции деконволюции метода максимальной энтропии (ММЭ) — метода с ярко выраженными нелинейными свойствами — принцип дифференциального картографирования может привести и к улучшению качества восстановления, особенно тогда, когда источник имеет яркие компактные компоненты на фоне достаточно слабой протяженной подложки. Один из примеров такого улучшения, достигнутого благодаря вычитанию яркой компоненты из функции видности, приведен в работе [2].

В чем же заключается причина улучшения качества восстановления? Дело в том, что после вычитания ярких компонент, восстановленных ММЭ на первом этапе, из исходной функции видности, мы получаем остаточную функцию видности, в которой в большей пропорции по отношению к компактной содержится слабая протяженная компонента. (Результат вычитания назовем остаточной функцией видности первого порядка.) Таким образом мы искусственно понижаем динамический диапазон карты, соответствующей остаточной функции видности, и, тем самым, облегчаем восстановление изображения на втором этапе.

Если из остаточной функции видности первого порядка вычесть яркие компоненты изображения, восстановленного на втором этапе, то получим остаточную функцию видности второго порядка, которой соответствует карта с еще более низким динамическим диапазоном. Продолжая таким образом, будем иметь остаточные функции видности все более высокого порядка.

Для получения искомой карты источника необходимо карту, восстановленную на последнем этапе, просуммировать со всеми компонентами, вычтенными из функции видности на предыдущих этапах.

Формально алгоритм двухэтапного дифференциального картографирования можно представить следующим образом:

$$F_{\rm vis}^{(2)} = F_{\rm vis}^{(0)} - F_{\rm vis}^{(1)},$$

где

$F_{\rm vis}^{(0)}$ — исходная функция видности искомого распределения яркости по источнику $I(x)$; $F_{\rm vis}^{(1)}$ — функция видности, соответствующая яркой компоненте $I^{(1)}(x)$, восстановленной на первом этапе; $F_{\rm vis}^{(2)}$ — остаточная функция видности, соответствующая карте источника $I^{(2)}(x)$, восстанавливаемой на втором этапе.

Результирующая карта будет иметь вид

$$I(x) = I^{(1)}(x) + I^{(2)}(x).$$

2. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

На рис. 1, 2 представлены результаты моделирования принципа дифференциального картографирования с ипользованием обобщенного метода максимальной энтропии (ОММЭ) [3] (о предпочтительности использования ОММЭ вместо ММЭ будет сказано ниже в разд. 4).

Эксперимент, соответствующий рис. 1, был проведен для трех модельных источников, представляющих собой одну или несколько ярких точечных компонент на фоне более слабой протяженной подложки, имеющей форму гауссианы. На рисунках приведены имена источников: “Model−1", “Model−2"и “Model−3". У источника “Model−1"амплитуда подложки составляет 20% от амплитуды яркой точечной компоненты, а у источника “Model−2-— всего 2%. Более сложным является источник “Model−3", имеющий четыре яркие компоненты но фоне 20% гауссианы. Для формирования функции видности было использовано симметричное покрытие UV−плоскости с почти равномерным заполнением по всей апертуре, включающим 98 точек. С целью изучения эффекта улучшения восстановления благодаря применению принципа дифференциального картографирования в чистом виде, зашумливания функции видности не производилось.

На рис. 1, 2 для каждого из источников, размещенных по вертикали, приводятся карты, следующие в определенном порядке по горизонтали (слева направо): карта модельного источника; “грязное"изображение, полученное в результате обратного преобразования Фурье исходной функции видности; изображение, восстановленное на первом этапе дифференциального картографирования с использованием ОММЭ; результирующая карта дифференциального картографирования. (Отметим, что везде на приведенных изображениях нижнему уровню контурной линии соответствует 1% от пикового значения потока.)

Очевидно, карты, полученные на первом этапе, можно рассматривать как предел возможностей для ОММЭ, примененного к исходной функции видности. Вычитание же из функции видности ярких компонент, восстановленных на первом этапе, позволило заметно улучшить качество восстановления на втором этапе. В табл. 1 приведены количественные характеристики качества восстановления изображений рис. 1 после первого и второго этапов дифференциального картографирования.

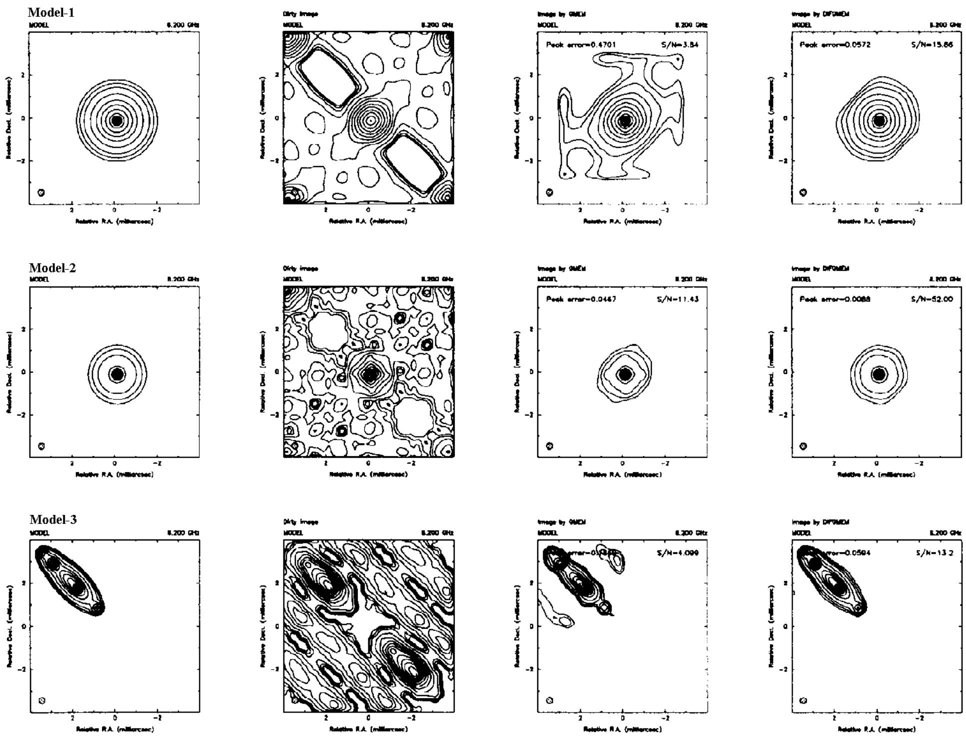


Рис. 1.

Таблица 1

Характеристики восстановленных изображений

Источник	Макс. ошибка на I этапе	Отношение “С/ш" на I этапе	Макс. ошибка на II этапе	Отношение “С/ш" на II этапе
Model−1	0.4701	3.84	0.0572	15.86
Model−2	0.0467	11.43	0.0088	52.00
Model−3	0.1890	4.10	0.0594	13.20

Как видно из таблицы, применение дифференциального ОММЭ по сравнению с простым ОММЭ позволило повысить отношение “сигнал/шум"выходного изображения в 3−5 раз, причем наибольший эффект достигнут для источника “Model−2", имеющего наибольший динамический диапазон.

На рис. 2 представлены результаты восстановления изображений источников “Model−1"и “Model−2"при другом заполнении UV−плоскости, включающем 98 точек, распределенных равномерно внутри центральной

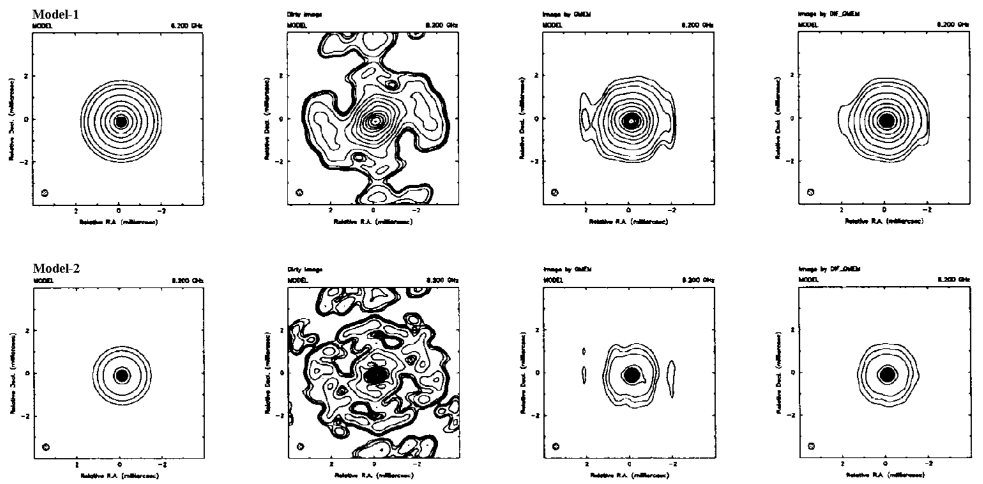


Рис. 2.

области, составляющей примерно четвертую часть от апертуры. Этот случай картографирования оказался более тяжелым по сравнению с предыдущим из-за полного отсутствия в функции видности высоких пространственных гармоник источника. Как видно из рисунка, дифференциальный ОММЭ позволил получить более качественные результаты по сравнению с простым ОММЭ и в этом случае.

3. КАРТОГРАФИРОВАНИЕ РАДИОИСТОЧНИКА 0059+581

С целью демонстрации возможностей дифференциального обобщенного метода максимальной энтропии для обработки реальных данных приведем результаты построения карт внегалактического радиоисточника 0059+581, активно используемого в качестве реперного источника в РСДБ геодезических программах [4]. Этот источник интересен тем, что обнаруживает быструю переменность как полного потока, так и структуры. Карты, построенные с использованием ОММЭ для ряда дат на интервале времени с июня 1994 г. по декабрь 1995 г., приведены на рис. 3. Изменение полного потока на этом же интервале показано на рис. 4 работы [5].

На интервале времени от середины до конца 1994 г. наблюдался почти линейный спад полного потока с $\sim$ 4 Ян до 1,5–2 Ян. Как видно из рис. 3, по данным за 26 июня 1994 г., когда наблюдался максимальный в рассматриваемом временном интервале поток, не удалось получить карту с динамическим диапазоном, достаточным для обнаружения протяженной детали. Благодаря применению простого ОММЭ удалось обнаружить протяженные детали, начиная лишь с 4 октября 1994 г., когда они стали составлять относительно полного потока долю, достаточную для обнаружения избранным методом восстановления. На рис. 4 для сравнения приводятся карты, полученные как простым, так и дифференциальным ОММЭ, для трех дат, когда полный поток был достаточно велик и доля потока, приходящегося на протяженные детали, была незначительной. Как видно из рисунка, применение дифференциального картографирования позволило настолько повысить динамический диапазон карт, что протяженные структуры явно выделились. Отметим, что везде на приведенных изображениях минимальному уровню контурной линии соответствует 1 % от пикового значения потока. Параметры карт, полученных с использованием дифференциального ОММЭ, приводятся ниже в табл. 2.

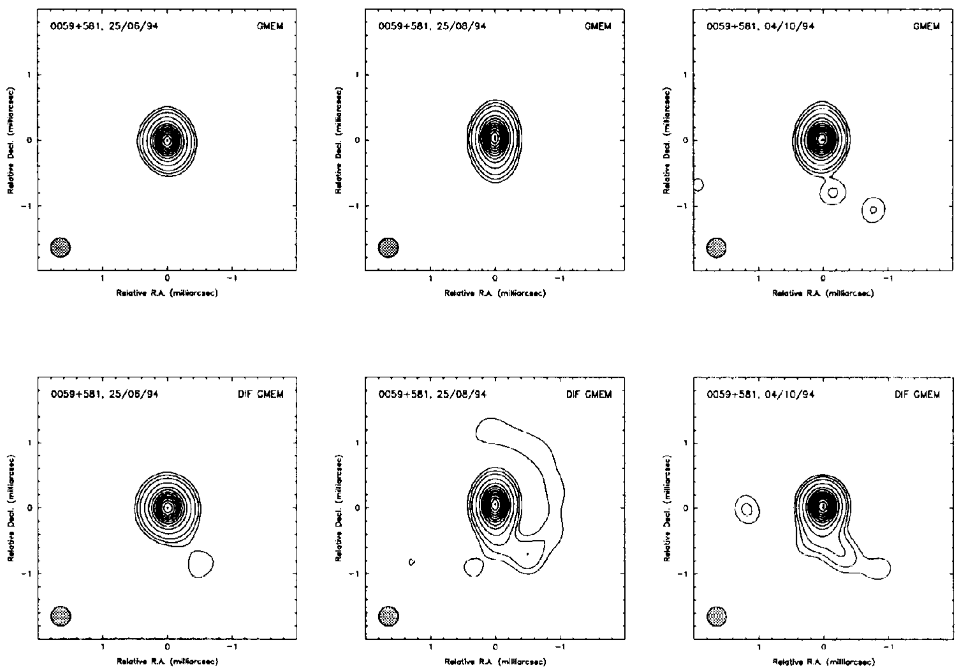


Рис. 4.

Таблица 2

Параметры карт источника 0059+581

Дата	Поток полн.(Ян)	Поток пик.(Ян)	Фактор согласия		
			ампл.	фаз.	полн.
25.06.94	4.23	2.40	6.138	1.919	5.098
25.08.94	3.40	1.76	4.016	2.828	3.616
04.10.94	3.28	1.86	7.892	4.559	7.405

4. ОБ ИСПОЛЬЗОВАНИИ ОММЭ ВМЕСТО ММЭ

Главным критерием для выбора метода максимальной энтропии (классического или обобщенного (ММЭ или ОММЭ)) является вид искомого распределения. В случае, когда искомое изображение является комплексным, имеющим как вещественную, так и мнимую часть, может быть использован только ОММЭ [6], полученный модифицированием классического метода специально для восстановления комплексных функций.

В случае, когда априори известно, что изображение является вещественным и неотрицательным, можно использовать классический ММЭ. Но при этом надо иметь в виду, что в случае больших ошибок в данных ОММЭ по сравнению с ММЭ, как показано в [7], позволяет получить более качественные карты с существенно меньшим уровнем нелинейных искажений. В настоящее время в ИПА РАН ведутся работы по картографированию компактных внегалактических источников по наблюдениям астрометрических и геодезических РСДБ программ. Эти данные не предназначены напрямую для картографирования с астрофизическими целями и, как правило, плохо калиброваны. Поэтому с целью получения удовлетворительных карт рекомендуется вместо ММЭ использовать ОММЭ.

В данной работе еще один довод в пользу применения ОММЭ вытекает из тех соображений, что остаточной функции видности, полученной после вычитания из исходной функции видности яркой компоненты, восстановленной на первом этапе алгоритма, может соответствовать изображение с отрицательными значениями. Это может произойти в том случае, если на первом этапе яркая компонента восстановилась с завышенным значением амплитуды, что вполне вероятно для любых нелинейных методов. Во избежание нежелательных искажений изображения в ситуации, когда ищется неотрицательное решение по данным, предполагающим наличие отрицательных компонент, следует использовать обобщенный метод максимальной энтропии вместо классического.

5. ВЫВОДЫ

Нелинейные дифференциальные методы картографирования, использующие в качестве операции деконволюции метод максимальной энтропии, позволяют, по сравнению с традиционными методами, заметно повысить качество восстановления изображений источников, содержащих яркие точечные компоненты на фоне слабой подложки. Карты, полученные с использованием дифференциального метода максимальной энтропии, характеризуются более высоким динамическим диапазоном. При дифференциальном картографировании, во избежание возможных нелинейных искажений, более предпочтительным является использование обобщенного метода максимальной энтропии вместо классического.

Работа выполнена при поддержке Российского фонда фундаментальных исследований по гранту № 96–02–19177 и Миннауки по подпрограмме "Астрономия. Фундаментальные космические исследования"по гранту № 2.1.1.3.

ЛИТЕРАТУРА

1. Taylor G. The Difmap Cookbook. — Pasadena: California Institute of Technology, 1994.
2. Cornwell T. Very Long Baseline Interferometry and the VLBA. //ASP Conference Series / Eds. J. A. Zensus, P. J. Diamond, P. J. Napier, 1995. V. 82. P. 227.
3. Байкова А. Т. //Сообщения ИПА РАН, 1993. № 58.
4. Байкова А. Т., Пятунина Т. Б., Финкельштейн А. М. //Труды ИПА РАН, 1997. Вып. 1. С. 22.
5. Пятунина Т. Б., Байкова А. Т., Финкельштейн А. М. //Труды ИПА РАН, 1997. Вып. 1. С. 64.
6. Байкова А. Т. //Изв. ВУЗов. Радиофизика, 1991. Т. 34. № 8. С. 919.
7. Байкова А. Т. //Изв. ВУЗов. Радиофизика, 1995. Т. 38. № 12. С. 1267.

Институт прикладной астрономии
РАН, г. Санкт-Петербург, Россия

Поступила в редакцию
13 февраля 1998 г.

DIFFERENCE MAXIMUM ENTROPY METHOD

A. T. Bajkova

The difference maximum entropy method based on the Fourier transform linearity and intended for image restoretion from visibility function differences is considered. The method effectiveness is shown as applied to image reconstruction of sources with bright compact components embeded in weak, rather extended details. Simulation results as well as maps of the source 0059+581 obtained for three dates by both usual and the difference maximum entropy method are given. It is shown that the difference maximum entropy method allows to increase considerably the dynamic range of the images.

УДК 535.41

СТАТИСТИКА ФОТООТСЧЁТОВ ПРИ РЕГИСТРАЦИИ ДЕТЕКТОРОМ С ПЕРЕМЕННОЙ КВАНТОВОЙ ЭФФЕКТИВНОСТЬЮ

А. С. Мазманишвили

Рассмотрена задача о распределении отсчетов оптического фотодетектора, регистрирующего неполяризованное поле излучения, обладающее свойствами нормального марковского процесса. Известная формула Манделя, описывающая вероятность $P(n)$ регистрирования n отсчетов за интервал наблюдения $(0,T)$, обобщена на случай детектора с зависящей от времени квантовой эффективностью $\epsilon(t)$.

1. ПОСТАНОВКА ЗАДАЧИ

Флуктуации числа фотоэлектрических отсчетов содержат информацию о состоянии поля излучения. При описании результатов опытов по распределению фотоотсчетов оптическими детекторами пользуются ставшей уже классической формулой Манделя. Эта формула для распределения $P(n)$ числа n фотоотсчетов (фотоэлектронов) на интервале регистрации $(0,T)$ имеет вид [1, 2]

$$P(n) = \left\langle \frac{1}{n!}\Omega^n \exp(-\Omega) \right\rangle . \tag{1}$$

В этой формуле

$$\Omega = \int\limits_0^T \epsilon(t)|\alpha(t)|^2 dt, \tag{2}$$

$\alpha(t)$ — комплексная амплитуда излучения, $\epsilon(t)$ — квантовая чувствительность (эффективность) фотодетектора. В том случае, когда регистрируется излучение со случайными свойствами, в формуле (1) необходимо выполнить усреднение по ансамблю всех возможных значений поглощенной детектором энергии Ω оптического поля излучения, реализующееся на рассматриваемом временном интервале $(0,T)$, или по всем возможным на интервале $(0,T)$ реализациям случайного амплитуды $\alpha(t)$ поля излучения. В (1) операция усреднения обозначена уголковыми скобками $\langle\cdot\rangle$. Другими словами, если $p(\Omega)$ — плотность распределения случайной величины Ω, то

$$P(n) = \int\limits_0^\infty \frac{1}{n!}\Omega^n \exp(-\Omega) p(\Omega) d\Omega. \tag{3}$$

Таким образом, в общем случае распределение отсчетов описывается взвешенным распределением Пуассона, т. е. операция усреднения по ансамблю приводит в общем случае к отклонению $P(n)$ от распределения Пуассона.

Большинство практически распространенных выражений относятся к случаю, когда $\epsilon(t) \equiv \mathrm{const}$. Между тем, если $\epsilon(t) \not\equiv \mathrm{const}$, то конкретный вид $p(\Omega)$, а с ней и $P(n)$ изменится. Конструкция формулы Манделя позволяет рассмотреть общий случай, когда $\epsilon(t) \not\equiv \mathrm{const}$.

Целью настоящей работы является получение общей формулы для распределения фотоотсчетов детектором с квантовой эффективностью $\epsilon(t)$, зависящей от текущего времени произвольным физически реализуемым образом. При этом будет принято, что поле излучения обладает свойствами нормального марковского комплекснозначного стохастического процесса с интенсивностью σ и шириной спектральной линии ν [3]. Такой процесс (процесс Орнштейна—Уленбека) описывается уравнением

$$\frac{d}{dt}\alpha(t) = \nu\alpha(t) + u(t), \quad \alpha(0) = \alpha_0, \tag{4}$$

с порождающим процессом $u(t)$ типа белого шума, при этом он имеет равновесную плотность распределения вида

$$w(\alpha_0) = \frac{1}{\pi\sigma}\exp\left(-\frac{|\alpha_0|^2}{\sigma}\right) \tag{5}$$

и в каждом сечении его первые моменты следующие:

$$\langle\alpha(t)\rangle = 0, \quad \langle|\alpha(t)|^2\rangle = \sigma. \tag{6}$$

2. ПРОИЗВОДЯЩАЯ ФУНКЦИЯ ФОТООТСЧЁТОВ

В задачах статистической оптики свойства распределения фотоотсчетов $P(n)$ удобно описывать посредством производящей функции [1, 2]. Для распределения (3) она имеет вид

$$Q_\epsilon(\lambda, T) = \langle\exp(-\lambda\Omega)\rangle = \int\limits_0^\infty \exp(-\Omega)p(\Omega)d\Omega, \tag{7}$$

и, если известна функция $Q_\epsilon(\lambda, T)$, то

$$P(n) = \frac{(-1)^n}{n!}\cdot\frac{d^n}{d\lambda^n}Q_\epsilon(\lambda, T)\bigg|_{\lambda=1}. \tag{8}$$

Таким образом, формально описание статистики отсчетов сводится к вычислению следующего континуального интеграла:

$$Q_\epsilon(\lambda, T) = \int w(\alpha_0)d^2\alpha_0\int d^2\alpha_T\Psi_\lambda(\alpha_0, 0; \alpha_T, T), \tag{9}$$

где

$$\Psi_\lambda(\alpha_0, 0; \alpha_T, T) = \langle\alpha_0, 0|\exp\left(-\lambda\int\limits_0^T \epsilon(t)|\alpha(t)|^2dt\right) \quad — \tag{9'}$$

условная амплитуда перехода из состояния $\langle\alpha_0, 0|$ в состояние $|\alpha_T, T\rangle$, усредненная по всем возможным траекториям процесса $\alpha(t)$, реализующимся за интервал регистрации $0 \leq t \leq T$, и по всем возможным начальным α_0 и конечным состояниям α_T.

3. ВЫЧИСЛЕНИЕ ПРОИЗВОДЯЩЕЙ ФУНКЦИИ ФОТООТСЧЁТОВ

Ввиду гауссового свойства процесса $\alpha(t)$ воспользуемся для нахождения среднего (9) методом Карунена–Лоэва [1]. Запишем безусловное среднее (9) с помощью континуального преобразования Фурье

$$Q_\epsilon(\lambda,T)=\int D^2f(t)\exp\left(-\int\limits_0^T|f(t)|^2dt\right)\left\langle\exp\left(\sqrt{-\lambda}\cdot 2\mathrm{Re}(A)\right)\right\rangle, \tag{10}$$

где

$$A=\int\limits_0^T\sqrt{\epsilon(t)}\alpha^*(t)f(t)dt.$$

В силу гауссовости $\alpha(t)$ величина A также гауссова с $\langle A\rangle=0$ и

$$\langle|A|^2\rangle=\sigma\int\limits_0^T\int\limits_0^T\sqrt{\epsilon(t)\epsilon(t^{'})}f^*(t)f(t^{'})dtdt^{'}. \tag{11}$$

Поэтому

$$Q_\epsilon(\lambda,T)=\int D^2f(t)\exp\left(-\int\limits_0^T|f(t)|^2dt-\lambda\sigma\left|\int\limits_0^T\sqrt{\epsilon(t)}f(t)dt\right|^2\right). \tag{12}$$

Алгоритм метода Карунена–Лоэва включает вычисление характеристических чисел интегрального оператора с соответствующим корреляционным ядром $K(t,t^{'})$, отвечающим нормальному марковскому процессу

$$f(t)=\lambda\int\limits_0^T K(t,t^{'})f(t^{'})dt^{'}; \tag{13}$$

при этом из (12) вытекает, что ядро имеет вид

$$K(t,t^{'})=\epsilon(t)\sigma\exp(-\nu|t-t^{'}|). \tag{14}$$

В рамках этого метода необходимо найти дисперсионное уравнение, определяющее набор полюсов функции $Q_\epsilon(\lambda,T)$, связанных с набором характеристических чисел интегрального оператора.

Детерминант Фредгольма $D(\lambda)$ ядра $K(t,t^{'})$ выражается через характеристические числа $\{\lambda_n\}_1^\infty$ в виде разложения Адамара

$$D(\lambda)=\prod_{n=1}^\infty(1-\lambda/\lambda_n), \tag{15}$$

что дает для производящей функции

$$Q_\epsilon(\lambda,T)=D(0)/D(-\lambda). \tag{16}$$

Из уравнения (13) вытекает его аналог, записанный в дифференциальной форме,

$$\frac{d^2}{dt^2}f(t)=r^2(t)f(t),\quad r^2(t)=\nu^2-2\lambda\sigma\nu\epsilon(t), \tag{17}$$

а также краевые условия, которым подчиняется функция $f(t)$,

$$\left(\frac{d}{dt}f(t)-\nu f(t)\right)\bigg|_{t=0}=0, \quad \left(\frac{d}{dt}f(t)+\nu f(t)\right)\bigg|_{t=T}=0. \tag{18}$$

Для получения формального решения уравнения (17) разобьем интервал $(0,T)$ на N частей длиной $\Delta_n=T/N$ каждая и заменим функцию $r(t)$ на ее кусочно—постоянный аналог такой, что на каждом n—ом участке $r(t)=r_n=r(n\Delta)$.Теперь рассмотрим вместо (17) последовательность систем уравнений

$$\frac{d}{dt}\begin{pmatrix} f_n(t) \\ g_n(t) \end{pmatrix}=\begin{pmatrix} 0 & r_n \\ r_n & 0 \end{pmatrix}\begin{pmatrix} f_n(t) \\ g_n(t) \end{pmatrix}, \quad n=1,\,2,\,\ldots,\,N, \tag{19}$$

с начальными условиями $f_1(0)=f_0$ и $g_1(0)=g_0$.

Формальное решение уравнения (17) на n—ом участке следующее:

$$\begin{pmatrix} f_n(t) \\ g_n(t) \end{pmatrix}=\begin{pmatrix} \mathrm{ch}(S_{n-1}+r_nt) & \mathrm{sh}(S_{n-1}+r_nt) \\ \mathrm{sh}(S_{n-1}+r_nt) & \mathrm{ch}(S_{n-1}+r_nt) \end{pmatrix}\begin{pmatrix} f_0 \\ g_0 \end{pmatrix}, \tag{20}$$

где $S_{n-1}=\sum\limits_{k=1}^{n-1}\Delta_k r_k$.

Переходя теперь к пределу $N\to\infty$, запишем формальное решение в виде

$$\begin{aligned} f(t)=C_1\exp\left(\int_0^t\sqrt{\nu^2-2\lambda\sigma\nu\epsilon(s)}ds\right)+ \\ +C_2\exp\left(-\int_0^t\sqrt{\nu^2-2\lambda\sigma\nu\epsilon(s)}ds\right) \end{aligned} \tag{21}$$

с некоторыми константами C_1 и C_2.

Искомое дисперсионное уравнение можно получить, если потребовать для формального решения (21) выполнения условий (18). Рассматривая (18) как однородную систему из двух линейных алгебраических уравнений относительно C_1 и C_2, потребуем существования у нее нетривиального решения. Это и дает дисперсионное уравнение вида

$$D(\lambda)\equiv(g_0+\nu)(g_{_T}+\nu)\exp(G)-(g_0-\nu)(g_{_T}-\nu)\exp(-G)=0, \tag{22}$$

где

$$\begin{gathered} g_0=\sqrt{\nu^2-2\lambda\sigma\nu\epsilon(0)}, \quad g_{_T}=\sqrt{\nu^2-2\lambda\sigma\nu\epsilon(T)}, \\ G=\int\limits_0^T\sqrt{\nu^2-2\lambda\sigma\nu\epsilon(t)}dt. \end{gathered} \tag{22$'$}$$

Учитывая условие нормировки $Q_\epsilon(0)=1$, из найденного определителя Фредгольма (22) получаем следующее выражение для искомой производящей функции фотоотсчетов:

$$Q_\epsilon(\lambda,T)=\frac{2\nu(r_0+r_{_T})\exp(\nu T)}{(r_0+\nu)(r_{_T}+\nu)\exp(R)-(r_0-\nu)(r_{_T}-\nu)\exp(-R)}\,, \tag{23}$$

где

$$\begin{gathered} r_0=\sqrt{\nu^2+2\lambda\sigma\nu\epsilon(0)}, \quad r_{_T}=\sqrt{\nu^2+2\lambda\sigma\nu\epsilon(T)}, \\ R=\int\limits_0^T\sqrt{\nu^2+2\lambda\sigma\nu\epsilon(t)}dt. \end{gathered} \tag{23$'$}$$

Формула (23) является основным результатом настоящей работы. В частном случае постоянной во времени квантовой эффективности фотодетектора, когда $\epsilon(t) \equiv 1$, найденное выражение для производящей функции фотоотсчетов переходит в известное (см., например, [2, 4])

$$Q_\epsilon(\lambda, T)\Big|_{\epsilon=1} = \frac{2\nu r \exp(\nu T)}{(r+\nu)^2 \exp(rT) - (r-\nu)^2 \exp(-rT)}\,,$$
$$r = \sqrt{\nu^2 + 2\lambda\sigma\nu}. \tag{24}$$

Плотность распределения $p(\Omega)$ значений случайной поглощенной фотодетектором энергии Ω может быть, согласно (7), определена из (23) с помощью обратного преобразования Лапласа.

4. СВОЙСТВА РАСПРЕДЕЛЕНИЯ ФОТООТСЧЁТОВ ПРИ РЕГИСТРАЦИИ ДЕТЕКТОРОМ С ПЕРЕМЕННОЙ КВАНТОВОЙ ЭФФЕКТИВНОСТЬЮ

Результат (23) имеет место при любых физически реализуемых зависимостях для квантовой эффективности $\epsilon(t)$ фотодетектора. Из (23) можно получить, что среднее число фотоотсчетов равно

$$\langle n\rangle = -\frac{d}{d\lambda} Q_\epsilon(\lambda, T)\Big|_{\lambda=0} = \sigma \int\limits_0^T \epsilon(t)dt, \tag{25}$$

а второй факториальный момент равен

$$\langle n(n-1)\rangle = \frac{d^2}{d\lambda^2} Q_\epsilon(\lambda, T)\Big|_{\lambda=0} =$$
$$= \langle n\rangle^2 + (\sigma^2/\nu) \int\limits_0^T \epsilon^2(t)dt - \frac{1}{2}(\sigma/\nu)^2 \epsilon_0 \epsilon_T (1 - \mathrm{e}^{-2\nu T}). \tag{26}$$

Отсюда следует выражение для дисперсии числа фотоотсчетов

$$\langle n^2\rangle - \langle n\rangle^2 = \sigma \int\limits_0^T \epsilon(t)dt + (\sigma^2/\nu) \left(\int\limits_0^T \epsilon^2(t)dt - \epsilon_0 \epsilon_T \frac{1 - \mathrm{e}^{-2\nu t}}{2\nu} \right). \tag{27}$$

В формулах для среднего числа фотоотсчетов (25) и дисперсии (27) зависимость от квантовой эффективности $\epsilon(t)$ как от функции текущего времени указана явно. Отметим, что в выражениях для моментов помимо функции $\epsilon(t)$ присутствуют ее значения ϵ_0 и ϵ_T, отвечающие границам интервала регистрации.

Величина $\delta = (\langle n^2\rangle - \langle n\rangle^2 - \langle n\rangle)\langle n\rangle^{-2}$ (параметр группировки фотоотсчетов [2, 4]) в случае идеального распределения Пуассона тождественно равна нулю, поэтому ее удобно рассматривать как индикатор близости к идеальному случаю. Рассмотрим простой пример для квантовой эффективности $\epsilon(t)$ вида

$$\epsilon(t) = 1 - \frac{t}{t + T_\mathrm{d}} \tag{28}$$

с некоторой постоянной $T_{\rm d}$ фотокатода детектора. Такая зависимость не отклоняется от идеальной тем лучше, чем больше параметр $T_{\rm d}$. После интегрирования в (27) получим

$$\delta = \frac{1}{1+T/T_{\rm d}}\delta_\infty, \tag{29}$$

где δ_∞ — значение параметра группировки в случае $T_{\rm d} \to \infty$, когда временные характеристики детектора не проявляются. Ситуация же, когда $(1+T/T_{\rm d})^{-1} \ll 1$, отвечает по существу детектированию поля излучения за малый интервал $T_{\rm d}$, при этом реализуются лишь состояния $\alpha(t)$, близкие к начальному α_0, и, таким образом, эффективно регистрируется поле в квазикогерентном состоянии. Поэтому на выходе фотодетектора с малой длительностью $T_{\rm d}$ квантовой эффективности будет зарегистрирован поток отсчетов, в котором явление их группировки окажется заниженным тем сильней, чем меньше $T_{\rm d}$.

В случае идеального фотодетектора на формирование статистики отсчетов влиял, согласно (24), параметр $\eta_1 = \nu T = T/T_c$, где $T_c = \nu^{-1}$ — время когерентности излучения. Детектор с неидеальным фотокатодом изменяет статистическую картину тем сильней, чем меньше величина $T_{\rm d}$. Включение в рассмотрение дополнительного параметра $\eta_2 = T/T_{\rm d}$ нарушает автомодельность распределения $P(n)$.

5. ВРЕМЕННЫЕ ИНТЕРВАЛЫ МЕЖДУ ФОТООТСЧЁТАМИ ПРИ РЕГИСТРАЦИИ ДЕТЕКТОРОМ С ПЕРЕМЕННОЙ КВАНТОВОЙ ЭФФЕКТИВНОСТЬЮ

Плотность распределения $g(\tau)$ временных интервалов $\{\tau\}$ между последовательными фотоотсчетами при пуассоновском законе является экспоненциальной. В случае некогерентного поля излучения с комплексной амплитудой $\alpha(t)$ для плотности $g(\tau)$ имеем [5]

$$g(\tau) = \int D^2\alpha(\tau)|\alpha(\tau)|^2 \exp\left(-\int\limits_0^\tau |\alpha(t)|^2 dt\right). \tag{30}$$

Пользуясь определением производящей функции, запишем

$$g(\tau) = -\frac{d}{d\tau}Q_\epsilon(\lambda,\tau)\bigg|_{\lambda=1}, \tag{31}$$

где $Q_\epsilon(\lambda,\tau)$ определяется выражением (7). С помощью (23), найдем

$$g(\tau) = \epsilon(\tau)\frac{4\nu^2\sigma(\rho_0+\rho_\tau){\rm e}^{\nu\tau}[(\rho_0+\nu){\rm e}^{B}+(\rho_0-\nu){\rm e}^{-B}]}{[(\rho_0+\nu)(\rho_\tau+\nu){\rm e}^{B}-(\rho_0-\nu)(\rho_\tau-\nu){\rm e}^{-B}]^2}, \tag{32}$$

$$\rho_0 = \sqrt{\nu^2+2\sigma\nu\epsilon(0)}, \quad \rho_\tau = \sqrt{\nu^2+2\sigma\nu\epsilon(\tau)},$$
$$B = \int_0^\tau \sqrt{\nu^2+2\sigma\nu\epsilon(t)}dt. \tag{32'}$$

В формуле (32) видно доминирующее влияние квантовой эффективности $\epsilon(\tau)$ на формирование статистики временных интервалов $\{\tau\}$ между фотоотсчетами. В предельном случае малого значения параметра $T_{\rm d}$ и большом времени когерентности T_c поля излучения распределение временных интервалов (32) переходит в

$$g(\tau) = \sigma\epsilon(\tau)\exp\left(-\sigma\int\limits_0^\tau \epsilon(t)dt\right), \tag{33}$$

т. е. фактически описывает лишь характеристики фотокатода детектора.

6. ВЫВОДЫ

Таким образом, в работе рассмотрена задача о распределении отсчетов оптического фотодетектора, регистрирующего неполяризованное поле излучения, обладающее свойствами нормального марковского процесса. Известная формула Манделя, описывающая вероятность $P(n)$ регистрирования n отсчетов за интервал наблюдения $(0, T)$, обобщена на случай детектора с квантовой эффективностью $\epsilon(t)$, зависящей от времени. Найденное выражение для производящей функции фотоотсчетов $Q_\epsilon(\lambda, T)$ позволило аналитически проанализировать влияние временных характеристик детектора на свойства распределения $P(n)$. Изучен вклад двух временных масштабов T_c (регистрируемого поля) и T_{d} (детектора) в образование распределения $P(n)$ фотоотсчетов случайного поля излучения $\alpha(t)$ за интервал регистрации T.

Развитый в работе подход для нахождения производящей функции (7) фотоотсчетов возможно реализовать при решении других задач квантовой оптики и статистической радиофизики, в которых необходимо учитывать временные свойства приемных регистрирующих устройств.

В заключение благодарю акад. А. И. Ахиезера за поддержку работы.

ЛИТЕРАТУРА

1. Вольф Э., Мандель Л. //УФН, 1966. Т. 88. № 4. С. 619.
2. Peřina J. Quantum Statistics of Linear and Nonlinear Optical Phenomena. — Dordrecht: Kluwer, 1991. – 412 p.
3. Тихонов В. И., Миронов М. А. Марковские процессы. — М.: Сов. радио, 1977. – 488 с.
4. Хелстром Л. Квантовая теория проверки гипотез и оценивания. — М.: Мир, 1979. – 342 с.
5. Мазманишвили А. С. //Изв. ВУЗов. Радиофизика, 1986. Т. 29. № 12. С. 1504.

Харьковский государственный
политехнический университет,
Украина

Поступила в редакцию
16 февраля 1998 г.

THE PHOTOCOUNTS STATISTICS IN THE CASE OF THE DETECTOR WITH TIME–DEPENDENT QUANTUM EFFICIENCY

A. S. Mazmanishvili

The problem of optical detector photocounts distribution evaluation is considered for the normal Markovian field. The well–known Mandel formula, discribing the probability $P(n)$ of registering n photocounts for an interval of observation $(0, T)$, is extended over the case of the detector with time–dependent quantum effeciency $\epsilon(t)$.

УДК 621.372.832

МОДИФИЦИРОВАННЫЙ МЕТОД ПОЧТИ ПОЛНОГО ОБРАЩЕНИЯ СИНГУЛЯРНОГО ИНТЕГРАЛЬНОГО ОПЕРАТОРА В ТЕОРИИ ЭКРАНИРОВАННЫХ ПОЛОСКОВЫХ ЛИНИЙ ПЕРЕДАЧИ

А. С. Арефьев, В. А. Неганов

Проводится модификация метода почти полного обращения оператора, позволяющая рассматривать экранированные полосковые структуры на основе аппарата интегральных уравнений второго рода. Исследуются высшие собственные волны экранированной несимметричной полосковой линии передачи с учетом несимметрии расположения полоски относительно центра экрана.

1. ВВЕДЕНИЕ. СУТЬ ПРОБЛЕМЫ

Полосковые волноведущие структуры составляют основу современных планарных и объемных интегральных схем (ИС) СВЧ [1]. Проблемы построения систем математического моделирования и автоматизированного проектирования ИС СВЧ в значительной степени определяются наличием эффективных алгоритмов расчета полосково—щелевых структур. В области анализа ИС СВЧ в настоящее время доминируют прямые проекционные методы [2], к достоинствам которых следует отнести их универсальность и относительную простоту численной реализации. Однако практическое осуществление таких методов наталкивается на ощутимые трудности, связанные с медленной сходимостью, а в ряде случаев — неустойчивостью соответствующих алгоритмов. Последнее обстоятельство обусловлено тем, что проекционные схемы применяются к интегральным уравнениям первого рода, которые, как известно, относятся к классу некорректно поставленных задач [3]. Таким образом, краевые задачи электродинамики, допускающие формулировку в терминах интегральных уравнений первого рода, нуждаются в применении новых, численно—аналитических методов решения, опирающихся на учет специфики исследуемой структуры.

Одним из эффективных методов расчета экранированных планарных систем является метод почти полного обращения интегрального оператора (МПО), идея которого развита в работах [4—8]. Метод опробован на целом ряде щелевых структур, таких как волноводно—щелевая линия, компланарный волновод, экранированная несимметричная двухщелевая линия передачи. Однако при расчете полосковых резонаторов и линий передачи с использованием МПО возникает ряд трудностей. Основная проблема здесь заключается в выборе оператора обращения. Дело в том, что ядро одного из интегральных уравнений первого рода, возникающих при моделировании полосковых структур, имеет особенность вида [7, 8]

$$\sum_{m=1}^{\infty} \frac{1}{m} \sin\frac{\pi m x}{a} \sin\frac{\pi m x'}{a} = \frac{1}{2} \ln\left| \frac{\sin\pi\dfrac{x+x'}{2a}}{\sin\pi\dfrac{x-x'}{2a}} \right|. \tag{1}$$

Для интегрального уравнения с особенностью (1) в настоящее время не известен обратный оператор. Поэтому в [7, 8] данная особенность путем дифференцирования интегрального уравнения по переменной x преобразуется к виду

$$2\sum_{m=1}^{\infty} \cos\frac{\pi m x}{a} \sin\frac{\pi m x'}{a} = \frac{\sin(\pi x'/a)}{\cos(\pi x/a) - \cos(\pi x'/a)}\,. \tag{2}$$

Затем, в результате применения соотношений Швингера [9], уравнение сводится к интегралу типа Коши, для которого известны формулы обращения. Далее осуществляется регуляризация уравнения с особенностью (2) и достаточно просто записывается общий вид решения для тока на полоске. Однако при дифференцировании (1) исчезают неявно присутствующие константы. Как результат, после обращения оператора, при записи общего вида решения возникает неопределенная постоянная, которую невозможно определить из интегрального уравнения с особенностью (2). Для достижения однозначности решения преобразованной системы интегральных уравнений приходится задействовать дополнительное условие. С этой целью обычно используется подстановка решения в исходное интегральное уравнение с особенностью (1) при некотором значении переменной x. Но можно ли считать последовательным подобный шаг? Стоило ли возводить МПО на отрицании прямой проекционной схемы, чтобы в конечном счете обратиться за помощью к процедуре, куда более слабой с точки зрения ее математической обоснованности? Заметим, что с аналогичными трудностями приходится сталкиваться и при расчете индуктивной диафрагмы в прямоугольном волноводе [9].

Таким образом, прямое применение МПО к полосковым структурам (линии передачи, резонаторы) приводит к существенным трудностям и, как следствие, обычно приходится ограничиваться первым или вторым приближениями [7, 8]. При этом остается без внимания вопрос о внутренней сходимости метода. Особенно это актуально при исследовании высших типов собственных волн.

Ниже для устранения указанных трудностей предлагается использовать соотношения

$$\begin{gathered}\sum_{m=1}^{\infty}\frac{1}{m}\sin\frac{\pi m x}{a}\sin\frac{\pi m x'}{a}=\sum_{m=1}^{\infty}\frac{1}{m}\cos\frac{\pi m x}{a}\cos\frac{\pi m x'}{a}+\ln\left|2\sin\pi\frac{x+x'}{2a}\right|,\\ \sum_{m=1}^{\infty}\frac{1}{m}\cos\frac{\pi m x}{a}\cos\frac{\pi m x'}{a}=-\frac{1}{2}\ln\left|2\left(\cos\frac{\pi x'}{a}-\cos\frac{\pi x}{a}\right)\right|,\end{gathered}\tag{3}$$

которые позволяют привести задачу о решении интегрального уравнения с особенностью (1) к решению интегральных соотношений с особенностью типа $\ln\left|\cos\frac{\pi x'}{a}-\cos\frac{\pi x}{a}\right|$, для которых известен обратный оператор, и потому отпадает необходимость в дифференцировании.

2. ФОРМУЛИРОВКА ЗАДАЧИ. ВЫБОР ОПЕРАТОРА ОБРАЩЕНИЯ

Проиллюстрируем модифицированный метод на примере обобщенной экранированной несимметричной полосковой линии передачи (ЭНПЛ): токопроводящая полоска в плоскости $y = 0$ между двумя областями $y < 0$ и $y > 0$, каждая из которых может содержать как изотропные, так и анизотропные слои (на рис. 1 показано поперечное сечение линии). Линия рассматривается в приближении бесконечно тонкой полоски; потери в металле не учитываются.

Задача о собственных волнах ЭНПЛ может быть сведена к системе интегральных уравнений первого рода [7, 8]

$$\int_{x_1}^{x_2}\hat{K}(x,x')\vec{J}(x')dx'=0 \qquad (x_1<x<x_2) \tag{4}$$

относительно неизвестной векторной функции

$$\vec{J}(x)=\left\{\begin{array}{c}J_z(x)\\ J'_x(x)\end{array}\right\},$$

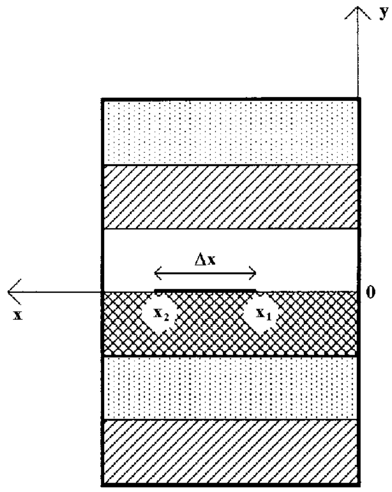


Рис. 1.

где J_z, J_x — составляющие плотности тока на полоске; $J_x' = \dfrac{\partial}{\partial x} J_x$. Элементы ядра представляют собой ряды по тригонометрическим системам:

$$
\begin{aligned}
K_{11}(x,x') &= \frac{2}{a}\sum_{n=1}^{\infty} Z_{n11}\sin\frac{\pi n x}{a}\sin\frac{\pi n x'}{a}\,,\\
K_{12}(x,x') &= -\frac{2}{\pi}\sum_{n=1}^{\infty} \frac{1}{n} Z_{n12}\sin\frac{\pi n x}{a}\sin\frac{\pi n x'}{a}\,,\\
K_{21}(x,x') &= \frac{2}{a}\sum_{n=1}^{\infty} Z_{n21}\cos\frac{\pi n x}{a}\sin\frac{\pi n x'}{a}\,,\\
K_{22}(x,x') &= -\frac{x'}{a} Z_{022} - \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{1}{n} Z_{n22}\cos\frac{\pi n x}{a}\sin\frac{\pi n x'}{a}\,;
\end{aligned}
\tag{5}
$$

выражения для элементов тензора поверхностного импеданса Z_{nij}

Особенности функций $K_{ij}(x,x')$ определяются асимптотическим поведением импедансов при неограниченном возрастании индекса n [7, 8]:

$$
\begin{aligned}
&\lim_{n\to\infty} n Z_{n11} = q_1, \qquad \lim_{n\to\infty} \frac{1}{n} Z_{n22} = q_3,\\
&\lim_{n\to\infty} Z_{n12} = \lim_{n\to\infty} Z_{n21} = q_2,
\end{aligned}
\tag{6}
$$

где константы q_i определяются геометрическими и физическими параметрами линии. На основании (6) легко убедиться в том, что элементы ядра (5) имеют логарифмические особенности $(K_{1j}(x,x'),\ j=1,\,2)$ и сингулярности типа Коши $(K_{2j}(x,x'),\ j=1,\,2)$.

Первый этап алгоритма МПО заключается в выделении особенности в ядре интегрального уравнения, в результате чего в правой части (4) окажется интеграл, соответствующий оператору обращения. Основная идея проводимой здесь модификации метода состоит в том, чтобы в качестве оператора обращения выбрать оператор, аналогичный тому, который возникает при моделировании щелевой линии передачи.

Разложим ядро интегрального уравнения на регулярную и сингулярную части:

$$\hat{K}(x,x') = \hat{K}^{(r)}(x,x') + \hat{K}^{(s)}(x,x').$$

Элементы сингулярной составляющей, соответствующей оператору обращения, выберем следующим образом:

$$\begin{aligned}
K_{11}^{(s)}(x,x') &= \frac{2}{a}q_1\sum_{n=1}^{\infty}\frac{1}{n}\cos\frac{\pi n x}{a}\cos\frac{\pi n x'}{a} = -\frac{q_1}{a}\ln\left|2\left(\cos\frac{\pi x'}{a}-\cos\frac{\pi x}{a}\right)\right|,\\
K_{12}^{(s)}(x,x') &= -\frac{2}{\pi}q_2\sum_{n=1}^{\infty}\frac{1}{n}\cos\frac{\pi n x}{a}\cos\frac{\pi n x'}{a} = \frac{q_2}{\pi}\ln\left|2\left(\cos\frac{\pi x'}{a}-\cos\frac{\pi x}{a}\right)\right|,\\
K_{21}^{(s)}(x,x') &= -\frac{2}{a}q_2\sum_{n=1}^{\infty}\sin\frac{\pi n x}{a}\cos\frac{\pi n x'}{a} = -\frac{q_2}{a}\frac{\sin(\pi x/a)}{\cos(\pi x'/a)-\cos(\pi x/a)},\\
K_{22}^{(s)}(x,x') &= \frac{2}{\pi}q_3\sum_{n=1}^{\infty}\sin\frac{\pi n x}{a}\cos\frac{\pi n x'}{a} = \frac{q_3}{\pi}\frac{\sin(\pi x/a)}{\cos(\pi x'/a)-\cos(\pi x/a)}.
\end{aligned}\tag{7}$$

Естественно, в двух последних рядах, как и в представлениях (5) для соответствующих элементов исходного ядра, сходимость понимается в обобщенном смысле. Разложение

$$\sum_{n=1}^{\infty}\left(\cos\frac{\pi n x}{a}\sin\frac{\pi n x'}{a}+\sin\frac{\pi n x}{a}\cos\frac{\pi n x'}{a}\right) = \frac{1}{2}\operatorname{ctg}\pi\frac{x+x'}{2a},\tag{8}$$

а также соотношения (3) дают основания утверждать, что элементы тензора $\hat{K}^{(r)}(x,x')$ ограничены на квадрате $[x_1<x<x_2]\times[x_1<x'<x_2]$.

3. ИНТЕГРАЛЬНЫЕ УРАВНЕНИЯ ВТОРОГО РОДА

Последующие этапы МПО [7, 8] не нуждаются в корректировке, поэтому мы остановимся на них достаточно кратко. За выделением особенности в ядре следует ортогонализующая подстановка

$$\cos\frac{\pi x}{a} = c + su, \qquad \cos\frac{\pi x'}{a} = c + sv,\tag{9}$$

где

$$\left\{\begin{array}{c} c\\ s\end{array}\right\} = \frac{1}{2}\left(\cos\frac{\pi x_1}{a}\pm\cos\frac{\pi x_2}{a}\right),$$

позволяющая перейти от “экранных"тригонометрических систем к степенным функциям u^m, v^m и, далее, к многочленам Чебышева, ортогональным на новом интервале определения интегральных уравнений $(-1<u<1)$. После применения преобразований (9), как легко убедиться, в представлении элементов ядра $\hat{K}^{(s)}$ оператора обращения будут явным образом присутствовать логарифмическая особенность вида $\ln|v-u|$ и сингулярность типа Коши $(v-u)^{-1}$.

Для интегральных уравнений

$$\frac{1}{\pi}\int_{-1}^{1}\ln|v-u|\varphi(v)dv = f(u),$$

$$\frac{1}{\pi}\int_{-1}^{1}\frac{1}{v-u}\psi(v)dv = g(u) \qquad (-1<u<1)$$

известны следующие формулы обращения:

$$\begin{aligned}\varphi(u) &= \frac{1}{\pi\sqrt{1-u^2}}\left[\int\limits_{-1}^{1}\frac{\sqrt{1-v^2}}{v-u}f'(v)dv - \frac{1}{\ln 2}\int\limits_{-1}^{1} f(v)\frac{dv}{\sqrt{1-v^2}}\right],\\ \psi(u) &= \frac{a_0}{\sqrt{1-u^2}} - \frac{1}{\pi\sqrt{1-u^2}}\int\limits_{-1}^{1}\frac{\sqrt{1-v^2}}{v-u}g(v)dv.\end{aligned} \tag{10}$$

В последнем соотношении присутствует неопределенная постоянная a_0.

Применение процедуры обращения (10) к интегральным соотношениям с ядром $\hat{K}^{(s)}$ приводит к векторному интегральному уравнению второго рода

$$\hat{q}\vec{\eta}(u) = \int\limits_{-1}^{1}\hat{G}(u,v)\vec{\eta}(v)dv + \frac{1}{\sqrt{1-u^2}}\vec{A}_0 \qquad (-1<u<1), \tag{11}$$

где введены обозначения

$$\hat{q} = \begin{pmatrix} q_1 & q_2\\ q_2 & q_3\end{pmatrix}, \qquad \vec{A}_0 = \begin{pmatrix} 0\\ a_0\end{pmatrix},$$

$$\vec{\eta}(u) = \begin{pmatrix}\eta_1(u)\\ \eta_2(u)\end{pmatrix} = \frac{1}{\sqrt{1-(c+su)^2}}\begin{pmatrix} J_z\left(\dfrac{a}{\pi}\arccos(c+su)\right)\\ -\dfrac{a}{\pi}J_x'\left(\dfrac{a}{\pi}\arccos(c+su)\right)\end{pmatrix}. \tag{12}$$

Элементы ядра уравнения (11) имеют вид

$$\begin{aligned}\left\{\begin{matrix}G_{11}(u,v)\\ G_{12}(u,v)\end{matrix}\right\} &= \frac{1}{\pi^2\sqrt{1-u^2}}\left[\pi\left\{\begin{matrix}q_1\\ q_2\end{matrix}\right\}\frac{\ln 2s}{\ln 2}+\right.\\ &+2\sum_{n=1}^{\infty}\left[\left\{\begin{matrix}nZ_{n11}\\ Z_{n12}\end{matrix}\right\}\left(s\xi_n(u)-\frac{1}{n\ln 2}\tau_n\right)\sqrt{1-(c+sv)^2}\times\right.\\ &\left.\left.\times\, U_{n-1}(c+sv) - \left\{\begin{matrix}q_1\\ q_2\end{matrix}\right\}\left(s\Omega_n(u)-\frac{1}{n\ln 2}\Lambda_n\right)T_n(c+sv)\right]\right],\\ \left\{\begin{matrix}G_{21}(u,v)\\ G_{22}(u,v)\end{matrix}\right\} &= \frac{s}{\pi^2\sqrt{1-u^2}}\left[\left\{\begin{matrix}0\\ Z_{022}\xi_0(u)\arccos(c+sv)\end{matrix}\right\}+\right.\\ &+\ 2\sum_{n=1}^{\infty}\left[\left\{\begin{matrix}Z_{n21}\\ \frac{1}{n}Z_{n22}\end{matrix}\right\}\xi_n(u)\sqrt{1-(c+sv)^2}U_{n-1}(c+sv)-\right.\\ &\left.\left.-\ \left\{\begin{matrix}q_2\\ q_3\end{matrix}\right\}\Omega_n(u)T_n(c+sv)\right]\right],\end{aligned} \tag{13}$$

где $T_n,\ U_n$ — многочлены Чебышева первого и второго рода,

$$\Omega_n(u) = \int\limits_{-1}^{1}\frac{\sqrt{1-v^2}}{v-u}U_{n-1}(c+sv)dv,$$

$$\xi_n(u) = -\int_{-1}^{1} \frac{\sqrt{1-v^2}}{v-u} \frac{T_n(c+sv)}{\sqrt{1-(c+sv)^2}} dv,$$

$$\Lambda_n = \int_{-1}^{1} T_n(c+sv) \frac{dv}{\sqrt{1-v^2}},$$

$$\tau_n = \int_{-1}^{1} \sqrt{1-(c+sv)^2}\, U_{n-1}(c+sv) \frac{dv}{\sqrt{1-v^2}}.$$

В связи с тем, что интегралы $\Omega_n(u)$ и Λ_n вычисляются в квадратурах, для унификации численных расчетов удобно использовать следующие представления:

$$\xi_0(u) = \sum_{m=1}^{\infty} m\chi_m \Omega_m(u),$$

$$\xi_n(u) = \frac{1}{n} \sum_{m=1}^{\infty} m r_m(n) \Omega_m(u) \qquad (n = 1,\, 2,\, \ldots), \tag{14}$$

$$\tau_n = \sum_{m=0}^{\infty} r_m(n) \Lambda_m,$$

где

$$\chi_m = \frac{2}{a\,(1+\delta_{0m})} \int_0^a \frac{\pi x}{a} \cos \frac{\pi m x}{a} dx,$$

$$r_m(n) = \frac{2}{a\,(1+\delta_{0m})} \int_0^a \sin \frac{\pi n x}{a} \cos \frac{\pi m x}{a} dx,$$

δ_{nm} — символ Кронекера.

4. АЛГЕБРАИЗАЦИЯ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ ДЛЯ ДВУХСЛОЙНОЙ СТРУКТУРЫ С НЕСИММЕТРИЧНЫМ РАСПОЛОЖЕНИЕМ ПОЛОСКИ

Представление неизвестных функций в виде линейных комбинаций ортогональных многочленов

$$\eta_j(u) = \frac{1}{\sqrt{1-u^2}} \sum_{n=0}^{N} \eta_{jn} T_n(u) \quad (j = 1,\, 2) \tag{15}$$

и учет в бесконечных суммах (13), (14) первых N слагаемых позволяют свести интегральное уравнение (11) к однородной системе линейных алгебраических уравнений относительно постоянных коэффициентов η_{jn}.

Наличие в (11) неопределенного параметра a_0 требует привлечения некоторого дополнительного условия. В качестве такового выбирается условие на ребрах металлической полоски

$$J_x|_{x=x_1,x_2} = 0,$$

непосредственно приводящее к равенству

$$\eta_{20} = 0.$$

Рассмотрим экранированную несимметричную полосковую линию передачи на однослойной подложке с несимметричным расположением полоски относительно боковых стенок экрана (рис. 2). Такая линия передачи рассматривается, по-видимому, впервые. Параметры задачи были выбраны следующим образом: $\dfrac{y_1}{a} = 0{,}125$; $\dfrac{y_2 - y_1}{a} = 0{,}375$; проницаемости диэлектрических слоев $\varepsilon^{(1)} = 9$, $\varepsilon^{(2)} = 1$, $\mu^{(1)} = \mu^{(2)} = 1$; нормированное волновое число $k \times a = 3{,}2\dfrac{\pi}{3}$. На рис. 3 представлена зависимость постоянных распространения собственных волн ЭНПЛ γ от ширины полоски $\triangle x$ при симметричном ее расположении внутри экрана. Здесь кривые 2 и 4 описывают подполосочные H_{10} и H_{30}-волны, которым соответствует модель прямоугольного волновода с сечением $\triangle x \times y_1$ и проницаемостью диэлектрического заполнения $\varepsilon^{(1)}$. Линия 3 на интервале $\triangle x < \dfrac{a}{2}$ отвечает подполосочной H_{20}-волне, а при уменьшении ширины полоски переходит в экранную LM_{11}-волну, характерную для частично заполненного прямоугольного волновода без металлической полоски. Кривая 5 при $\dfrac{\triangle x}{a} \to 1$ соответствует волне H_{10} прямоугольного волновода сечением $a \times (y_2 - y_1)$ без диэлектрического заполнения.

На рис. 4 построена зависимость постоянных распространения от координаты центра полоски x_0 при фиксированной ее ширине $\left(\dfrac{\triangle x}{a} = 0{,}2\right)$. Нами рассмотрен интересный случай, когда одна из высших собственных волн при определенном расположении полоски $\left(0{,}16 < \dfrac{x_0}{a} < 0{,}31\right)$ переходит в разряд нераспространяющихся ($\mathrm{Re}\,\gamma = 0$). Таким образом, перемещение полоски внутри экрана при определенных условиях может рассматриваться как один из способов подавления высших собственных волн. Из рис. 3 и 4 следует, что характеристики основной волны слабо зависят от геометрии полоски.

На рис. 5 приведены распределения компонент тока на полоске для основной (а) и высшей (б) собственных волн ЭНПЛ. $\left(\dfrac{x_0}{a} = 0{,}25\,,\ \dfrac{\triangle x}{a} = 0{,}2\right)$. Кривые, соответствующие основной моде, как этого и следовало ожидать для квази-T-волны, характеризуются большим различием масштабов поперечной и продольной компонент тока. Отметим также, что все интегралы, возникающие при переходе от (12), (15) к истинному распределению $J_x(x)$, легко могут быть найдены аналитически.

Рис. 6 и 7 содержат информацию о сходимости модифицированного МПО соответственно для основной и высшей собственных волн ЭНПЛ в случае $\dfrac{x_0}{a} = 0{,}25$, $\dfrac{\triangle x}{a} = 0{,}2$. Здесь приведены зависимости постоянных распространения от числа N слагаемых в представлениях (15). Центры интервалов, ограниченных пунктирными линиями, соответствуют значениям γ, рассчитанным в приближении $N = 30$. Радиус интервала δ для основной волны составляет 0,1 %, для волны высшего типа — 1 % от его среднего значения. Как видно из рисунков, стабилизация численного алгоритма для первой моды происходит уже при $N > 6$. При построении характеристик ЭНПЛ (рис. 3, 4) в суммах (15) было учтено $N = 20$ слагаемых, распределению токов на полосках (рис. 5) соответствует $N = 6$.

5. ЗАКЛЮЧЕНИЕ

Таким образом, предложенная модификация МПО позволила на основе интегральных уравнений второго рода построить более строгую математическую модель экранированной несимметричной полосковой линии передачи, позволяющую анализировать собственные волны высших типов. Проведенный анализ внутренней сходимости метода показал необходимость в более высоких приближениях при

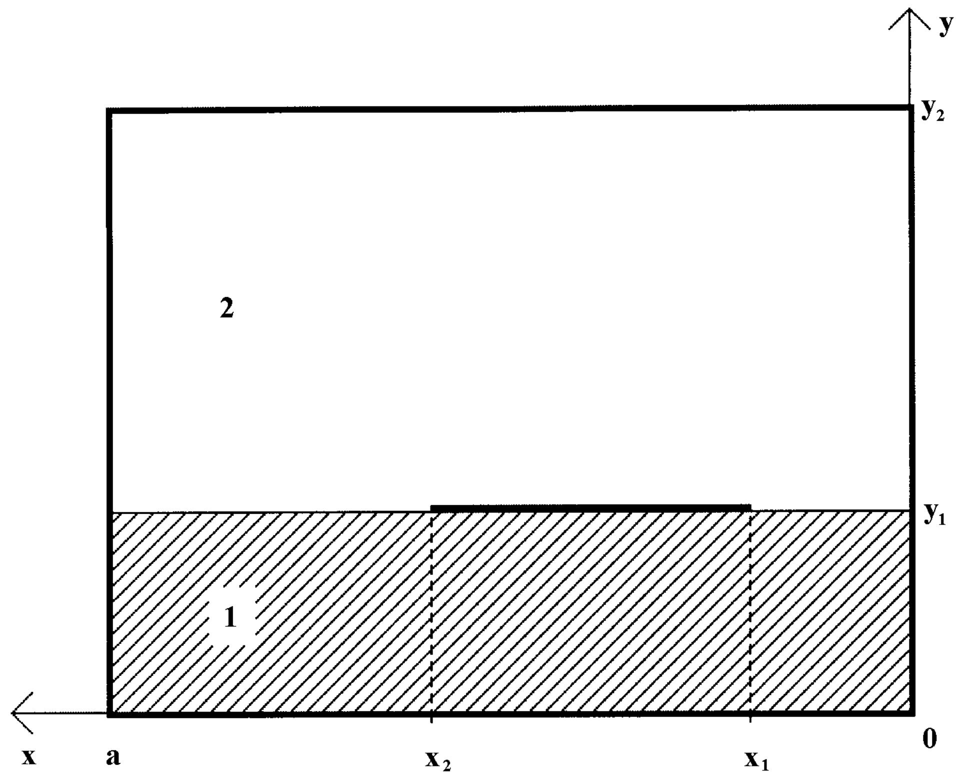


Рис. 2.

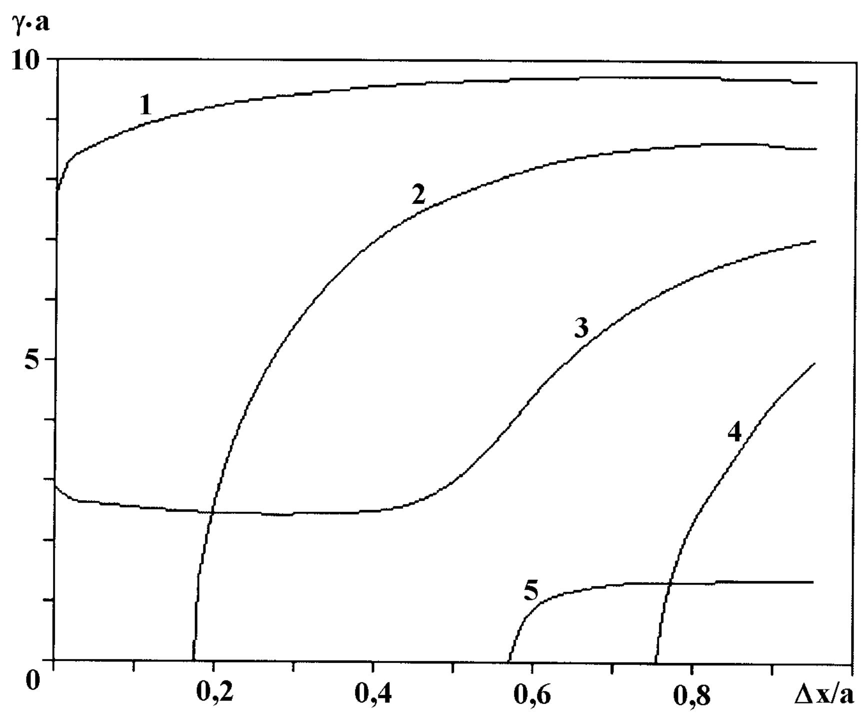


Рис. 3.

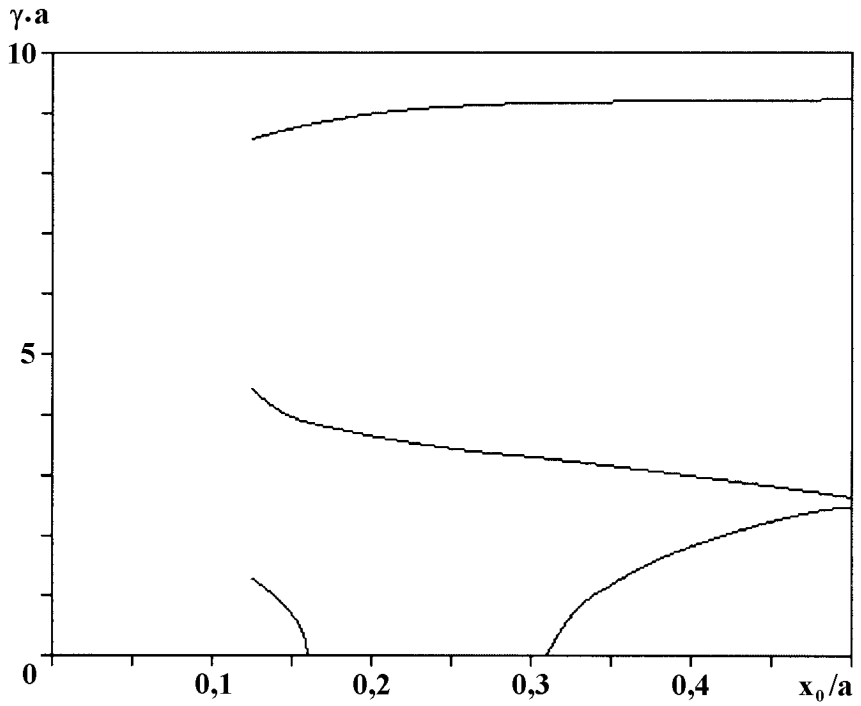


Рис. 4.

исследовании характеристик высших мод. Учитывая то обстоятельство, что в научной литературе отсутствует информация о сходимости по высшим собственным волнам, с нашей точки зрения, необходим критический пересмотр численных результатов, полученных для высших мод полосковых линий передачи с помощью прямых проекционных схем без предварительного выделения особенностей в интегральных уравнениях.

ЛИТЕРАТУРА

1. Гвоздев В. И., Нефедов Е. И. Объемные интегральные схемы СВЧ. — М.: Наука, 1985. – 256 с.
2. Автоматизированное проектирование устройств СВЧ/ В. В. Никольский, В. П. Орлов, В. Г. Феоктистов и др. — М.: Радио и связь, 1982. – 272 с.
3. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. — М.: Наука, 1986. – 288 с.
4. Неганов В. А. //Радиотехника и электроника, 1986. Т. 31. № 3. С. 479.
5. Неганов В. А., Нефедов Е. И. // ДАН СССР, 1988. Т. 299. № 5. С. 1124.
6. Неганов В. А. //Радиотехника и электроника, 1989. Т. 34. № 11. С. 2251.
7. Неганов В. А. Электродинамическая теория полосково—щелевых структур СВЧ. — Самара: Изд-во Самарского ун-та, 1991. – 240 с.
8. Неганов В. А., Нефедов Е. И., Яровой Г. П. Полосково—щелевые структуры сверх- и крайневысоких частот. — М.: Наука. Физматлит, 1996. – 304 с.
9. Левин Л. Теория волноводов: Пер. с англ./ Под ред. В. И. Вольмана. — М.: Радио и связь, 1981. – 312 с.

Поволжский институт
информатики, радиотехники и
связи, г. Самара, Россия

Поступила в редакцию
15 февраля 1998 г.

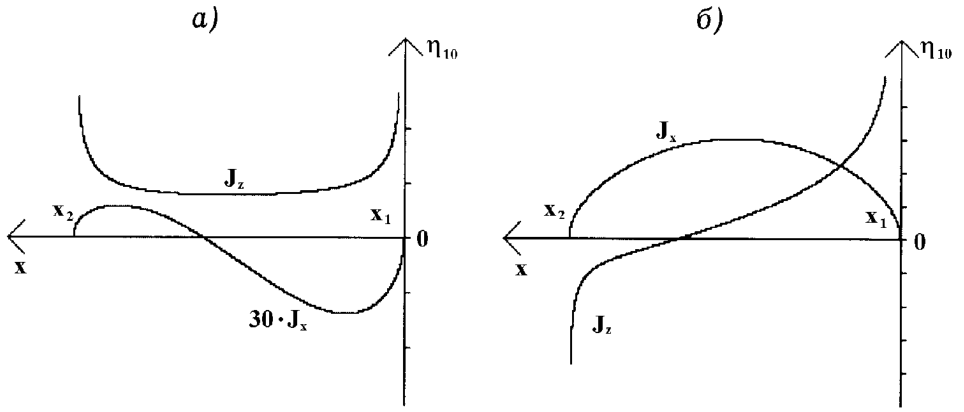


Рис. 5.

THE MODIFIED METHOD OF ALMOST-COMPLETE INVERSION OF SINGULAR INTEGRAL OPERATOR IN THE THEORY OF SHIELDED STRIP TRANSMISSION LINES

A. S. Aref'ev, V. A. Neganov

The modification of the method of operator almost-complete inversion, which enables one to examine shielded strip structures by means of integral equations of the second kind, is carried out. The higher modes of the shielded asymmetric strip line are investigated taking into account the asymmetry of strip location relative to the shield center.

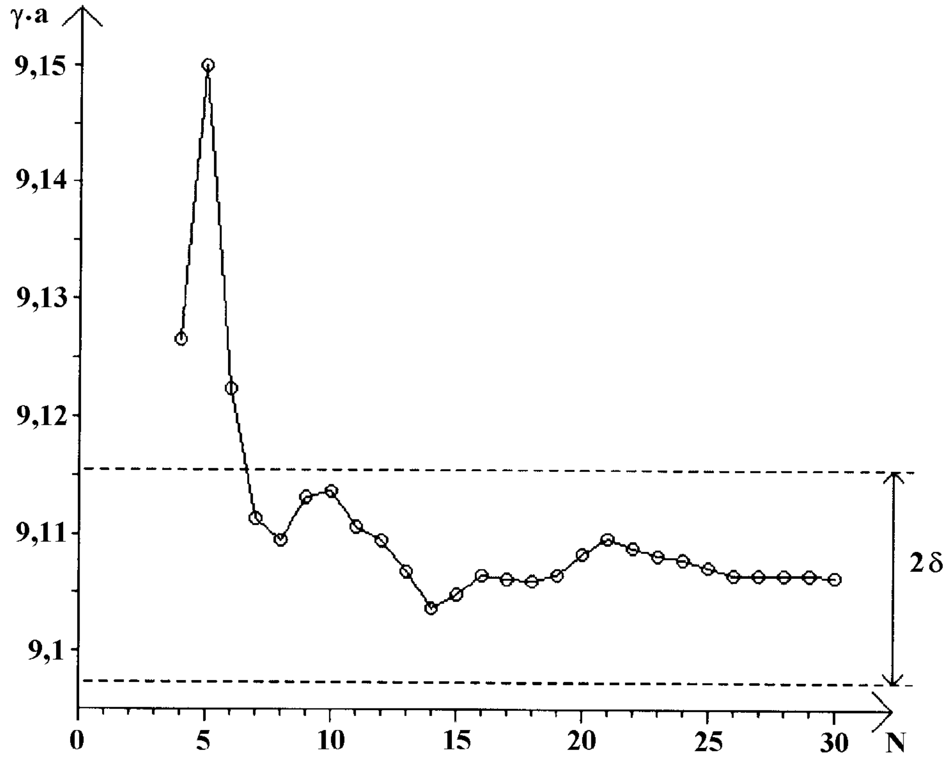

γ·a
9,15
9,14
9,13
9,12
9,11
9,1
2δ
0
5
10
15
20
25
30
N

Рис. 6.

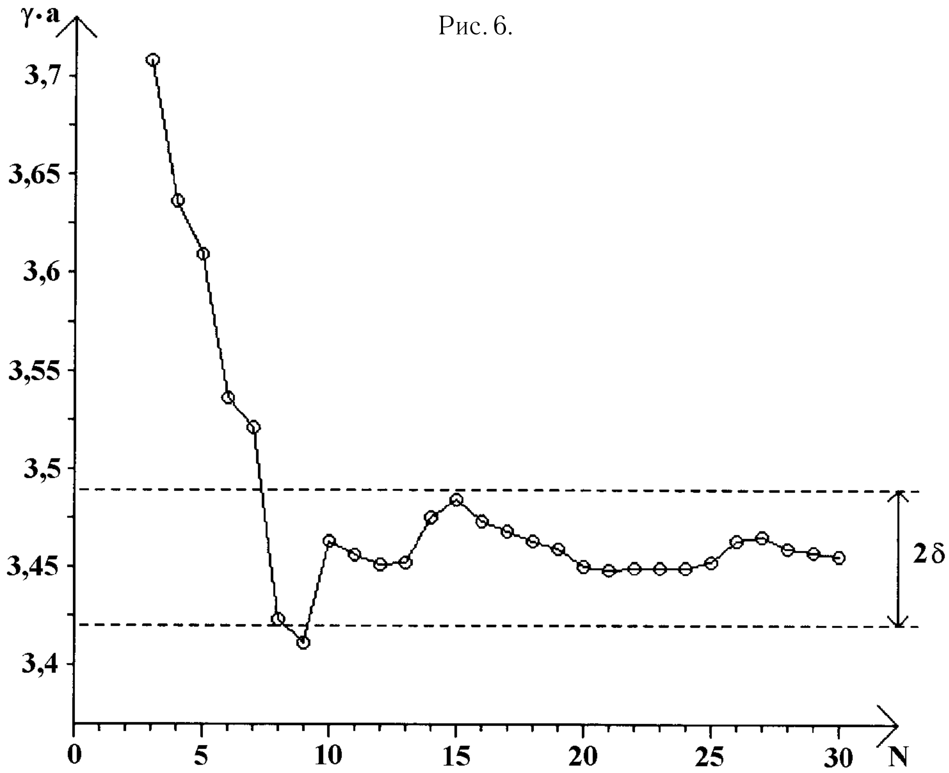

γ·a
3,7
3,65
3,6
3,55
3,5
3,45
3,4
2δ
0
5
10
15
20
25
30
N

Рис. 7.

УДК 621.372.8.001.24

ЭЛЕКТРОДИНАМИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ ЭЛЕКТРОМАГНИТНЫХ ПОЛЕЙ В Н-ВОЛНОВОДАХ

Г. Ф. Заргано, К. В. Вдовенко, Г. П. Синявский

Изложена методика электродинамического моделирования и визуализации структуры электромагнитных полей Н- и Е-волн в Н-волноводе на ЭВМ. Собственные числа и собственные функции Н-волновода определялись методом частичных областей с учетом особенности электромагнитного поля на ребре. В качестве независимого параметра при моделировании использовалась длина силовой линии электромагнитного поля.

ВВЕДЕНИЕ

Волноводы со сложной формой поперечного сечения, к которым относятся Н-волноводы, находят все более разнообразное применение при конструировании различных СВЧ устройств специального назначения. Так Н-волноводы используются в измерительной аппаратуре и в приемных трактах, когда необходимы широкополосность, малое значение характеристического сопротивления, малые габариты и вес аппаратуры. Они успешно применяются при разработке всевозможных электронных СВЧ-приборов и ферритовых устройств. В этом случае требуются типы волн, имеющие продольные составляющие электрического и магнитного полей. С учетом оптимальности потерь и габаритов Н-волноводы целесообразнее использовать в сантиметровом диапазоне длин волн. При одинаковой полосе частот одномодового режима предельная мощность Н-волновода в два раза больше предельной мощности П-волновода соответствующих размеров [1–5].

Сложность рассматриваемой структуры и приближенное решение задачи о собственных числах и собственных функциях Н-волноводов делают актуальной задачу электродинамического моделирования в них пространственных структур электромагнитных полей. При этом знание картин электромагнитных полей высших типов волн особенно необходимо не только при проектировании устройств на регулярных отрезках Н-волноводов, но и при анализе параметров различных неоднородностей, селективных устройств и многомодового режима работы. С использованием возможностей аппаратных средств современных ПЭВМ появились новые направления в изучении электромагнитных полей сложных структур и машинном проектировании на этой основе сложных СВЧ устройств [6–8].

Моделирование и визуализация пространственной структуры на ПЭВМ различных типов Н- и Е-волн в Н-волноводе связаны со сложностью теоретических и численных расчетов критических волновых чисел и всех шести компонент электромагнитного поля как функций пространственных координат. Все компоненты электромагнитного поля должны рассчитываться с высокой степенью точности, чтобы обеспечивалась хорошая “сшиваемость"полей на смежных границах частичных областей. В противном случае невозможно, даже при однородном диэлектрическом заполнении Н-волновода, обеспечить гладкость силовых линий поля, при их переходе через смежную границу частичных областей.

1. МЕТОД РАСЧЁТА СОБСТВЕННЫХ ЧИСЕЛ И СОБСТВЕННЫХ ФУНКЦИЙ

В качестве метода расчета критических волновых чисел и компонент электромагнитных полей в Н-волноводе с воздушным заполнением ($\varepsilon = 1$, $\mu = 1$) без потерь использовался хорошо зарекомендовавший себя метод частичных областей (МЧО) с учетом особенности поведения электромагнитного поля на ребре [3, 9].

В соответствии с МЧО поперечное сечение Н-волновода разбивается на четыре одинаковые Г-образные области (рис. 2), в которых выделяются две частичные прямоугольные области, в каждой из которых записываются решения двумерного уравнения Гельмгольца для продольных компонент электромагнитного поля — H_z для Н-волн, E_z для Е-волн — в следующем виде:

$$W^{\mathrm{a}\nu}(x,y) = \sum_{m=0}^{M} A_m^{\mathrm{a}\nu} X_m^{\mathrm{a}\nu}(x) Y_m^{\mathrm{a}\nu}(y), \quad \mathrm{a} = \mathrm{I},\ \mathrm{II}, \tag{1}$$

где

$$W^{\mathrm{a}\nu}(x,y) = \begin{cases} H_z^{\mathrm{a}}(x,y) & \text{для Н-волн } (\nu = h), \\ E_z^{\mathrm{a}}(x,y) & \text{для Е-волн } (\nu = e), \end{cases}$$

$$X_m^{\mathrm{Ih}}(x) = \cos p_m^{\mathrm{I}}(l-x), \quad X_m^{\mathrm{Ie}}(x) = \sin p_m^{\mathrm{I}}(l-x),$$

$$X_m^{\mathrm{IIh}}(x) = \cos(p_m^{\mathrm{II}}x - 0{,}5\pi n_2), \quad X_m^{\mathrm{IIe}}(x) = \sin(p_m^{\mathrm{II}}x + 0{,}5\pi n_2),$$

$$Y_m^{\mathrm{ah}}(y) = \chi_m^{\mathrm{a}} \cos(\alpha_m^{\mathrm{a}} y - 0{,}5\pi n_1), \quad Y_m^{\mathrm{ae}}(y) = \chi_m^{\mathrm{a}} \sin(\alpha_m^{\mathrm{a}} y + 0{,}5\pi n_1),$$

$$\alpha_m^{\mathrm{I}} = \frac{\pi}{h}\,(m + 0{,}5n_1)\,, \quad \alpha_m^{\mathrm{II}} = \frac{\pi}{c}\,(m + 0{,}5n_1)\,, \quad (p_m^{\mathrm{a}})^2 = (k^{n\nu})^2 - (\alpha_m^{\mathrm{a}})^2,$$

$$\chi_m^{\mathrm{I}} = \left(\frac{2 - \delta_{0\alpha^{\mathrm{I}}}}{h}\right)^{\frac{1}{2}}, \quad \chi_m^{\mathrm{II}} = \left(\frac{2 - \delta_{0\alpha^{\mathrm{II}}}}{c}\right)^{\frac{1}{2}}, \quad \delta_{0\alpha^{\mathrm{a}}} = \begin{cases} 1, & \alpha^{\mathrm{a}} = 0, \\ 0, & \alpha^{\mathrm{a}} \neq 0, \end{cases}$$

где $n_i = 0$ ($i = 1,\ 2$), если на i-ом отрезке Г-образного контура задано граничное условие типа электрической стенки, $n_i = 1$, если условие типа магнитной стенки; $k^{n\nu}$ — критическое волновое число n-й волны (например, для волны типа $\mathrm{H_{oe}}$ или $\mathrm{E_{oe}}$ оно равно $k_{\mathrm{oe}}^{n\nu} = 2\pi/\lambda_{\mathrm{oe}}^{n\nu}$); индекс n показывает порядковый номер волны; индекс “e"(“even-— четный) — граничное условие соответствует электрической стенке; индекс “o"(“odd-— нечетный) — магнитной (первый подстрочный индекс относится к изменению поля по оси x, второй — по оси y).

Амплитудные коэффициенты $A_m^{\mathrm{a}\nu}$ имеют вид

$$\begin{aligned} &A_m^{\mathrm{a}\nu} = F_m^{\mathrm{a}\nu} \sum_{i=0}^{N} V_i \Phi_{im}^{\mathrm{a}\nu}\,, \\ &F_m^{\mathrm{ah}} = \left(\frac{\partial X_m^{\mathrm{ah}}}{\partial x}\,|_{x=g}\right)^{-1}, \quad F_m^{\mathrm{ae}} = (X_m^{\mathrm{ae}}(g))^{-1}, \\ &\Phi_{im}^{\mathrm{ah}} = J_{2i+1/6+n_1}(\alpha_m^{\mathrm{a}}c)/(\alpha_m^{\mathrm{a}})^{1/6}, \quad \Phi_{im}^{\mathrm{ae}} = J_{2i+13/6-n_1}(\alpha_m^{\mathrm{a}}c)/(\alpha_m^{\mathrm{a}})^{7/6}, \end{aligned} \tag{2}$$

где $J_n(\alpha_m^{\mathrm{a}}c)$ — функция Бесселя 1-го рода.

Неизвестные коэффициенты V_i находятся из решения линейной однородной алгебраической системы

$$\sum_{i=0}^{N} V_i D_{ji}(k^{n\nu}) = 0, \quad j = 0,\ 1,\ 2, \ldots,\ N, \tag{3}$$

где

$$D_{ji}(k^{n\mathrm{h}}) = \sum_{a=\mathrm{I,\,II}} (-1)^{\mathrm{a}} \sum_{m=0}^{M} X_m^{\mathrm{ah}}(g)\, F_m^{\mathrm{ah}}\, \Phi_{jm}^{\mathrm{ah}}\, \Phi_{im}^{\mathrm{ah}} \text{ — для H-волн},$$

$$D_{ji}(k^{ne}) = \sum_{a=\mathrm{I,\,II}} (-1)^{e} \sum_{m=0}^{M} \left(\frac{\partial X_m^{\mathrm{ae}}(x)}{\partial x}\,|_{x=g} \right) F_m^{\mathrm{ae}}\, \Phi_{jm}^{\mathrm{ae}}\, \Phi_{im}^{\mathrm{ae}} \text{ — для E-волн}.$$

Приравнивая определитель $|D_{ji}(k^{n\nu})|$ системы (3) нулю, получаем трансцендентные уравнения для определения собственных чисел $k_{\mathrm{oe}}^{n\nu}$, $k_{\mathrm{ee}}^{n\nu}$, $k_{\mathrm{eo}}^{n\nu}$ и $k_{\mathrm{oo}}^{n\nu}$.

Неизвестные коэффициенты V_i с точностью до постоянного множителя определяем из решения линейной неоднородной алгебраической системы уравнений, которую получаем из однородной системы (3), отбрасывая l-ое уравнение и перенося l-ый столбец в правую часть системы:

$$\sum_{i=0}^{N} V_i D_{ji}(k^{n\nu}) = -V_l D_{jl}(k^{n\nu}), \quad j = 0,\, 1,\, 2, \ldots,\, N, \quad j \neq l. \tag{4}$$

Постоянный множитель (коэффициент V_l) определяется из условия нормировки мощности падающей волны в поперечном сечении H-волновода s:

$$\sum_{a=\mathrm{I,\,II}} \int\limits_{s_a} W_i^{\mathrm{a}\nu}(x,y) W_j^{\mathrm{a}\nu}(x,y)\, ds_a = Q_i^2 \delta_{ij} = 1, \tag{5}$$

где $Q_i = \left\{ \sum\limits_{a=\mathrm{I,\,II}} \int\limits_{s_a} [W_i^{\mathrm{a}\nu}(x,y)]^2 ds_a \right\}^{1/2}$ — норма собственных функций n-го типа волны, δ_{ij} — символ Кронеккера, s_a — площадь частичной области.

Часть сомножителей, входящих в выражение (2), включены и учитываются при их нахождении в (4). Это

$$\frac{(-1)^i \pi c^{5/6} \Gamma(2i + 1/3 + n_1)}{(n_1 + 2i)!\, \Gamma(1/6) \cdot 2^{1/6}} \quad \text{— для H-волн},$$

$$\frac{(-1)^i \pi c^{1\frac{5}{6}} \Gamma(2i + 7/3 + n_1)}{(n_1 + 2i)!\, \Gamma(7/6) \cdot 2^{7/6}} \quad \text{— для E-волн}.$$

Численный анализ показал, что при проведении расчетов компонент электромагнитных полей H-волноводов с достаточной для практики точностью (погрешность меньше 1%) можно ограничиться значениями $N = 3$, $\mathrm{M} = 100$. Погрешность вычисления критических волновых чисел в данном приближении составляла $1 \cdot 10^{-7}$. Максимальная “несшиваемость"компонент полей на линии раздела частичных областей составила 1,2%, при этом почти на всем интервале “сшивания"несовпадение меньше 1%.

Некоторые результаты расчетов нормированных критических волновых чисел двух низших типов H-волн $k_{\mathrm{oe}}^{\mathrm{Ih}} \cdot l$ и $k_{\mathrm{ee}}^{\mathrm{Ih}} \cdot l$ в зависимости от размеров c/l и $2g/l$ H-волновода представлены соответственно на рис. 1 и рис. 2. Эти критические волновые числа в основном определяют ширину рабочей полосы частот одномодового режима работы H-волновода при $h/l = 0{,}463$. Изменение ширины выступа $2g/l$ в большей степени влияет на $k_{\mathrm{ee}}^{\mathrm{Ih}} \cdot l$, чем на $k_{\mathrm{oe}}^{\mathrm{Ih}} \cdot l$. При этом для всех значений c/l при $g = 0{,}5l$ $k_{\mathrm{oe}}^{\mathrm{Ih}} \cdot l$ достигает минимума. Критическое волновое число $k_{\mathrm{ee}}^{\mathrm{Ih}} \cdot l$ волны $\mathrm{H}_{\mathrm{ee}}^{1}$ возрастает с увеличением размера $2g/l$, достигая максимума вблизи $2g = 0{,}5l$, затем убывает до минимального значения, после чего опять возрастает. Такая закономерность наблюдается для всех размеров c/l. Однако, чем меньше c/l, тем больше интервал изменения $k_{\mathrm{ee}}^{\mathrm{Ih}} \cdot l$. На рис. 3 в виде графиков представлена зависимость полосы пропускания одномодового режима, определяемая значениями величин $k_{\rm oe}^{\rm Ih}\cdot l$, $k_{\rm ee}^{\rm Ih}\cdot l$ и $k_{\rm eo}^{\rm Ih}\cdot l$, в зависимости от изменения размеров $2g/l$ и c/l. В области значений, в которой $k_{\rm eo}^{\rm Ih}\cdot l < k_{\rm ee}^{\rm Ih}\cdot l$, она определяется выражением $\Delta = k_{\rm eo}^{\rm Ih}\cdot l - k_{\rm oe}^{\rm Ih}\cdot l$. Для других значений размеров выступов Н-волновода полоса пропускания будет определяться, как $-\Delta = k_{\rm ee}^{\rm Ih}\cdot l - k_{\rm oe}^{\rm Ih}\cdot l$. Штриховая линия на рис. 3 показывает, критическое волновое число какой волны определяет верхнюю границу полосы пропускания одномодового режима ($k^{\rm Ih}\cdot l = k^{\rm Ih}\cdot l$)

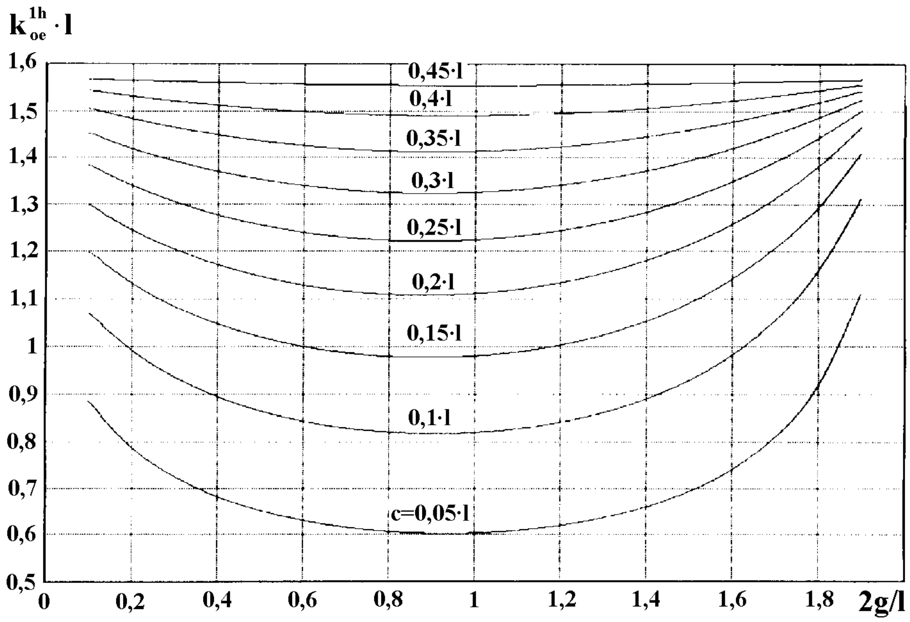


Рис. 1. Зависимость нормированного критического волнового числа $k_{\rm oe}^{\rm Ih}\cdot l$ основной Н-волны от ширины и высоты гребней Н-волновода.

2. МЕТОДИКА МОДЕЛИРОВАНИЯ СТРУКТУРЫ ЭЛЕКТРОМАГНИТНОГО ПОЛЯ

В основе моделирования и визуализации на ПЭВМ электромагнитных полей любых типов волн на рабочей частоте лежит система дифференциальных уравнений, описывающих сложную пространственную структуру силовых линий как электрического, так и магнитного поля, представленная в параметрической форме. В качестве независимого параметра использовалась длина силовой линии $L\,[0,\infty]$ электромагнитного поля. Это связано с тем, что при использовании в качестве независимой переменной одной из координат соответствующие производные остальных координат могут обращаться в бесконечность или быть тождественно равны нулю. Таким образом, учитывая, что $x = x(L)$, $y = y(L)$, $z = z(L)$, имеем следующую систему уравнений:

$$\frac{1}{W_x}\cdot\frac{dx}{dL} = \frac{1}{W_y}\cdot\frac{dy}{dL} = \frac{1}{W_z}\cdot\frac{dz}{dL} = F(L), \quad \vec{W} = \left\{ \begin{array}{c} \vec{H} \\ \vec{E} \end{array} \right\}, \tag{6}$$

где: $F(L)$ — произвольная ограниченная функция, неравная нулю для всей области определения.

Система (6) является системой обыкновенных дифференциальных уравнений 1-го порядка относительно координат точек силовой линии. Представим ее в следующем виде:

$$\left\{ \begin{array}{l} \dfrac{dx}{dL} = F(L)W_x(L), \\ \dfrac{dy}{dL} = F(L)W_y(L), \\ \dfrac{dz}{dL} = F(L)W_z(L), \end{array} \right. \tag{7}$$

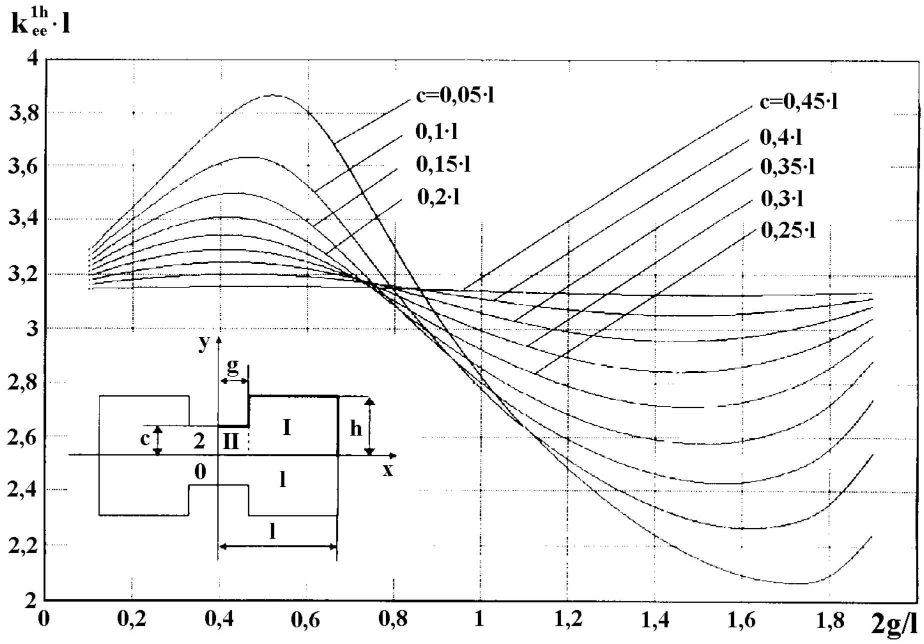


Рис. 2. Зависимость нормированного критического волнового числа $k_{\rm ee}^{\rm Ih}\cdot l$ второй Н-волны от ширины и высоты гребней Н-волновода.

где $F(L) = 1/(W_x^2 + W_y^2 + W_z^2)^{1/2}$.

Использование длины силовой линии L в качестве независимого параметра позволяет проводить моделирование силовой линии независимо от того, является она замкнутой или нет. Так как ее длина всегда положительна и увеличивается при движении вдоль силовой линии, не возникает проблем со сменой независимой переменной при поворотах силовых линий и смене знака приращения независимой переменной.

При реализации изложенной методики моделирования на ПЭВМ использовалась векторная модель, в которой каждая точка пространственной структуры представляется радиус—вектором, восстановленным из некоторого центра, принятого за начало системы координат. Начало системы координат было выбрано в центре симметрии Н-волновода. Изображение пространственной структуры поля получается путем параллельного проектирования всех векторов, задающих структуру поля на картинную плоскость (экран монитора ПЭВМ). Векторная модель позволяет, используя матричный аппарат векторной графики, легко просчитать повороты, растяжение или сжатие объекта в пространстве, а также его проектирование на плоскость, что является необходимым при анализе сложной структуры электромагнитных полей. Система уравнений (7) решалась численно методом Рунге—Кутта 4-го порядка. Относительная погрешность решения системы уравнений (7) составляла 0,1%. Программа моделирования электромагнитных полей написана на языке Borland C++ для ПЭВМ, совместимых с IBM PC. Она имеет модульную структуру. В качестве входных данных программы использовались рассчитанные критические волновые числа $k^{n\nu}$ и амплитудные коэффициенты $A_m^{\mathrm{a}\nu}$ разложения электромагнитных полей в Н-волноводе (2).

Выбор начальной точки для расчета и визуализации одной силовой линии достаточно произволен. Однако, плотность силовых линий должна быть пропорциональна напряженности поля. Чтобы силовые линии поля имели законченный вид и их система наиболее объективно отражала структуру электромагнитного поля, необходимо начинать построение Н-поля для Н-волн в максимуме H_z компоненты поля при $z = 0$, Н-поля для Е-волн — в максимуме E_z компоненты поля при $z = \pm\dfrac{\lambda}{4}$. Начальную точку при построении Е-поля необходимо выбирать на металле.

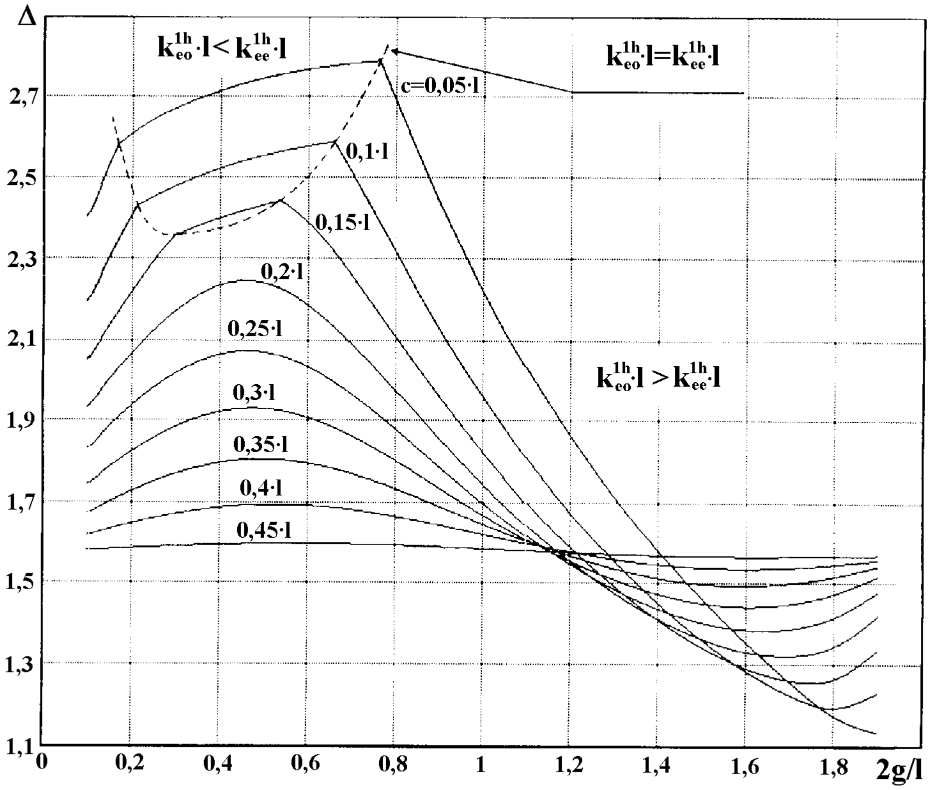


Рис. 3. Зависимость нормированной полосы пропускания одномодового режима от ширины и высоты гребней H-волновода.

3. АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

На рис. 4 изображена структура электромагнитных полей первых H-волн всех типов в H-волноводе на рабочей частоте $k \cdot l = 1{,}5k^{n\nu} \cdot l$ в плоскостной и объемной проекциях. На рис. 5 в аналогичном виде изображена структура первых E-волн. При расчетах и моделировании использовался стандартный H-волновод с полосой пропускания 2,4:1 ($h = 0{,}463l$, $c = 0{,}185l$, $g = 0{,}25l$) [10].

Визуализация картин электромагнитного поля показывает (рис. 4, 5), что по сравнению с прямоугольным волноводом металлические выступы сильно искажают структуру поля всех типов H- и E-волн. Выступы как бы "выдавливают"поле в боковые области H-волновода. При этом силовые линии электрического поля сгущаются вблизи ребер выступов волновода. В структуре электромагнитного поля появляются дополнительные вариации. Одновременно появляется вырождение по критическим волновым числам и структуре силовых линий E-волн. Необходимо заметить, что при некоторых размерах H-волновода волна может оказаться неосновной волной волновода.

Представляют практический интерес картины полей E-волн, магнитное поле которых сосредоточено в боковых областях, а силовые линии электрического поля образуют своеобразные эквипотенциальные трубки. Такую структуру можно использовать в конструкциях электронных приборов СВЧ, в которых осуществляется взаимодействие электронного потока с продольным электрическим полем.

Разработанный алгоритм и программа визуализации и моделирования структуры полей электромагнитных волн универсальны и позволяют строить картины полей в любой открытой или закрытой структуре, если поддаются определению все шесть компонент поля как функции пространственных координат.

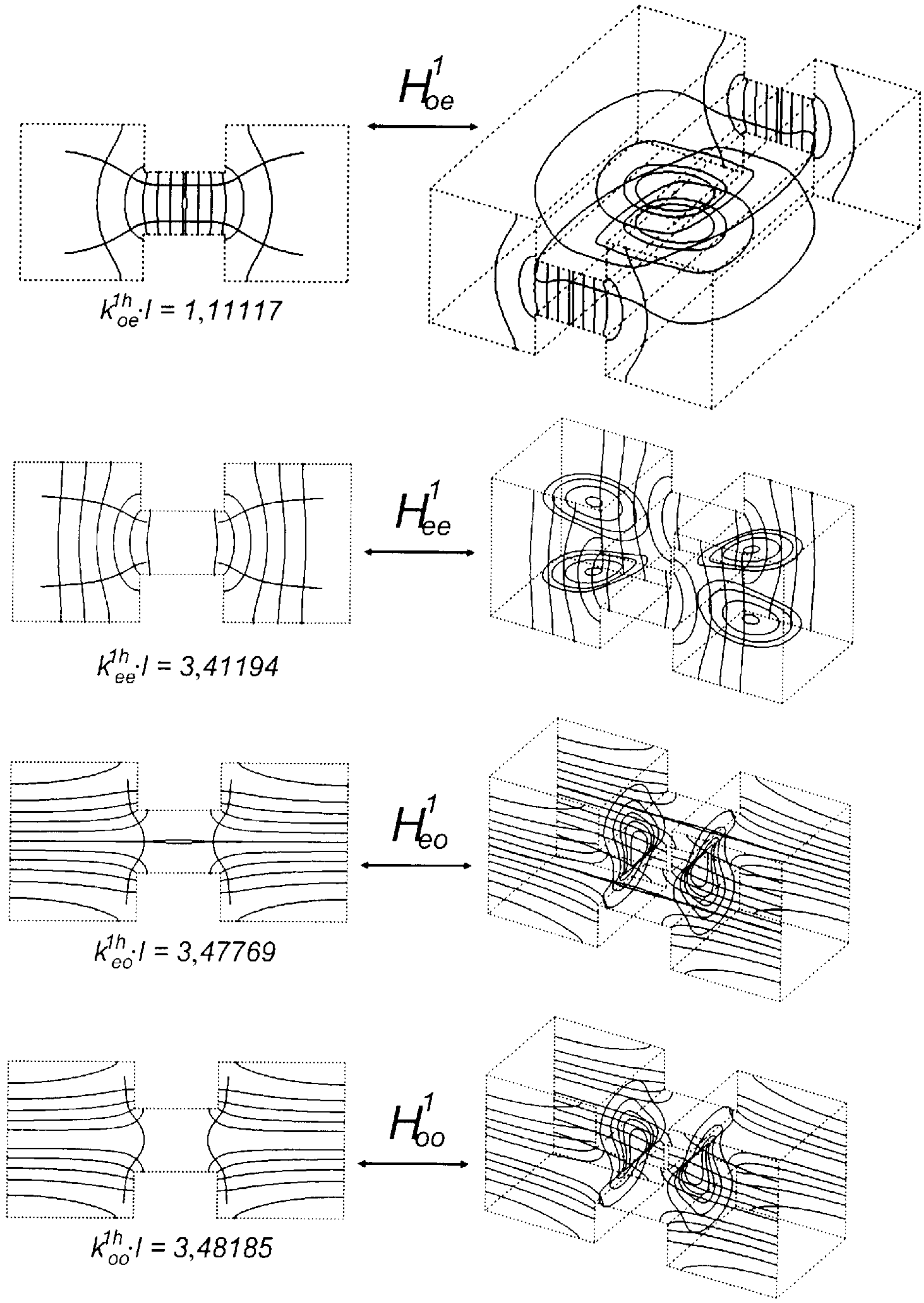


Рис. 4. Пространственная и плоскостная структура электромагнитных полей первых Н-волн всех типов в Н-волноводе.

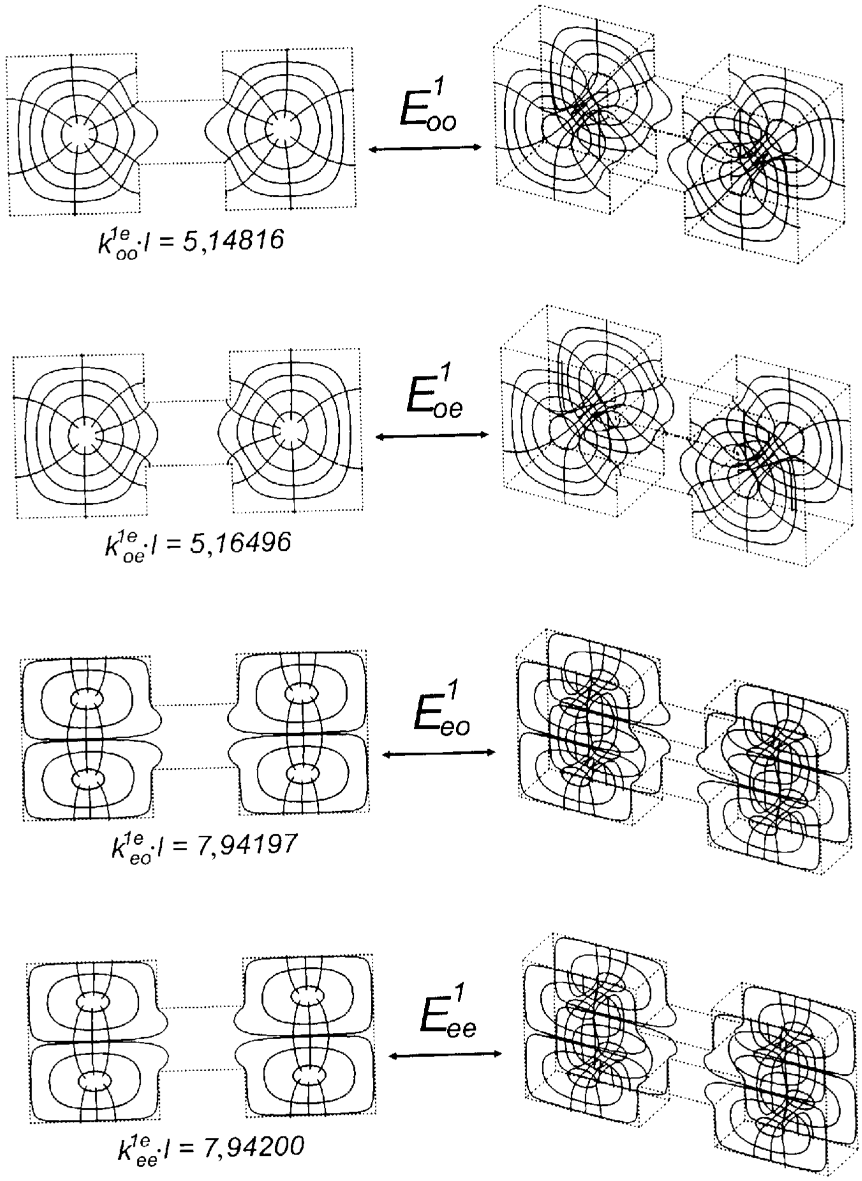


Рис. 5. Пространственная и плоскостная структура электромагнитных полей первых E-волн всех типов в H-волноводе.

ЛИТЕРАТУРА

1. Helszajn J., Caplin M. //Microwave Enginering Europe, 1997. № 5. P. 49.
2. Hoefer W. J. R., Burton M. N. //IEEE Trans. MTT, 1982. V. 30. № 12. P. 2190.
3. Заргано Г. Ф., Ляпин В. П., и др. Волноводы сложных сечений. — М.: Радио и связь, 1986. – 124 с.
4. Ланчук А. С., Фиалковский А. Т., Юнисов Л. Е. //Электронная техника. Сер. Электроника СВЧ, 1985. Вып. 10(382). С. 34.
5. Utsumi Y. //IEEE Trans. MTT, 1985. V. 33. № 2. P. 111.
6. Заргано Г. Ф., Синявский Г. П., Ткаченко В. П. //Изв. ВУЗов. Радиофизика, 1987. Т. 30. № 11. С. 1350.
7. Донченко В. А., Заргано Г. Ф., Синявский Г. П. //Изв. ВУЗов. Радиофизика, 1997. Т. 40. № 10. С. 1286.
8. Никольская Т. И., Никольский В. В. //Межвуз. сб. научных трудов МИРЭА. — М., 1990. — 160 с.
9. Lypin V. P., Mikhalevsky V. S., Sinyavsky G. P. //IEEE Trans. MTT, 1982. V. 30. № 7. P. 1107.
10. Фельдштейн А. П., Явич Л. Р., Смирнов В. П. Справочник по элементам волноводной техники. — М.: Госэнергоиздат, 1963.

Ростовский госуниверситет,
г. Ростов-на-Дону, Россия

Поступила в редакцию
19 января 1997 г.

ELECTRODYNAMIC MODELING OF THE SPATIAL STRUCTURE OF THE ELECTROMAGNETIC FIELDS IN H-WAVEGUIDE

G. F. Zargano, K. V. Vdovenko, G. P. Sinyavskij

A method of electrodynamic computer modeling of H and E wave electromagnetic fields spatial structures in H-waveguide is described. The vector eigenfunctions and the cutoff wavenumbers were obtained using the method of partial domains taking into account singularities of the electromagnetic field at the edge. The length of electromagnetic field line was used as an independent parameter for modeling.

УДК 621.373.7

РАСПРОСТРАНЕНИЕ ОПТИЧЕСКИХ ИМПУЛЬСОВ В ПЕРИОДИЧЕСКИХ НЕЛИНЕЙНЫХ ВОЛОКНАХ

И. О. Золотовский, Д. И. Семенцов

Исследуются условия формирования и распространения солитонных импульсов в периодическом двухмодовом нелинейном волокне с учетом дисперсии материала волокна и межмодовой дисперсии взаимодействующих мод при условии их фазового синхронизма.

Вопросам распространения и преобразования мод в волоконных световодах с различным профилем показателя преломления в сечении световода, но однородных по его длине, посвящено достаточно много работ [1, 2]. Важным типом неоднородности является также периодическое по длине распределение показателя преломления, широко используемое в планарных пассивных и активных волноводах [3, 4]. В [5, 6] показана возможность создания периодического распределения показателя преломления по длине световода, легированного германием, после воздействия на него непрерывного излучения аргонового лазера вблизи длины волны, близкой к 0,5 мкм. В [7] одномодовый световод с наведенной в нем решеткой использовался как узкополосный брегговский фильтр. В [8, 9] показана возможность сжатия гауссова импульса в линейном и нелинейном режимах его распространения в периодическом двухмодовом волокне в условиях фазового синхронизма мод, формирующих импульс. В настоящей работе проводится анализ зависимости дисперсионных свойств данного типа световода от его параметров и вводимого в него импульса, а также влияния эффективной дисперсии волокна на условия формирования в нем солитонных импульсов.

1. Рассмотрим двухмодовое волокно с показателем преломления, зависящим от координат следующим образом:

$$n(r,z) = n_0\{1 - f(r)[1 + m\varphi(z)]\}^{1/2}, \tag{1}$$

где n_0 — показатель преломления на оси волокна; функции $f(r)$ и $\varphi(z)$ определяют распределение оптической неоднородности по сечению и длине волокна; параметр $m \ll 1$ определяет глубину модуляции показателя преломления по длине волокна. Пусть на вход волокна с координатой $z = 0$ подается импульс длительностью τ_0 с пиковым значением входной амплитуды A_0. Поле в световоде представим в виде суперпозиции двух собственных мод невозмущенного периодичностью волокна:

$$\mathbf{E}(t,r,z) = \sum_j \mathbf{e}_j A_j(t,z) U_j(r) \exp[i(\omega_0 t - \beta_j z)], \quad j = 1,\, 2, \tag{2}$$

где β_j — константа распространения соответствующей моды; $U_j(r)$ — профильная функция, описывающая радиальное распределение поля моды по сечению волновода; $\mathbf{e}_j$ — единичные векторы поляризации мод; $A_j(t,z)$ — временные огибающие (амплитуды) собственных мод; ω_0 — центральная частота вводимого волнового пакета. Эффективная связь между распространяющимися в волокне модами имеет место при условии их фазового синхронизма, которое с учетом периодичности волокна и импульсного его возбуждения должно выполняться на центральной частоте и имеет вид

$$\beta_1(\omega_0) - \beta_2(\omega_0) - 2\pi\iota/\Lambda = 0, \tag{3}$$

где Λ — период функции $\varphi(z)$, ι — номер гармоники разложения этой функции в ряд Фурье. В однородном по длине двухмодовом волокне выполнение условия фазового синхронизма $\beta_1(w) - \beta_2(w) = 0$ не может быть реализовано, ввиду различия констант распространения в диапазоне рабочих частот (например, для мод LP_{01} и LP_{02}). Использование периодичности позволяет легко добиться выполнения условия фазового синхронизма на одной из гармоник функции $\varphi(z)$. В области параметров, где это условие близко к выполнению, уравнения связанных волн для временных огибающих мод в импульсе, записанные в координатах бегущего времени $\tau = t - z/u$ (где u — групповая скорость волнового пакета), межмодовой расстройки групповых скоростей, материальной дисперсии и нелинейных эффектов (фазовой самомодуляции и кроссмодуляции) могут быть представлены следующим образом [1, 9]:

$$\begin{aligned} \frac{\partial A_1}{\partial z} - \frac{1}{\upsilon_1}\frac{\partial A_1}{\partial \tau} - i\frac{d}{2}\frac{\partial^2 A_1}{\partial \tau^2} &= -i\sigma_{12}A_2 - iR(\gamma_{11}|A_1|^2 + 2\gamma_{12}|A_2|^2)A_1, \\ \frac{\partial A_2}{\partial z} + \frac{1}{\upsilon_2}\frac{\partial A_2}{\partial \tau} - i\frac{d}{2}\frac{\partial^2 A_2}{\partial \tau^2} &= -i\sigma_{21}^*A_1 - iR(2\gamma_{21}|A_1|^2 + \gamma_{22}|A_2|^2)A_2. \end{aligned} \tag{4}$$

Здесь введены обозначения $\upsilon_j^{-1} = \dfrac{u_1 - u_2}{u_j(u_1 + u_2)}$, где $u_j = (\partial\beta_j/\partial\omega)_0^{-1}$ — групповая скорость j-й моды; материальная дисперсия волокна d; параметр нелинейности световода $R = \dfrac{\omega n_2}{2c}$, где n_2 — коэффициент нелинейности; коэффициенты модовой связи и параметры фазовой самомодуляции и кроссмодуляции определяются соответствующими интегралами перекрытия профильных функций волноводных мод

$$\begin{aligned} \sigma_{ij} &= \left(k^2 n_0^2 m \int \mathbf{e}_1\mathbf{e}_j f(r) U_i^* U_j r\,\mathrm{d}r\right)\cdot\left(2\beta_i \int U_i^* U_i r\,\mathrm{d}r\right)^{-1}, \\ \gamma_{ij} &= \left(\int U_i^2 U_j^2 r\,\mathrm{d}r\right)\cdot\left(\int U_i^2 r\,\mathrm{d}r\right)^{-1}, \end{aligned} \tag{5}$$

где $k = \omega/c$, ω и c — частота и скорость света в вакууме. Уравнения (4) должны решаться совместно с начальными условиями для временных огибающих мод A_j, определяемыми условиями возбуждения световода. Достаточно общий вид начальных условий представляется соотношением $A_2(\tau, 0) = \xi A_1(\tau, 0)$, где параметр ξ определяет тип возбуждения волокна. При $\xi = 1$ имеет место симметричное, а при $\xi = -1$ антисимметричное возбуждение волокна, а при $\xi = 0$ и $\xi^{-1} = 0$ реализуется одномодовое его возбуждение.

Введем характерные длины — дисперсионную $L_\mathrm{d} = \tau_0^2|d|^{-1}$, групповой расстройки $L_\mathrm{g} = \tau_0\upsilon$, межмодового $L_\sigma = \sigma^{-1}$ и нелинейного $L_\mathrm{n} = = |R|^{-1}|A_0|^{-2}$ взаимодействий ($2\upsilon = \upsilon_1 + \upsilon_2$, $2\sigma = \sigma_{12} + \sigma_{21}^*$). В рассматриваемом нами случае $L_\sigma \ll L_\mathrm{d}$, L_g, L_n, поэтому в отсутствие потерь изменение мощности импульса на длине L_σ, обусловленное материальной дисперсией волокна, дисперсией групповых скоростей мод и кубической нелинейностью, пренебрежимо мало. Поэтому можно считать, что на длине L_σ с достаточной степенью точности выполняется условие $|A_1|^2 + |A_2|^2 = \mathrm{const}$, из которого для коэффициентов, определяемых соотношениями (5), следует справедливость соотношений $\sigma_{12} \cong \sigma_{21}^* = \sigma$, $\gamma_{11} \cong \gamma_{22} = \gamma_c$ и $2\gamma_{21} \cong 2\gamma_{12} = \gamma_k$.

Решение уравнений (4) в условиях фазового синхронизма мод будем искать в приближении медленно меняющихся амплитуд. Представим временную огибающую соответствующей моды в виде суммы двух парциальных импульсов:

$$A_j = (-1)^{j+1} a_1(\tau, z)\exp(i\sigma z) + a_2(\tau, z)\exp(-i\sigma z), \tag{6}$$

где a_f — медленно меняющиеся с координатой z амплитуды. Подставляя (6) в (4), получаем следующие уравнения для амплитуд парциальных импульсов ($f=1,2$):

$$\frac{\partial a_f}{\partial z} - \frac{iD_f}{2}\frac{\partial^2 a_f}{\partial \tau^2} + iR(\gamma_k+\gamma_c)(|a_1|^2+|a_2|^2)a_f = 0. \tag{7}$$

Из (6) и (7) следует, что исходный импульс может быть представлен как комбинация парциальных импульсов a_1 и a_2, для которых эффективная дисперсия волокна определяется выражением

$$D_f = (-1)^f d_{\mathrm{m}} + d\,, \tag{8}$$

где введена межмодовая дисперсия групповых скоростей мод $d_{\mathrm{m}} = \dfrac{1}{\upsilon^2\sigma}$. Знак материальной дисперсии d определяется типом реализуемой материальной дисперсии кварцевых волокон на центральной частоте импульса — нормальной или аномальной. При соответствующем подборе параметров υ, σ и d для одного из парциальных импульсов в волокне можно создать волноведущую среду с нулевой эффективной дисперсией, для которой на рабочей частоте выполняется условие $D_f=0$.

2. Зависимость межмодовой дисперсии от параметров световода определим для градиентного волокна. Радиальное распределение показателя преломления примем параболическим, т. е. будем его описывать функцией $f(r)=\Delta(r/r_0)^2$, где r_0 — радиус сердцевины волокна, а для параметра Δ справедливо соотношение $\Delta = (1-n_{\mathrm{ob}}^2/n_0^2)^{1/2}$, где n_{ob} — показатель преломления оболочки волокна. Постоянная распространения соответствующей аксиально—симметричной LP_{0j}-моды в этом случае определяется выражением [10]

$$\beta_j = n_0 k\left[1-\frac{2(2j-1)\Delta}{n_0 k r_0}\right]^{1/2}. \tag{9}$$

Учитывая (9), для межмодовой дисперсии LP_{01} и LP_{02} мод получаем

$$d_{\mathrm{m}} \cong \frac{1}{4\sigma}\left[\left(\frac{\partial\beta_2}{\partial\omega}\right)^{-1} - \left(\frac{\partial\beta_1}{\partial\omega}\right)^{-1}\right]^2 = \frac{4}{\sigma}\left(\frac{\Delta^2 c}{n_0\omega^2 r_0^2}\right)^2. \tag{10}$$

Радиальное распределение поля LP_{0j} моды в этом случае задается в виде

$$U_j = C_j L_{j-1}^{(0)}\left(\frac{2r^2}{w^2}\right)\exp\left(-\frac{r^2}{w^2}\right), \tag{11}$$

где эффективный радиус волокна $w=\left(\dfrac{2r_0}{n_0 k\Delta}\right)^{1/2}$, $L_{j-1}^{(0)}$ — полиномы Лагерра, задаваемые для LP_{01} и LP_{02} мод выражениями $L_0^{(0)}=1$ и $L_1^{(0)}=1-\dfrac{2r^2}{w^2}$. С учетом приведенных соотношений для коэффициента связи указанных мод получаем $\sigma=\dfrac{m\Delta}{2r_0}$. При этом выражение для эффективной дисперсии волокна может быть представлено следующим образом:

$$D_f = d + (-1)^f d_{\mathrm{mo}}\left(\frac{\omega_{02}}{\omega}\right)^4, \quad d_{\mathrm{mo}} = \frac{2\Delta^7 r_0 n_0^2}{t_{11}^4 c^2 m}, \tag{12}$$

где d_{mo} — выражение для модовой дисперсии на частоте отсечки $\omega_{02}=\dfrac{\sqrt{2}t_{11}c}{r_0 n_0\Delta}$ моды LP_{02}, $t_{11}=3{,}832$ — корень функции Бесселя первого рода [11].

Рассмотрим условия возникновения нулевой эффективной дисперсии волокна вблизи частоты отсечки моды LP_{02}. Пусть $n_0 = 1{,}45$, $r_0 \cong \cong 8{,}5 \cdot 10^{-6}$ м и $\Delta \cong 0{,}1$. При этом частота отсечки $\omega_{02} \cong 1{,}3 \cdot 10^{15}$ с$^{-1}$, что соответствует области минимума оптических потерь в кварцевых волокнах. На этой частоте значение материальной дисперсии световода $d \cong -2{,}5 \cdot 10^{-26}$ с^2/м. Равенства нулю эффективной дисперсии D_f в случае симметричного возбуждения световода можно добиться при глубине модуляции показателя преломления $m = 1{,}5 \cdot 10^{-5}$. Меняя параметры волокна и вводимого в него излучения можно менять как модовую дисперсию, так и величину и знак эффективной дисперсии световода в достаточно широком диапазоне частот.

3. Из (6) следует, что при симметричном или антисимметричном двухмодовом возбуждении световода ($A_{20} = \pm A_{10}$) амплитуда одного из парциальных импульсов равна нулю ($a_1 = 0$ при симметричном возбуждении, $a_2 = 0$ — при антисимметричном). При этом эти уравнения вырождаются в одно нелинейное уравнение Шредингера [1, 2]

$$\frac{\partial a_f}{\partial z} - \frac{iD_f}{2} \cdot \frac{\partial^2 a_f}{\partial \tau^2} + iR(\gamma_k + \gamma_c)|a_f|^2 a_f = 0, \tag{13}$$

где $f = 1$ для антисимметричного и $f = 2$ для симметричного возбуждения волокна. Данное уравнение описывает динамику импульса в кубически нелинейной среде с эффективной дисперсией D_f. Характерной чертой описываемого уравнением (13) процесса распространения волнового пакета в такой среде является самовоздействие, приводящее как к уширению волнового пакета, так и сжатию. Одним из важных эффектов, имеющих место в волноведущей среде и описываемых уравнением (13), является формирование устойчивых волновых пакетов — шредингеровских солитонов, возникновение которых связано с балансом действия нелинейности среды и дисперсии [1, 2]. В рассматриваемом нами случае двухмодового периодического волокна его дисперсионные свойства определяются эффективной дисперсией D_f, обусловленной материальной дисперсией и межмодовой связью. Представляя решение (13) в виде $a_f(\tau, z) = |a_f(\tau)| \exp(-i\Gamma z)$, приходим к следующему уравнению для модуля амплитуды парциального импульса:

$$\frac{D_f}{2}\left(\frac{\partial^2 |a_f|}{\partial \tau^2}\right) + \Gamma |a_f| - R(\gamma_k + \gamma_c)|a_f|^3 = 0. \tag{14}$$

Первый интеграл этого уравнения имеет вид

$$D_f \left(\frac{\partial |a_f|}{\partial \tau}\right)^2 + 2\Gamma |a_f|^2 - R(\gamma_k + \gamma_c)|a_f|^4 = C, \tag{15}$$

где стационарному процессу распространения импульса отвечает константа интегрирования $C = 0$. В случае аномальной эффективной дисперсии ($D_f < 0$) и фокусирующих свойств волноведущей среды ($R > 0$) уравнение (15) имеет решение, определяющее так называемые “светлые"солитоны секанс-гиперболической формы. Полное решение уравнения (13) для амплитуды парциального импульса в этом случае имеет вид

$$a_f(\tau, z) = a_{f0} \operatorname{sech}(\tau/\tau_{\text{и}}) \exp(-i\Gamma z), \tag{16}$$

где фаза, длительность импульса (солитона) и его начальная амплитуда связаны соотношениями

$$2\Gamma = R(\gamma_k + \gamma_c) a_{f0}^2 = |D_f|/\tau_{\text{и}}^2, \tag{17}$$

откуда следует, что длительность солитона существенным образом определяется величиной эффективной дисперсии

$$\tau_{\text{и}} = \left(\frac{|D_f|}{R(\gamma_k + \gamma_c) a_{f0}^2}\right)^{1/2}. \tag{18}$$

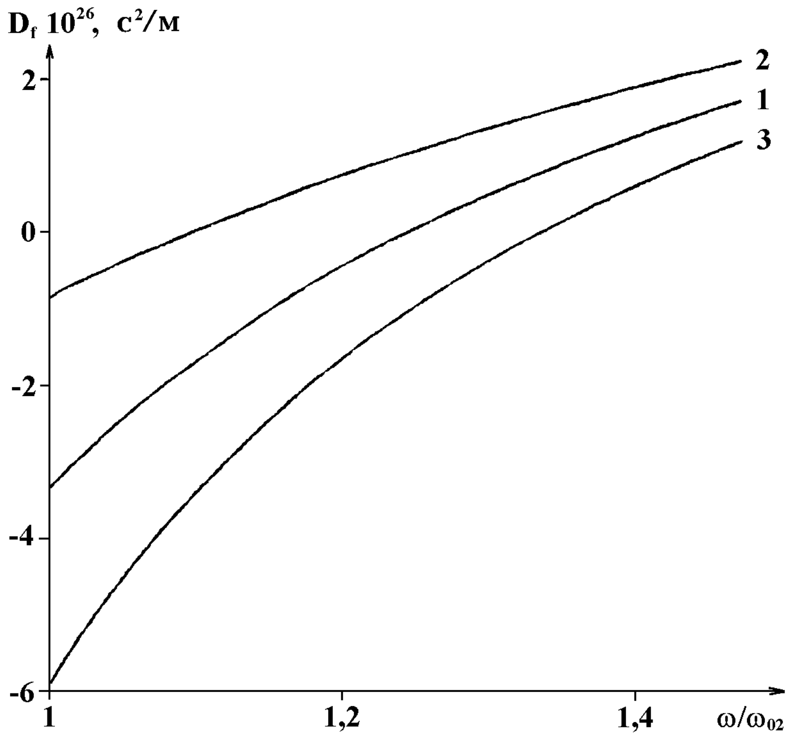


Рис. 1. Зависимость эффективной дисперсии волокна D_f от нормированной частоты ω/ω_{02}.

В волноводе с нормальной эффективной дисперсией ($D_f > 0$) и $R > 0$ уравнение (13) допускает стационарное решение вида

$$a_f(\tau, z) = a_{f0}\,\mathrm{th}(\tau/\tau_{\text{и}})\exp(-i\Gamma z), \tag{19}$$

где фаза, длительность и начальная амплитуда импульса связаны соотношениями $\Gamma = R(\gamma_k+\gamma_c)a_{f0}^2 = D_f/\tau_{\text{и}}^2$. Это решение описывает стационарное состояние, называемое “темным"солитоном, которое представляет собой провал в интенсивности излучения в точках с координатой $z = tu$.

Введем дисперсионную длину $L_D = \tau_{\text{и}}^2/|D_f|$, связанную с эффективной дисперсией волокна. Анализ (13) показывает, что самосжатие импульса имеет место в волокне с $RD_f < 0$ при выполнении условия $2L_{\text{n}} < (\gamma_k + \gamma_c)L_D$. Если $2L_{\text{n}} = (\gamma_k + \gamma_c)L_D$, дисперсионное расплывание импульса компенсируется его самосжатием. При этом плотность энергии солитона

$$W_{\text{s}_f} = 2\tau_{\text{и}}|a_f(0,z)|^2 = \frac{2|D_f|}{R\tau_{\text{и}}(\gamma_k+\gamma_c)}. \tag{20}$$

Если плотность вводимой в волокно энергии W_0 достаточно близка к W_{s_f}, то реализуется солитонный режим распространения импульса. При $W_0 < W_{\text{s}_f}$ импульс расплывается, а при $W_0 > W_{\text{s}_f}$ импульс сжимается. В последнем случае для длины и степени самосжатия импульса справедливы аппроксимационные соотношения $z_k\chi^2 \cong (0{,}32\chi + 1{,}1)L_D,\ \tau_k \cong \dfrac{\tau_0}{4{,}1\chi}$, где $\chi^2 = \dfrac{(\gamma_k+\gamma_c)L_D}{2L_{\text{n}}}$ [12].

4. На рис. 1 приведена частотная зависимость эффективной дисперсии D_f для симметричного (кривая 2) и антисимметричного (кривая 3) двухмодового возбуждения волокна с параметрами $\tau_0 = 10^{-11}$ с, $n = 1{,}45$, $n_2 = 2\cdot10^{-20}$ м^2/Вт, $d_{\text{m}0} = 2{,}5\cdot10^{-26}$ с^2/м, $\omega_{02} = 1{,}2\cdot10^{15}$ с$^{-1}$. Кривая 1 представляет собой частотную зависимость материальной дисперсии световода, построенной на основе соотношения $d(\omega) \cong k(\mathrm{d}^2 n/\mathrm{d}\omega^2)$, и отвечает эффективной дисперсии одномодового волокна. Показатель преломления материала волокна в рассматриваемом интервале частот $\omega \cong (1 \div 2) \cdot 10^{15}$ с$^{-1}$, лежащем вдали от резонансных частот кварца, достаточно хорошо описывается выражением [13]

$$n^2(\omega) = 1 + \sum \frac{B_i \omega_i^2}{\omega_i^2 - \omega^2}, \quad i = 1,\, 2,\, 3. \tag{21}$$

Здесь члены с коэффициентами $B_1 = 0{,}696$ и $B_2 = 0{,}407$ связаны с резонансами в УФ области (соответствующие резонансные частоты $\omega_1 \cong 2{,}63 \cdot 10^{16}$ с$^{-1}$, $\omega_2 \cong 1{,}55 \cdot 10^{16}$ с$^{-1}$, а с коэффициентом $B_3 = 0{,}897$ ($\omega_3 \cong 1{,}82 \cdot 10^{14}$ с$^{-1}$) — в инфракрасной области. Как следует из представленных кривых, в исследуемом диапазоне частот зависимость $D_f(\omega)$ носит монотонный характер и в точках дисперсионного нуля ω_{d_i} имеет место смена знака эффективной дисперсии для указанных типов возбуждения волокна. Симметричное двухмодовое возбуждение волокна смещает эту точку в ИК область, антисимметричное возбуждение — в УФ область.

На рис. 2 для световода с используемыми выше параметрами приведены частотные зависимости критической энергии W_{s_f} образования солитона (а) и коэффициента сжатия импульса $\tau_0/\tau_{\text{и}}$ (б) при значении входной амплитуды импульса $A_0 = 2 \cdot 10^5$ Вт/м^2. Видно, что область возможного существования солитонного решения в случае антисимметричного возбуждения двухмодового волокна (кривая 3) значительно больше, чем в случае одномодового волокна (кривая 1). Для симметричного возбуждения волокна (кривая 2) минимум плотности энергии образования солитона смещается в область дальнего ИК диапазона, что позволяет реализовать солитонный режим в области минимума оптических потерь.

На рис. 3 приведена зависимость амплитуды “светлого"солитона от нормированного времени $\tau/\tau_{\text{и}}$ (в бегущей системе координат) для введенных ранее параметров волокна и двух значений несущей частоты $\omega = 1{,}2 \cdot 10^{15}$ с-1 (сплошные линии) и $\omega = 1{,}3 \cdot 10^{15}$ с$^{-1}$ (пунктирные линии). Нумерация кривых соответствует нумерации предыдущих рисунков. Видно, что с ростом частоты амплитуда солитона уменьшается при всех значениях τ вплоть до точки дисперсионного нуля ω_{d_i}.

5. Анализ дисперсионных свойств световода в случае его произвольного начального возбуждения, определяемого параметром ξ, показывает, что выражение для эффективной дисперсии световода в этом случае принимает вид

$$D = d + \frac{2\xi}{\sigma v^2 (1 + \xi^2)}\,. \tag{22}$$

Выражения для длительности импульса и энергии образования солитона в этом случае, имеют вид, подобный выражениям (18) и (20) с учетом эффективной дисперсии, определяемой соотношением (22). Следует отметить, что в случае, если выполняется условие $|d| v^2 \sigma \le 1$, всегда можно подобрать $\xi = \dfrac{-1 \pm \sqrt{1 - (dv^2\sigma)^2}}{dv^2\sigma}$, при котором эффективная дисперсия световода становится равной нулю. В случае одномодового возбуждения волокна (ξ, $\xi^{-1} = 0$) дисперсия импульса фактически полностью определяется материальной дисперсией световода и не зависит от межмодовой дисперсии, т. е. $d_{\mathrm{m}} = 0$, $\quad D = d$. Таким образом, кривые 1 на приведенных выше рисунках можно считать относящимися к случаю одномодового возбуждения двухмодового волокна.

Наличие широкого набора параметров, управляющих дисперсионными свойствами двухмодового периодического волокна, делает удобным использование подобных волокон в качестве модельного объекта для исследования различного рода нелинейных эффектов. Кроме того, в отличие от одномодового волокна, для которого условие минимума материальной дисперсии выполняется только на одной частоте, в двухмодовых периодических волокнах выбором типа возбуждения можно получать дисперсионные минимумы на различных частотах. Это обстоятельство делает весьма перспективным прак-

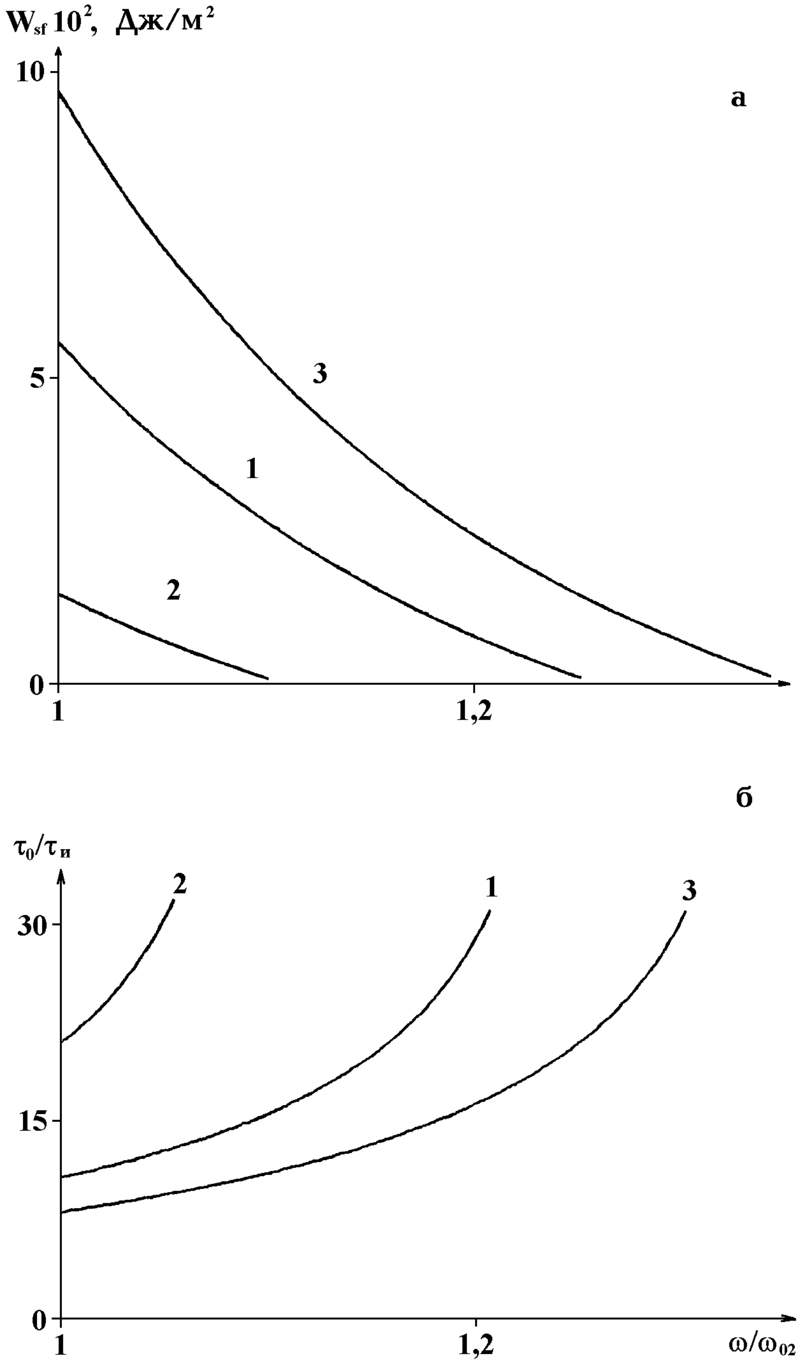


Рис. 2. Зависимость эффективной энергии образования солитона W_{sf} (а) и степени сжатия импульса $\tau_0/\tau_и$ (б) от нормированной частоты ω/ω_{02}.

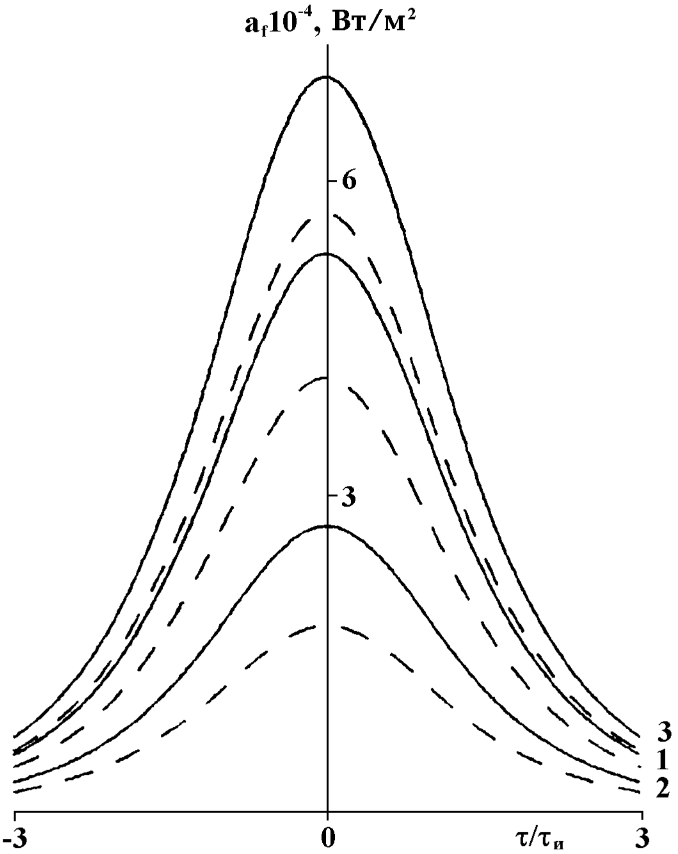


Рис. 3. Зависимость амплитуды солитона от нормированной длительности импульса $\tau/\tau_и$.

тическое применение двухмодовых волокон, позволяя создавать в световодах дополнительные “окна"с оптимальными дисперсионными параметрами.

ЛИТЕРАТУРА

1. Ахманов С. А., Выслоух В. А., Чиркин А. С. Оптика фемтосекундных лазерных импульсов. — М., 1988. – 312 с.
2. Agrawal G. P. Nonlinear fiber optics. — Boston etc: Academic Press, 1989. P. 392.
3. Прохоров А. М., Смоленский Г. А., Агеев А. М. //УФН, 1984. Т. 143. № 1. С. 33.
4. Семенцов Д. И., Широков А. А., Шутый А. М. //Радиотех. и электрон., 1994. Т. 39. № 10. С. 1524.
5. Hill K. O. et. al. //Appl. Phis. Lett., 1978. V. 32. P. 647.
6. Kawasaki B. S. et. al. //Opt. Lett., 1978. V. 3. P. 66.
7. Winful H. G. //Appl. Phis. Lett., 1985. V. 46. P. 527.
8. Выслоух В. А., Геворкян Л. П. //Изв. АН СССР, 1991. Т. 55. № 2. С. 323.
9. Золотовский И. О., Семенцов Д. И. //Опт. и спектр., 1998. Т. 84. № 1. С. 77.
10. Унгер Х. Г. Планарные и волоконные оптические волноводы. — М., 1980. – 656 с.
11. Янке Е., Эмде Ф., Леш Ф. Специальные функции. — М., 1977. – 342 с.
12. Дианов Е. М., Никонова З. С., Прохоров А. М., Серкин В. Н. //Письма в ЖТФ, 1986. Т.12. С.756.

13. Maltson I. H. //J. Opt. Soc. America, 1965. V. 55. P. 1205.

Государственный университет,
г. Ульяновск, Россия

Поступила в редакцию
30 января 1998 г.

OPTICAL PULSE PROPAGATION IN PERIODICAL NONLINEAR FIBERS

I. O. Zolotovskij, D. I. Sementsov

The conditions are studied of soliton pulse formation and propagation in periodical two-mode nonlinear fiber with regard to fiber material dispersion and intermode dispersion of interacting modes provided their phase synchronism.

УДК 612.766

МЕТОДЫ СПЛАЙН-АППРОКСИМАЦИИ В ЗАДАЧЕ АМПЛИТУДНО–ВРЕМЕННОГО АНАЛИЗА ПУЛЬСОВОЙ ВОЛНЫ

В. В. Бороноев, О. С. Ринчинов

Проведен анализ применимости методов восстановления производных сигнала кубическими и B-сплайнами в задаче амплитудно–временного анализа пульсовой волны. Установлено, что интерполирующие сплайны достаточно чувствительны к наличию в обрабатываемом сигнале шумов. Предложена и реализована схема предварительной обработки, основанная на применении специально сконструированного фильтра низких частот (ФНЧ) и позволяющая существенно повысить качество методов. Определены предельные уровни ошибки нахождения ординат характерных точек пульсовой волны, выявлены их зависимости как от уровня шума, так и от амплитуды исходного сигнала. Полученные результаты позволили скорректировать требования по качеству регистрирующего тракта автоматизированного пульсодиагностического комплекса (АПДК). На основе исследованных методов разработан прикладной пакет амплитудно–временного анализа пульсовой волны как составная часть АПДК.

1. ПОСТАНОВКА ЗАДАЧИ

Решение задачи автоматизации медицинской диагностики подразумевает наличие надежных, устойчивых методов выделения и анализа диагностически значимых параметров физиологических сигналов. В качестве таковых можно рассматривать их амплитудные, временные и амплитудно–временные характеристики [4]. Такой анализ дает достаточный количественный материал для определения критериев постановки диагноза, поскольку протекание любого физиологического процесса, регистрируемого в виде сигнала, напрямую связано с теми или иными характерными свойствами этого сигнала (амплитудой, фазой, геометрией контура). Анализируя изменения внешних характеристик сигнала, можно устанавливать их связи с соответствующими внутренними изменениями в протекании порождающих их процессов, определяя некоторые устойчивые области нормального или патологического развития в пространстве анализируемых параметров.

Амплитудные и временные характеристики сигнала однозначно определяются набором некоторых характерных точек, являющихся экстремумами и точками перегиба кривой сигнала. Задача корректного определения характерных точек облегчается свойством производной функции, согласно которому производная меняет знак в точке экстремума функции и направление в точке перегиба. Следовательно, существуют необходимые предпосылки для создания процедуры автоматизированного определения характерных точек, основанную на вычислении и анализе первой и второй производных сигнала.

Среди методов восстановления производных сигнала выделяются основанные на применении аппроксимирующих свойств сплайн-функций. Они отличаются простотой реализации, нетребовательностью к вычислительным ресурсам, устойчивостью получаемых решений. Нашей целью при проведении настоящих исследований являлась оценка возможности таких методов в задаче восстановления производных пульсового сигнала.

2. ЛОКАЛЬНЫЕ B-СПЛАЙНЫ ВТОРОЙ СТЕПЕНИ

В качестве основных для восстановления производных пульсового сигнала использовались методы, основанные на применении интерполирующих сплайнов разных порядков; были рассмотрены кубические сплайны и локальные B-сплайны. Постановка задачи, решение и вычислительные процедуры для кубических сплайнов подробно изложены в [2]. Ниже рассмотрены некоторые вопросы практической реализации метода локальных В-сплайнов.

Пусть $x(t)$ — непрерывный аналог экспериментального пульсового сигнала, рассматриваемого после дискретизации как сеточная функция $x_k = x(t_k),\ k = 0, 1, \ldots, N-1$; N —длина реализации в отсчетах.

Интерполяционную формулу функции $x(t)$ можно записать следующим образом:

$$x(t) = \sum_{k=0}^{N-1} b_k\, B(t-k), \tag{1}$$

где $B(t)$ — локальные В-сплайны вида

$$B(t) = \frac{1}{16}\left\{(t+3/2)_+^2 - 3(t+1/2)_+^2 + 3(t-1/2)_+^2 - (t-3/2)_+^2\right\}. \tag{2}$$

Символ “+"в (2) означает, что степень равна нулю при отрицательных значениях аргумента. Свойства введенной сплайн-функции $B(t)$:

- имеет финитный носитель наименьшей длины — $[-3/2,\ 3/2]$;
- неотрицательна на нем;
- непрерывна (имеет непрерывную первую производную);
- вторая производная является кусочно—постоянной функцией.

Предполагается, что коэффициенты b_k являются линейными комбинациями ближайших значений временного ряда x_k:

$$b_k = c_{-1}x_{k-1} + c_0x_k + c_1x_{k+1}\,. \tag{3}$$

Коэффициенты $c_j,\ j = -1, 0, 1$, ввиду локальности сплайна, не привязаны к узлам сетки; исходя из требований минимизации погрешности приближения функции сплайном второй степени, для них были приняты следующие значения:

$$c_{-1} = c_1 = -1, \quad c_0 = 10. \tag{4}$$

Пусть $t = p + \xi$, $\xi \in [-1/2,\ 1/2]$, $p = 0,\ 1, \ldots,\ N-1$. В этом случае функция $x(t)$ на p-ом участке разбиения представима как

$$x(p+\xi) = x_p(\xi) = b_{p-1}\, B(1+\xi) + b_p\, B(\xi) + b_{p+1}\, B(\xi-1). \tag{5}$$

Используя формулы (5) и (1), получаем следующее выражение для функции $x(t)$ на p-ом интервале сетки:

$$x_p(\xi) = a_0 + a_1\xi + a_2\xi^2. \tag{6}$$

Коэффициенты многочлена (6) имеют вид

$$\begin{aligned} a_0 &= \frac{1}{64}(b_{p-1} + 6b_p + b_{p+1}),\\ a_1 &= \frac{1}{16}(b_{p+1} - b_{p-1}),\\ a_2 &= \frac{1}{16}(b_{p+1} - 2b_p + b_{p-1}). \end{aligned} \tag{7}$$

Формулы (3) и (7) представляют собой основные формулы вычислительной процедуры для моделирования функции $x(t)$ B-сплайнами. Значения коэффициентов a_1, a_2 многочлена (6) являются соответственно оценками первой и второй производной в узлах сеточной функции x_k, $k = 0, 1, \ldots, N-1$.

Для корректного вычисления коэффициентов интерполяционной формулы на всей области определения, формулы (3), (7) необходимо дополнить выражениями для вычисления значений функции за пределами рассматриваемого интервала: x_{-1} и x_N. В нашем случае были использованы простейшие линейные зависимости вида

$$x_{-1} = 2x_0 - x_1, \quad x_N = 2x_{N-1} - x_{N-2}\,. \tag{8}$$

Реализованную описанным образом процедуру нахождения сплайн-функции второго порядка отличает простота программной реализации и высокое быстродействие, что наилучшим образом отвечает требованиям, предъявляемым к обработке длительных реализаций пульсового сигнала.

3. ИССЛЕДОВАНИЯ ПРИМЕНИМОСТИ СПЛАЙН-ИНТЕРПОЛЯЦИИ ДЛЯ ВОССТАНОВЛЕНИЯ ПРОИЗВОДНЫХ ПУЛЬСОВОГО СИГНАЛА

3.1. Процедура регистрации пульсовых сигналов

При проведении эксперимента, в его постановочной части, нами были предприняты меры для того, чтобы он в наибольшей степени соответствовал каноническим требованиям к процедуре диагностики по пульсу. Эти правила касаются предпочтительно времени съема сигналов, продолжительности самой процедуры, положения тела пациента при проведении исследования и пр.

Для получения исходных реализаций пульсовых сигналов в цифровой форме использовался автоматизированный пульсодиагностический комплекс (АПДК), разрабатываемый в лаборатории радиобиофизики БИЕН СО РАН. Аппаратная часть комплекса спроектирована таким образом, чтобы максимально удовлетворять требованиям пульсовой диагностики тибетской медицины. Съем пульсовых сигналов производится двумя трехпозиционными датчиками, связанными цифровым трактом с компьютером, с обеих рук пациента одновременно. В нашем эксперименте частота выборки исходных данных составляет 200 Гц для каждого регистрируемого сигнала. Таким образом, основной особенностью используемой методики является одновременный съем пульсовых сигналов в шести точках обоих запястий обследуемого.

Полученный сигнал является суперпозицией исходной пульсовой волны (несущего сигнала, являющегося, по сути, квазипериодическим и при сильном упрощении рассматриваемого как композиция, в основном, низкочастотных гармоник) и шумовых гармонических компонентов (аппаратные наводки, шумы квантования и др.), сумму которых приближенно можно рассматривать как некоторый аддитивный шум с нулевым средним и нормальным законом распределения (гауссов шум). Последний вносит определенные искажения в восстановление производных, что затрудняет точное определение начала временных фаз процессов цикла сердечной деятельности.

3.2. Количественная оценка качества восстановления производных

Первой задачей являлась оценка качества восстановления производных сигнала сплайнами в зависимости от уровня шумов. Для этого был сформирован модельный сигнал, форма и временные характеристики которого задавались набором наиболее характерных экстремумов реального пульсового сигнала. Полученная функция, заданная на неравномерной сетке, интерполировалась кубическими сплайнами, и на их основе формировался временной ряд с равномерным шагом, соответствующим шагу квантования, принятом в нашем эксперименте (5 мс). Такая процедура формирования модельного сигнала гарантирует кусочную монотонность (гладкость) производных, вытекающую из свойств сплайн–функций. Фрагменты модельного сигнала, его первой и второй производной приведены на рис. 1 (а–в), кривые 1.

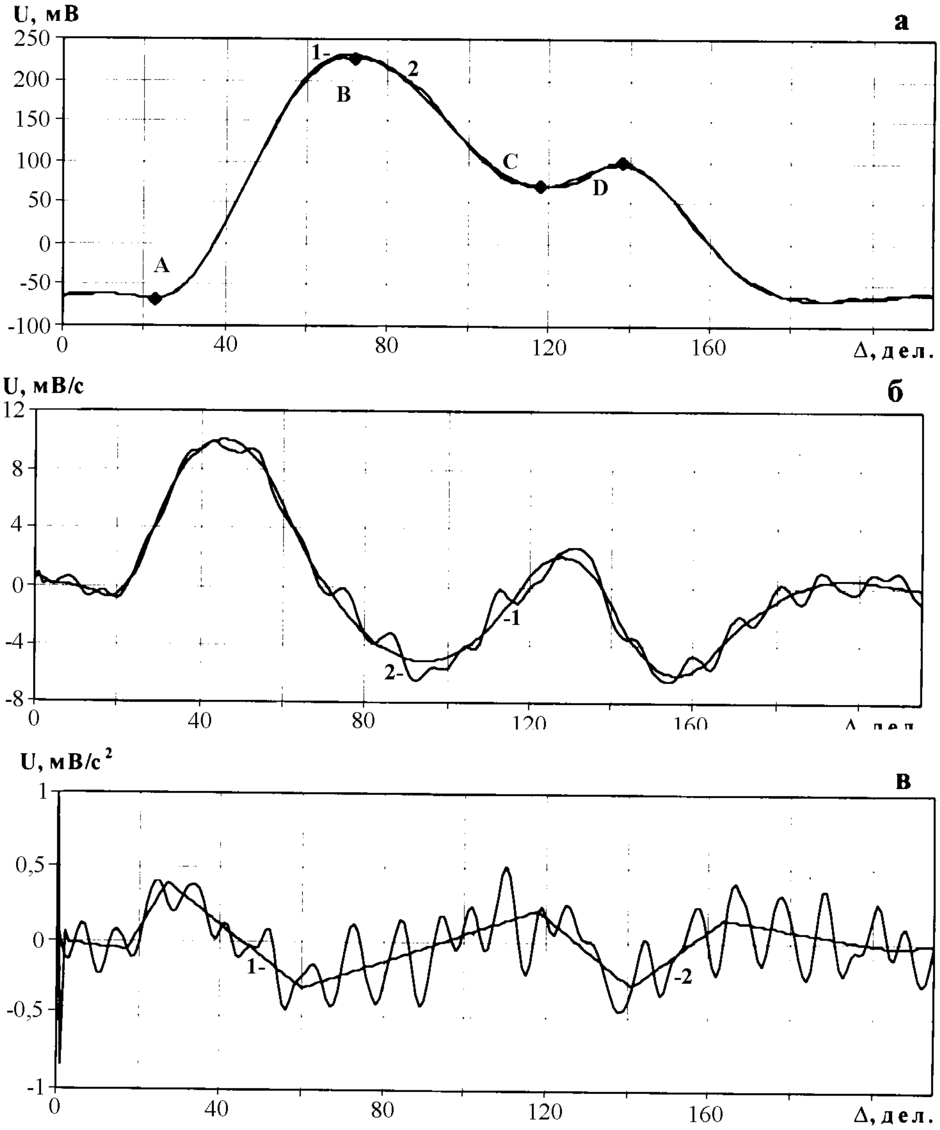


Рис. 1. а) Фрагмент модельных пульсовых сигналов, б) фрагмент первых производных модельных пульсовых сигналов, в) фрагмент вторых производных тех же сигналов. 1 — 0% шума, 2 — 5% шума, после фильтрации.

Характер второй производной, имеющей вид ломаной линии, объясняется тем, что исходный модельный сигнал построен на основе многочленов третьего порядка (кубических сплайнов). В ходе эксперимента модельный сигнал суммировался с гауссовым шумом, получаемым с помощью генератора случайных чисел; его максимальная амплитуда δ задавалась в процентах от максимальной амплитуды модельной пульсовой кривой.

Коэффициенты сплайнов вычислялись для каждого отсчета модельного (M) и зашумленного сигналов (S). Затем для полученных наборов данных определялись соответствующие коэффициенты корреляции ρ:

$$\rho_{\mathrm{MS}} = \frac{R(\vec{M},\vec{S})}{\sqrt{R(\vec{M},\vec{M})}\cdot\sqrt{R(\vec{S},\vec{S})}}\,, \tag{9}$$

где $R(\vec{M},\vec{S}) = \langle f(\vec{M})\cdot f(\vec{S})\rangle - \langle f(\vec{M})\rangle\cdot\langle f(\vec{S})\rangle$ — ненормированная взаимная пространственная корреляционная функция, функции $f(\vec{M})$ и $f(\vec{S})$ соответствуют модельному и зашумленному сигналам, соответственно, а

$$R(\vec{M},\vec{M}) = \langle f^2(\vec{M})\rangle - \langle f(\vec{M})\rangle^2$$

и

$$R(\vec{S},\vec{S}) = \langle f^2(\vec{S})\rangle - \langle f(\vec{S})\rangle^2 \ —$$

дисперсии функций $f(\vec{M})$ и $f(\vec{S})$. Аналогично определялись коэффициенты корреляции и для восстанавливаемых производных модельного и зашумленного сигналов (M', M'' и S', S'', соответственно).

Также по формуле (10) производилась оценка ошибки восстановления второй производной ε:

$$\varepsilon = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i'' - M_i'')^2}. \tag{10}$$

Полученные результаты приведены на рис. 2. На графиках а—в приведены коэффициенты корреляции соответственно модельного и зашумленного сигналов, их первых и вторых производных; на графике г показаны результаты расчетов ошибки восстановления второй производной ε; все данные приведены в зависимости от уровня шума. Показатели кубических сплайнов обозначены как кривые 1, B-сплайнов — кривые 2.

Было отмечено, что величина ошибки ε находится в прямой зависимости от величины стандартного отклонения σ соответствующей функции, возрастающего с увеличением амплитуды шумов.

Лучшие показатели B-сплайнов относительно кубических (наблюдаемые на рис. 2) объясняются, очевидно, их локальным характером. Поскольку в вычислении коэффициентов B-сплайна участвуют только три смежных узла сетки, на них мало сказывается суммарная ошибка задания сеточной функции на всем интервале ее определения. В то же время, качество второй производной, вычисленной непосредственно по коэффициентам интерполяционных сплайнов сигнала в случае кубических сплайнов и путем повторного определения коэффициентов сплайна для отсчетов первой производной (коэффициентов при первой степени многочлена (6)) в случае B-сплайнов, совершенно неудовлетворительно даже при весьма низких уровнях шума в обоих случаях. Это было подтверждено и при последующем анализе производных реального пульсового сигнала. Очевидно, что для получения более качественных производных, исходный сигнал необходимо подвергнуть предварительной обработке.

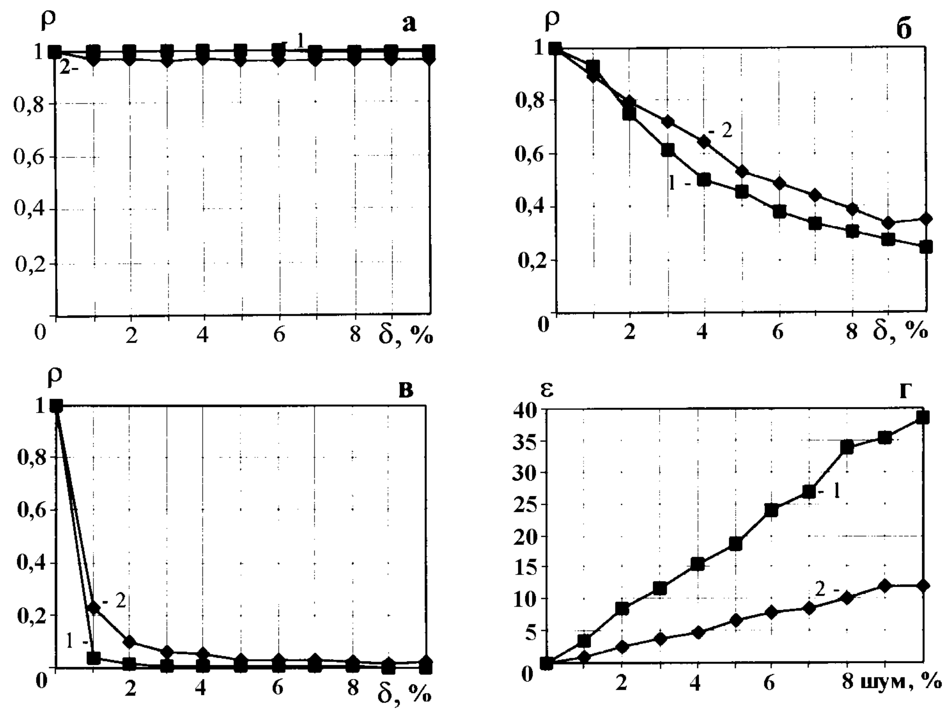


Рис. 2.

3.3. Методы повышения эффективности восстановления производных сигнала

3.3.1. Сглаживание

В теории сплайн—функций накоплен значительный опыт по применению свойств интерполирующих сплайнов для решения задачи сглаживания функции, заданной на сетке. Однако они малопригодны в нашем случае для анализа пульсовой кривой, поскольку, сводя решение задачи сглаживания исходного сигнала к решению задачи минимизации функционала вида [2]

$$J(S) = \int_a^b |S^{(q)}|^2 dx, \quad |S(x_i) - z_i^{(0)}| \le \varepsilon_i, \quad i = 0, \ldots, N, \tag{11}$$

требуют многократного (с заранее не определимой величиной цикла) выполнения процедуры определения сплайн-коэффициентов, а значит, и значительных ресурсов (вычислительных и временных). Кроме того, для реализации этих методов необходимо обладание некоторой априорной информацией, в частности, погрешностью задания исходной функции на каждом шаге $z_i^{(0)}$. В то же время трудно оценить амплитудные и фазовые искажения, вносимые подобной процедурой сглаживания в исходный сигнал.

Были исследованы на применимость наиболее распространенные методы сглаживания реального времени, основанные на использовании скользящего окна, часто применяемые для некоторых видов анализа пульса, в частности, для выделения и удаления низкочастотной волны (дыхательной компоненты). Оценка проводилась по двум показателям: коэффициенту корреляции модельного сигнала и зашумленного, обработанного скользящим окном, и смещению точек максимума сглаженного сигнала (соответствует точке В на рис. 1) и его производных относительно аналогичных точек модельных сигналов. Последний показатель является оценкой ошибки определения ординаты характерной точки. Оптимальной считалась ширина окна, которая при достаточно высоком уровне корреляции обеспечивала бы наименьшее смещение ординат максимумов. Результаты этого анализа для уровня шумов в 1% приведены в табл. 1, 2.

Таблица 1

Коэффициенты корреляции модельного сигнала и сглаженного методом скользящего среднего при уровне шума в 1% от амплитуды модельного сигнала (производные определялись методом кубических сплайнов)

Ширина окна, мс	ρ_{MS}	$\rho_{M'S'}$	$\rho_{M''S''}$
50	0,99	0,99	0,663
100	0,98	0,93	0,681
150	0,95	0,80	0,448

По приведенным результатам видно, что метод скользящего окна не обеспечивает приемлемого уровня точности для задач амплитудно—временного анализа. При малых размерах окна недостаточной является степень сглаживания, что выражается в низких значениях коэффициента

Таблица 2

Смещение ординаты первого максимума сглаженного сигнала и его производных относительно модельного в миллисекундах (отрицательный знак — смещение в сторону начала координат)

Ширина окна, мс	Δ	Δ'	Δ''
50	0	0	10
100	0	−30	−15
150	−15	−55	−65

корреляции, особенно для второй производной, при больших — существенным образом изменяются фазовые характеристики сигнала, при этом теряется информация о его реальных временных параметрах. При использовании метода скользящего окна может возникнуть значительная ошибка определения времени начала того или иного процесса, существенным образом искажающая результаты последующего анализа. Неудовлетворительными для автоматизированного определения временных параметров пульса являются формы получаемых производных, в которых присутствует множество дополнительных малоамплитудных пиков, затрудняющих поиск экстремумов реальной пульсовой волны. В связи с этим встает задача выбора метода сглаживания, во-первых, минимизирующего фазовые искажения сигнала, во-вторых, нетребовательного к вычислительным ресурсам, который можно было бы реализовать в нашем случае анализа квазиреального времени.

3.3.2. Фильтрация

Поскольку сглаживание является, по существу, процессом низкочастотной фильтрации, то становится оправданным применение цифрового фильтра низких частот (ФНЧ). Алгоритм быстрого преобразования Фурье позволяет, не занимая значительных вычислительных ресурсов, получить представление исходного сигнала в частотной области. Частотные отсчеты сигнала перемножаются с отсчетами ФНЧ соответствующего размера, спроектированного с учетом некоторых априорных представлений о пульсовом сигнале, при этом ФНЧ, обладая одинаковыми вещественной и мнимой частями, имеет линейную ФЧХ и заведомо не вносит фазовых искажений в восстанавливаемый после фильтрации сигнал. Нами был выбран фильтр с частотой среза АЧХ (на уровне 0,7) $f_c \approx 28$ Гц. Для получения гладкой огибающей была использована функция [5]

$$F_i = I_0 \left[\pi a \sqrt{\left(1 - \frac{i - N/2}{N/2}\right)^2}\right] \bigg/ I_0(\pi a)\,, \tag{12}$$

где $I_0(x)$ — цилиндрическая функция Бесселя первого рода нулевого порядка [3], $a = \mathrm{const}$ — параметр, задающий крутизну спада огибающей.

Недостатком данного подхода является низкое быстродействие при обработке длинных реализаций, поскольку метод предполагает последовательное применение прямого и обратного преобразования Фурье. Поэтому, в целях повышения быстродействия до приемлемого уровня, исходная реализация произвольной длины разбивалась на окна шириной в 1024 отсчета.

Результаты расчетов количественных показателей качества восстановления производных модельных сигналов после их обработки ФНЧ приведены на рис. 3. Порядок графиков, обозначения соответствуют рис. 2.

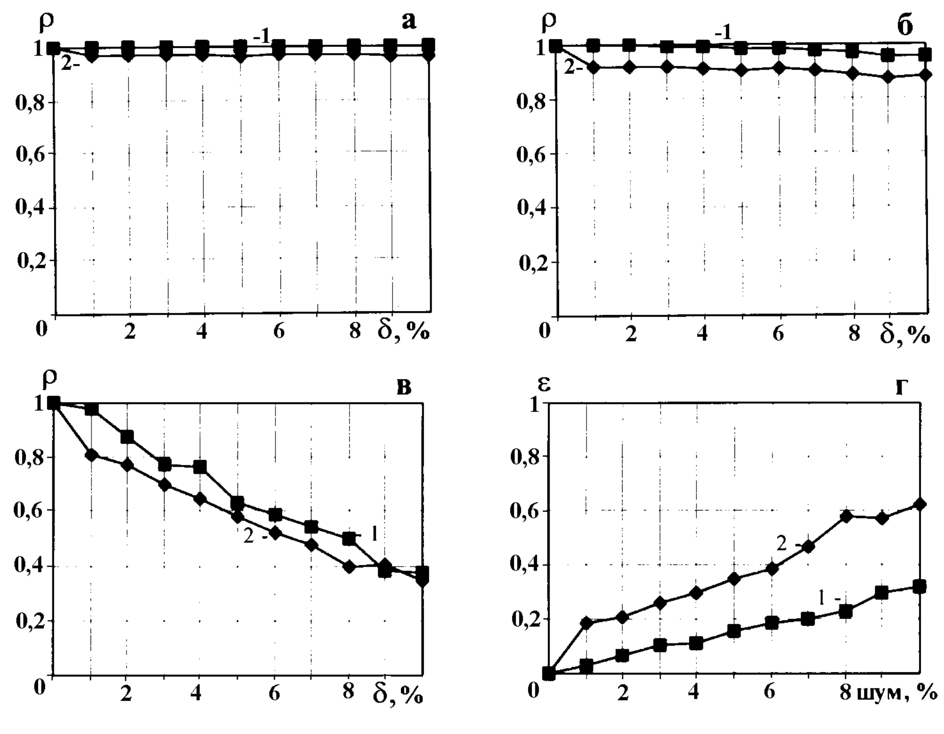


Рис. 3.

Из приведенных графиков видно, что качество восстановления производных сигнала значительно улучшилось, величина ошибки восстановления уменьшилась на два порядка (для кубических сплайнов). Относительно худшие результаты B-сплайнов, показанные в этом эксперименте, также имеют причиной их локальный характер. Кубические сплайны, коэффициенты которых вычисляются как функция по всей области определения исходной сеточной функции (длине реализации сигнала в нашем случае), обладают лучшими сглаживающими свойствами.

3.4. Выделение характерных точек

Проверка корректности выделения характерных точек проверялась в ходе следующего эксперимента. Для модельного (гладкого) сигнала вычислялись производные. По первой производной определялись комплексы характерных точек (A—D на рис. 1), соответствующих наиболее выраженным частям контура пульсограммы. Аналогичные характерные точки находились для сигналов с аддитивным шумом с максимальной амплитудой δ, предварительно подвергнутых обработке описанным выше ФНЧ. Определялись абсолютные значения смещения ординат Δ характерных точек (A—D) зашумленного сигнала относительно модельного. Для большей достоверности для каждого уровня шума проводилось несколько независимых измерений, по результатам которых определялось среднее значение и стандартное отклонение. Результирующие усредненные оценки приведены на рис. 4, 5.

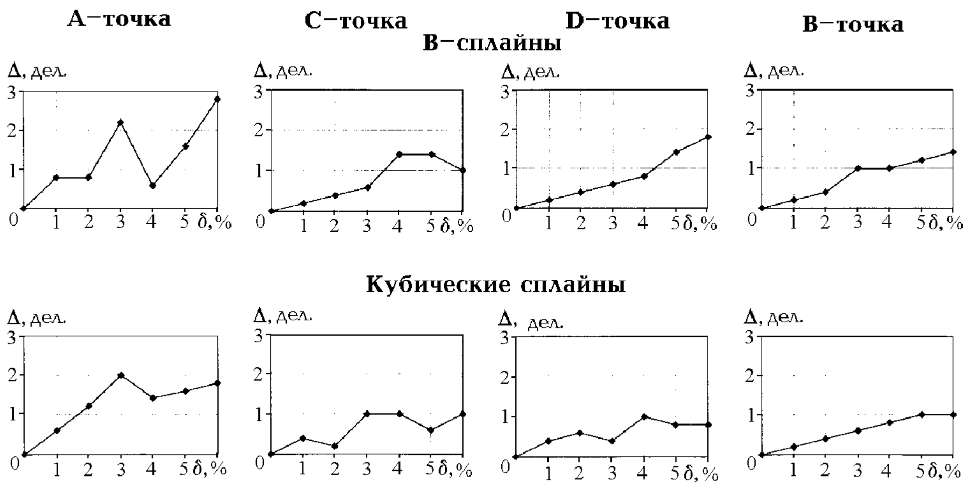


Рис. 4.

Графики на рис. 4, показывающие зависимости величины ошибки от уровня шума, ранжированы по возрастанию амплитуд определяемых характерных точек. Можно отметить, что для небольших амплитуд исходного сигнала ошибка носит в большей степени случайный характер, мало согласующийся с уровнем шума. С возрастанием амплитуды исходного сигнала распределение ошибки становится более упорядоченным, возрастает степень коррелированности с уровнем шума. Правило “больший шум — большая ошибка"становится актуальным для точки с максимальной амплитудой, хотя эта зависимость не имеет линейного характера. Распределения ошибки для кубических и B-сплайнов имеют общие тенденции и, в целом, достаточно близки друг другу.

Значения ошибок невелики при всех исследованных уровнях шума, что говорит о хорошей устойчивости предлагаемой схемы обработки пульсового сигнала, вносящей минимальные искажения в его фазовые характеристики; максимальная ошибка при максимальном исследованном уровне шума достигает 3 шагов квантования или 15 мс. Возникающие ошибки, в основном, связаны с тем, что шумовые компоненты, лежащие ниже частоты пропускания ФНЧ, присутствуют в качестве аддитивных гармонических функций в отфильтрованном сигнале (что можно отчетливо наблюдать на рис. 1б и 1в) и вызывают локальные изменения фазы сигнала.

На рис. 5 приведены графики, показывающие распределение средних ошибок восстановления характерных точек в зависимости от амплитуды последних. Точки ранжированы по возрастанию амплитуды (A, C, D, B). Здесь также можно отметить, что ошибка восстановления в среднем меньше для точек с большей амплитудой. В то же время, особенно для локальных B-сплайнов, видно, что данное утверждение действительно только при достаточно низких уровнях шума — до 3%; при более

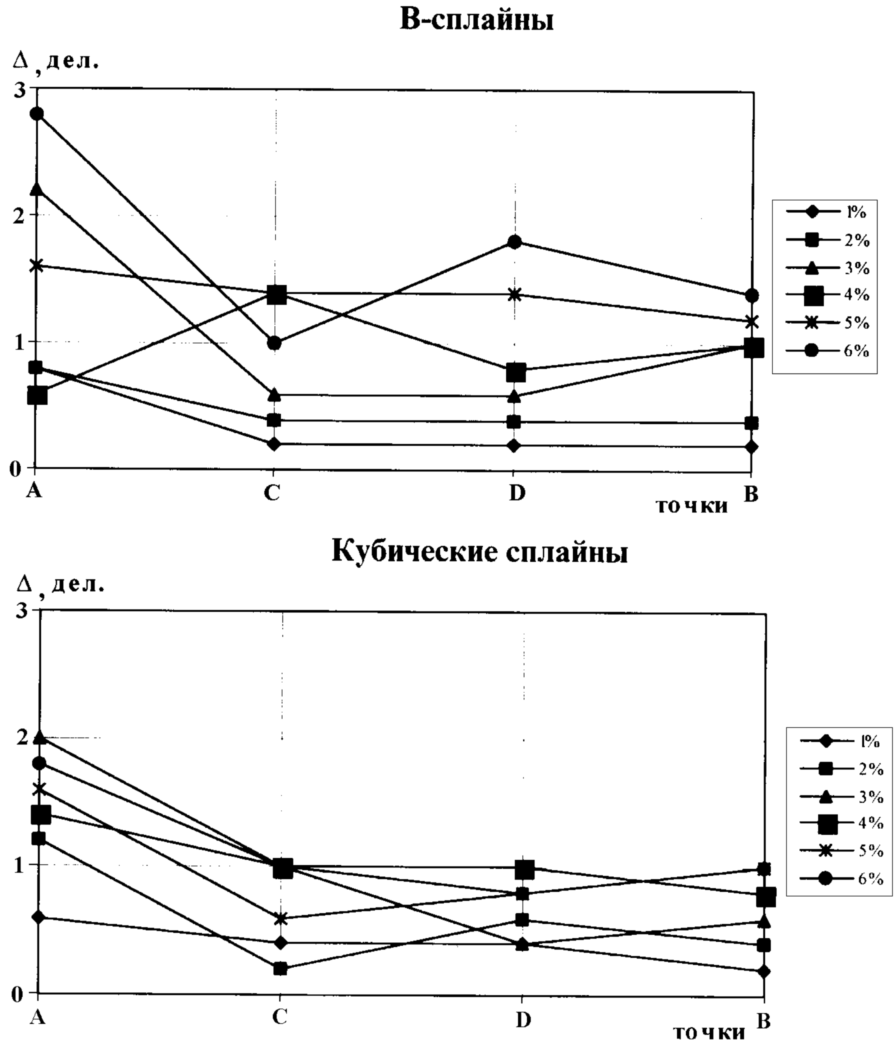


Рис. 5.

высоких шумах (больших амплитудах случайных компонентов) ошибка определения ординаты характерной точки приобретает все более стохастический характер, что затрудняет введение корректирующих поправок и, тем самым, последующий достоверный анализ. Для кубических сплайнов картина распределения ошибки выглядит более однородной, ее средние величины также несколько ниже. Эти результаты хорошо согласуются с рис. 3. После выполнения приведенной выше схемы обработки сигнала, кубические сплайны показывают лучшие результаты по сравнению с B-сплайнами, что еще раз позволяет утверждать об их большем сглаживающем эффекте.

Интересным является сопоставление полученных для сплайнов результатов с аналогичными для метода восстановления производных с помощью регуляризующего алгоритма Тихонова (в котором для восстановления производной решается интегральное уравнение), приведенными на рис. 6, на котором показаны зависимости величины смещения ординаты характерной точки в зависимости от уровня шумов; кривая 1 соответствует первой производной, кривая 2 — второй.

Как видно из графиков, метод Тихонова обеспечивает высокую точность определения характерных точек даже при очень больших уровнях шумов. Лучшие результаты для второй производной на рис. 6 объясняются тем, что для вычисления производных в этом методе используются интегральные уравнения с разными ядрами [1]. Однако, применение данного метода для анализа длинных реализаций

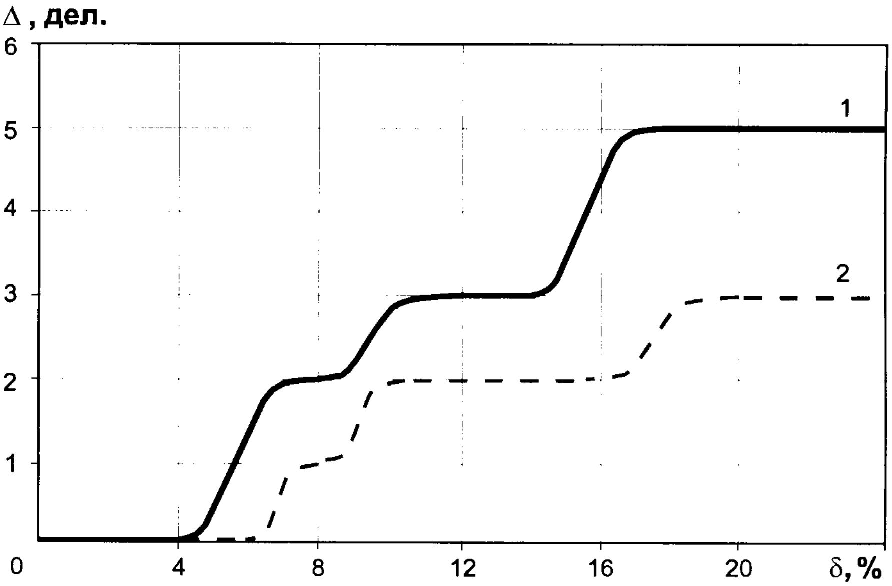


Рис. 6. Зависимость отклонения Δ от уровня шума входного сигнала δ для первой (1) и второй (2) производных [1].

пульсовых сигналов представляется затруднительным ввиду его большой ресурсоемкости.

На практике, реальные шумы являются суммой высокочастотных гармонических составляющих, и шумовые компоненты, в отличие от случайного шума, практически не наблюдаются в низкочастотной части спектра (в полосе пропускания ФНЧ); их влияние на восстановление производных и последующий анализ значительно меньше расчетных данных, приведенных на рис. 4, 5. Ошибки, возникающие в реальных условиях, в основном, связаны с меньшими амплитудами регистрируемого сигнала.

4. ВЫВОДЫ

По результатам исследования методов восстановления производных, основанных на интерполирующих сплайнах, можно сделать следующие выводы.

Интерполирующие сплайны достаточно чувствительны к наличию в обрабатываемом сигнале шумов. Расчетные оценки показали, что ошибка восстановления производных (10) находится в прямой зависимости от возрастания дисперсии выборки сигнала. Таким образом, для эффективного использования означенных методов необходимо подавление шумовых компонентов сигнала, выражающееся в нашем случае низкочастотной фильтрацией.

Даже при таких условиях преимущества применения сплайнов очевидны. Это, в первую очередь, их низкая ресурсоемкость, позволяющая осуществлять углубленный анализ длинных реализаций сигналов в режиме реального или квазиреального времени. Во-вторых, качество обработки экспериментальных данных можно существенно улучшать, осуществляя мероприятия общего порядка (фильтрация, децимация и т.д.). При выполнении предварительной обработки сплайн—функции дают достаточно хорошие результаты, сопоставимые и зачастую превосходящие другие методы восстановления производных, особенно для высоко—амплитудных участков пульсовой волны.

В то же время, поскольку амплитудные параметры восстанавливаемых производных сильно зависят от характеристик контура, можно отметить недостаточную разрешающую способность этих методов в мало—амплитудных сегментах пульсовой волны, в частности, в фазе диастолы — восстановленные производные не носят выраженный характер, имея вид сильно сглаженных кривых.

Для более тонкого анализа пульсовой волны, вероятно, применима комбинированная схема обработки, когда для реализации в целом применяется один из методов сплайн–аппроксимации, для отдельных участков фазы диастолы, наиболее информативных с диагностической точки зрения, — один из методов, обеспечивающих более высокое разрешение, например, регуляризирующий алгоритм Тихонова.

В целом, рассмотренные нами методы удовлетворяют решению задачи амплитудно–временного анализа пульсовых сигналов. По результатам их анализа скорректированы требования по качеству регистрирующего тракта автоматизированного пульсодиагностического комплекса (АПДК).

Авторы выражают особую благодарность д.ф.-м.н. Игорю Ильичу Орлову (Институт солнечной и земной физики СО РАН, г. Иркутск) за неоценимую поддержку нашей работы.

ЛИТЕРАТУРА

1. Бороноев В. В., Шабанова Е. В. //Измерительная техника, 1994. № 11. С. 60.
2. Вершинин В. В., Завьялов Ю. С., Павлов Н. Н. Экспериментальные свойства сплайнов и задача сглаживания. — Новосибирск: Наука, 1988. – 104 с.
3. Корн Г., Корн Т. Справочник по математике. — М.: Наука, 1973. – 832 с.
4. Логвинов В. С. — В кн.: Пульсовая диагностика тибетской медицины /Под ред. Ч. Ц. Цыдыпова. — Новосибирск: Наука, 1988. С. 90.
5. Сопряжение датчиков и устройств ввода данных с компьютерами IBM PC /Под ред. У. Томпкинса и Дж. Уэбстера. — М.: Мир, 1992. – 592 с.

Бурятский институт естественных наук СО РАН, Россия

Поступила в редакцию 20 февраля 1998 г.

SPLINES IN TIME–AMPLITUDE ANALYSIS OF HUMAN PULSE

V. V. Boronoyev, O. S. Rinchinov

The applicability of cubic and B-splines in the time–amplitude analysis of the human pulse is investigated. It is stated that interpolating splines are rather sensitive to noise. The scheme of preliminary data processing is discussed and realized which is based on the application of specially designed low–frequency filter which allows to raise significantly the spline method efficiency. There were estimated the limits of pulse wave typical point determination error which are dependent on both noise level and input signal amplitude. The results obtained allowed to make corrections of the registering subsystem of the Automated Pulse Diagnostics Complex. The methods investigated were used as the basis of APDC's time–amplitude analysis software.

УДК 621.391

ОПТИМАЛЬНОЕ ОЦЕНИВАНИЕ МОМЕНТА ПОЯВЛЕНИЯ ИМПУЛЬСНОГО СИГНАЛА СО СЛУЧАЙНОЙ СУБСТРУКТУРОЙ

А. П. Трифонов, О. В. Чернояров

Синтезирован алгоритм оптимального оценивания момента появления импульсного сигнала со случайной субструктурой, наблюдаемого на фоне гауссовского белого шума. Выполнен сравнительный анализ байесовского и максимально—правдоподобного измерителей момента появления. Приведены результаты статистического моделирования алгоритмов оценки.

ВВЕДЕНИЕ

Во многих практических приложениях статистической радиофизики необходимо определять момент появления импульсного сигнала. В [1] рассмотрен байесовский алгоритм оценивания момента появления квазидетерминированного импульса известной формы, наблюдаемого на фоне гауссовского белого шума. Методами статистического моделирования установлена его работоспособность и достаточно высокая эффективность. Однако в реальных условиях генерации, распространения и приёма радиоволн структура радиоимпульса может существенно искажаться вследствие влияния помех и случайных воздействий среды распространения [2—4]. Поэтому актуальной представляется задача синтеза и анализа алгоритмов оценивания момента появления импульсного сигнала со случайной субструктурой. В [3] рассмотрена оценка максимального правдоподобия (ОМП) момента появления импульсного сигнала со случайной субструктурой.

Там же найдены асимптотические (при неограниченном увеличении отношения сигнал/шум (ОСШ)) характеристики ОМП с учётом аномальных ошибок. В то же время известно, что применение байесовского подхода [1, 5, 6] в задачах обработки сигналов может обеспечить более высокую эффективность, чем максимально—правдоподобные (МП) алгоритмы. В связи с этим вызывает интерес синтез и анализ байесовского измерителя момента появления случайного импульсного сигнала. Ниже показано, что в рамках байесовского подхода [1, 5, 6] можно получить достаточно простой и эффективный алгоритм оценивания момента появления радиоимпульса со случайной субструктурой.

1. ОЦЕНКА МОМЕНТА ПОЯВЛЕНИЯ ИМПУЛЬСНОГО СИГНАЛА СО СЛУЧАЙНОЙ СУБСТРУКТУРОЙ

Пусть на интервале времени $[0;T]$ наблюдается аддитивная смесь импульсного сигнала со случайной субструктурой $s(t,\lambda_0)$ и помехи $n(t)$

$$x(t) = s(t,\lambda_0) + n(t). \tag{1}$$

Здесь момент появления $\lambda_0 \in [\Lambda_1, \Lambda_2]$ предполагается случайной величиной с априорной плотностью вероятности $W(\lambda)$.

Следуя [2—4], под импульсным сигналом со случайной субструктурой будем понимать отрезок реализации случайного процесса достаточно большой длительности τ

$$s(t,\lambda_0) = \xi(t)\, I[(t-\lambda_0)/\tau]. \tag{2}$$

В (2) обозначено: $I(\cdot)$ — индикатор единичной длительности, а $\xi(t)$ — реализация стационарного центрированного гауссовского случайного процесса. Спектральную плотность процесса $\xi(t)$, описывающего случайную субструктуру импульсного сигнала, представим как [7]

$$G(\omega) = \frac{\gamma}{2}\left[g\left(\frac{\vartheta-\omega}{\Omega}\right) + g\left(\frac{\vartheta+\omega}{\Omega}\right)\right], \quad \vartheta \gg \Omega. \tag{3}$$

Здесь γ — величина, Ω — ширина полосы частот, ϑ — центральная частота спектральной плотности. Функция $g(x)$ описывает форму спектральной плотности и нормирована так, что $\max g(x) = 1$, $\int\limits_{-\infty}^{\infty} g^2(x)\,dx = 1$. Полагаем, что случайный импульсный сигнал (2) полностью находится внутри интервала наблюдения, т. е. $0 \le \Lambda_1 - \tau/2 < \Lambda_2 + \tau/2 \le T$. Помеху $n(t)$ в (1), аналогично [1, 3], аппроксимируем гауссовским белым шумом с односторонней спектральной плотностью N_0. На основе наблюдаемой реализации (1) необходимо оценить случайный момент появления радиоимпульса со случайной субструктурой (2).

Будем считать, что длительность τ импульсного сигнала (2) значительно больше времени корреляции процесса $\xi(t)$, т.е. $\mu = \tau\Omega/2\pi \gg 1$. Тогда в соответствии с [3, 4] логарифм функционала отношения правдоподобия имеет вид

$$L(\lambda) = M(\lambda)/N_0 - (\tau\Omega/2\pi)\int\limits_{-\infty}^{\infty} \ln[1+qg(x)]\,dx,$$

$$M(\lambda) = \int\limits_{\lambda-\tau/2}^{\lambda+\tau/2} y^2(t)\,dt, \tag{4}$$

где

$$q = \gamma/N_0, \tag{5}$$

а $y(t)$ — отклик фильтра с передаточной функцией $H(\omega)$ на реализацию наблюдаемых данных (1). При этом выполняется условие $|H(\omega)|^2 = = f[(\vartheta-\omega)/\Omega] + f[(\vartheta+\omega)/\Omega]$, $f(x) = qg(x)/[1+qg(x)]$.

Согласно [5], ОМП $\hat{\lambda}$ момента появления случайного импульсного сигнала (2) определяется как положение абсолютного (наибольшего) максимума функционала $M(\lambda)$ (4):

$$\hat{\lambda} = \arg\sup M(\lambda), \quad \lambda \in [\Lambda_1, \Lambda_2]. \tag{6}$$

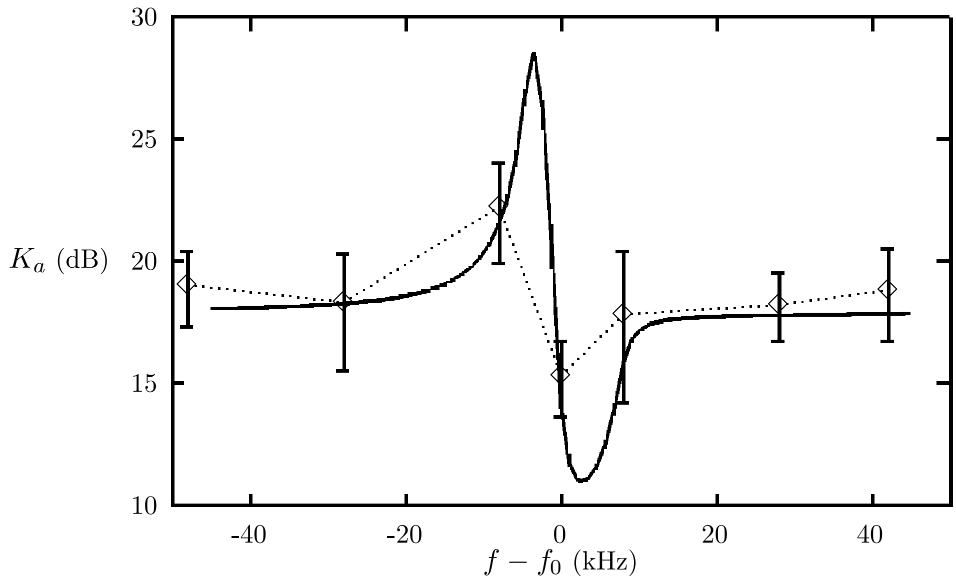
Из (6) следует, что структура ОМП не зависит от априорной плотности вероятности параметра λ_0.

Структурная схема МП измерителя момента появления импульсного сигнала со случайной субструктурой выделена на рис. 1 штриховой линией, где обозначено: 1 — ключ, открывающийся на время $[\Lambda_1 - \tau/2, \Lambda_2 + \tau/2]$; 2 — фильтр с передаточной функцией $H(\omega)$; 3 — квадратор; 4 — интегратор; 5 — линия задержки на время τ; 6 — экстрематор, фиксирующий в качестве оценки $\hat{\lambda}$ положение наибольшего максимума сигнала.

Точность оценки будем характеризовать безусловным рассеянием (средним квадратом ошибки) оценки момента появления

$$V(\hat{\lambda}) = \int\limits_{\Lambda_1}^{\Lambda_2} V(\hat{\lambda}|\lambda)W(\lambda)\,d\lambda,$$

где $V(\hat{\lambda}|\lambda_0)$ — условное рассеяние ОМП момента появления импульсного сигнала (2).



Обозначим: $m = (\Lambda_2 - \Lambda_1)/\tau$ — приведённая длина априорного интервала возможных значений момента появления случайного сигнала [6]. Очевидно, m определяет количество неперекрывающихся сигналов (2), которые могут быть размещены в интервале $[\Lambda_1, \Lambda_2]$. Если $m \gg 1$, то, согласно [5, 6], условное рассеяние ОМП $\hat{\lambda}$ (6) определяется приближённой формулой

$$V(\hat{\lambda}|\lambda_0) = P_0 V_0(\hat{\lambda}|\lambda_0) + (1 - P_0)\left[\frac{\Lambda_2^2 + \Lambda_2\Lambda_1 + \Lambda_1^2}{3} - (\Lambda_2 + \Lambda_1)\lambda_0 + \lambda_0^2\right], \tag{7}$$

точность которой возрастает с увеличением m. Здесь P_0 — вероятность, а $V(\hat{\lambda}|\lambda_0)$ — условное рассеяние надёжной оценки соответственно. Под надёжной понимается оценка $\hat{\lambda}$, для которой $|\hat{\lambda} - \lambda_0| < \tau$ [5, 6].

Иногда вместо вероятности надёжной оценки P_0 удобнее использовать вероятность аномальной ошибки $P_a = 1 - P_0 = P[|\hat{\lambda} - \lambda_0| > \tau]$ [5, 6]. Особенно это целесообразно для задач, в которых необходимо обеспечить достаточно малый уровень нормальных ошибок при фиксированном значении P_a. Величина вероятности аномальной ошибки как характеристики оценки момента появления случайного импульсного сигнала (2) представляет при этом самостоятельный интерес [6].

Воспользовавшись результатами [3], можем записать

$$\begin{aligned} P_0 \approx\ & 2\Psi z \exp\left(\frac{\Psi^2 z^2}{2} + \Psi z^2\right) \int\limits_{\kappa}^{\infty} \exp\left[-\frac{mx}{\kappa\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\kappa^2}\right)\right] \times \\ & \times \Big\{ \exp(-\Psi z x)\Phi[x - z(\Psi + 1)] - \\ & - \exp[3\Psi^2 z^2/2 + \Psi z(z - 2x)]\Phi[x - z(2\Psi + 1)] \Big\}\, dx, \end{aligned} \tag{8}$$

$$\begin{aligned} P_a \approx\ & 2\Psi z \exp\left(\frac{\Psi^2 z^2}{2} + \Psi z^2\right) \int\limits_{\kappa}^{\infty} \left\{ 1 - \exp\left[-\frac{mx}{\kappa\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\kappa^2}\right)\right]\right\} \times \\ & \times \Big\{ \exp(-\Psi z x)\Phi[x - z(\Psi + 1)] - \\ & - \exp[3\Psi^2 z^2/2 + \Psi z(z - 2x)]\Phi[x - z(2\Psi + 1)] \Big\}\, dx, \end{aligned} \tag{9}$$

$$V_0(\hat\lambda|\lambda_0) = \frac{13\tau^2}{2\Psi^2 z^4}, \tag{10}$$

где

$$\Psi = \frac{2}{\kappa^2+1},\ \kappa^2 = \int\limits_{-\infty}^{\infty} \frac{g^2(x)\,dx}{[1+qg(x)]^2},\ z^2 = \mu q^2 \Big\{ \int\limits_{-\infty}^{\infty} \frac{g^2(x)\,dx}{[1+qg(x)]} \Big\}^2 \ — \tag{11}$$

ОСШ для МП алгоритма [3, 4], а $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp(-t^2/2)\,dt$ — интеграл вероятности [7]. Точность формул (7)–(9) возрастает с увеличением μ, z, m, а формулы (10) — с увеличением μ и z.

Усредняя (7) по возможным значениям параметра λ_0 с априорной плотностью вероятности $W(\lambda)$, получаем безусловное рассеяние ОМП $\hat\lambda$:

$$V(\hat\lambda) = P_0 V_0(\hat\lambda) + (1-P_0)[(\Lambda_2^2 + \Lambda_2\Lambda_1 + \Lambda_1^2)/3 - (\Lambda_2+\Lambda_1)\lambda_{\rm pr} + \lambda_{\rm pr}^2 + D_{\rm pr}]. \tag{12}$$

Здесь $V_0(\hat\lambda)$ — безусловное рассеяние надёжной ОМП, совпадающее с условным рассеянием $V_0(\hat\lambda|\lambda_0)$ (10), а $\lambda_{\rm pr}$ и $D_{\rm pr}$ — априорные среднее значение и дисперсия момента появления соответственно.

На рис. 2 сплошными линиями нанесены зависимости нормированного условного рассеяния $\tilde V_{\rm c}(q) = 12V(\hat\lambda|\lambda_0)/(\Lambda_2-\Lambda_1)^2$ ОМП $\hat\lambda$ с учётом аномальных ошибок от величины q (5) при $\lambda_0 = (\Lambda_2+\Lambda_1)/2$. На рис. 3 сплошными линиями изображены аналогичные зависимости нормированного безусловного рассеяния $\tilde V_{\rm u}(q) = 6V(\hat\lambda)/(\Lambda_2-\Lambda_1)^2$, рассчитанные по формулам (8)–(12) для равномерной априорной плотности вероятности момента появления сигнала

$$W(\lambda) = \begin{cases} 1/(\Lambda_2-\Lambda_1), & \Lambda_1 \le \lambda \le \Lambda_2; \\ 0, & \lambda < \Lambda_1,\ \lambda > \Lambda_2. \end{cases} \tag{13}$$

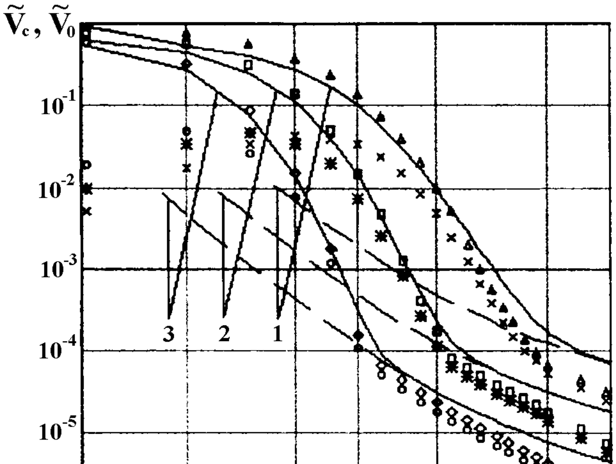


Рис. 2.

Штриховыми линиями на рис. 2, 3 показаны соответствующие зависимости нормированного условного $\tilde V_{0\rm c}(q) = 12V_0(\hat\lambda|\lambda_0)/(\Lambda_2-\Lambda_1)^2$ и безусловного $\tilde V_{0\rm u}(q) = 6V_0(\hat\lambda)/(\Lambda_2-\Lambda_1)^2$ рассеяния (10) надёжной ОМП. Наконец, на рис. 4 приведены теоретические зависимости вероятности аномальной ошибки P_a

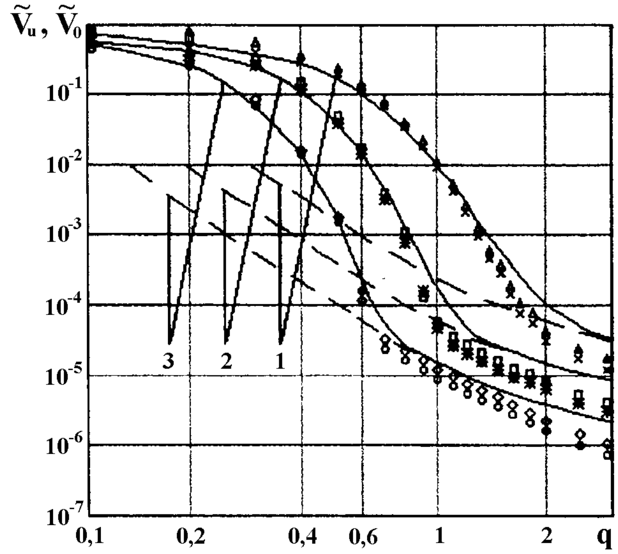


Рис. 3.

от величины q, рассчитанные по формуле (9). При расчёте графиков полагалось, что $g(x) = I(x)$ и $m = 20$. Кривые 1 на рис. 2—4 соответствуют $\mu = 50$, 2 — $\mu = 100$, 3 — $\mu = 200$.

Видно, что при недостаточно больших ОСШ (11) существенную роль начинают играть пороговые эффекты, связанные с относительно частым появлением аномальных ошибок. Вследствие этого условное $V(\hat{\lambda}|\lambda_0)$ (7) и безусловное $V(\hat{\lambda})$ (12) рассеяния ОМП $\hat{\lambda}$ с учётом аномальных ошибок резко возрастают по сравнению с соответствующими условным $V_0(\hat{\lambda}|\lambda_0)$ и безусловным $V_0(\hat{\lambda})$ рассеяниями (10) надёжной ОМП $\hat{\lambda}$. Значит, точность ОМП момента появления случайного импульсного сигнала (2) существенно ухудшается.

Байесовский алгоритм оценивания при квадратичной функции потерь минимизирует средний квадрат ошибки (рассеяние) оценки. Согласно [1, 5, 6], байесовская оценка (БО) момента появления запишется как

$$\lambda_{\text{Б}} = \frac{\int\limits_{\Lambda_1}^{\Lambda_2} \lambda W(\lambda) \exp[L(\lambda)]\, d\lambda}{\int\limits_{\Lambda_1}^{\Lambda_2} W(\lambda) \exp[L(\lambda)]\, d\lambda} = \frac{\int\limits_{\Lambda_1}^{\Lambda_2} \lambda W(\lambda) \exp[M(\lambda)/N_0]\, d\lambda}{\int\limits_{\Lambda_1}^{\Lambda_2} W(\lambda) \exp[M(\lambda)/N_0]\, d\lambda}, \tag{14}$$

где $M(\lambda)$ определяется из (4), а $W(\lambda)$ — априорная плотность вероятности параметра λ_0.

В большинстве случаев байесовские измерители параметров сигналов достаточно сложно реализуемы аппаратурно. Они, как правило, являются многоканальными, причём для полностью оптимальной реализации число каналов должно быть бесконечным [5, 6]. Однако, как следует из [1] и (14), одним из немногих исключений является байесовский измеритель момента появления импульсного сигнала известной формы или со случайной субструктурой. Это существенно облегчает аппаратурную реализацию байесовского алгоритма оценки, структурная схема которого изображена на рис. 1, откуда следует исключить блок 6. Остальные блоки обозначены: 7 — нелинейный элемент с экспоненциальной характеристикой; 8 — генератор линейно–изменяющегося напряжения. Величина сигнала на выходе делителя в момент времени $\Lambda_2 + \tau/2$ является байесовской оценкой $\lambda_{\text{Б}}$ (14). Очевидно, что структура байесовского измерителя момента появления несколько сложнее структуры МП измерителя. Отметим также, что байесовский измеритель момента появления случайного импульсного сигнала может быть достаточно просто получен из байесовского измерителя момента появления квазидетерминированного

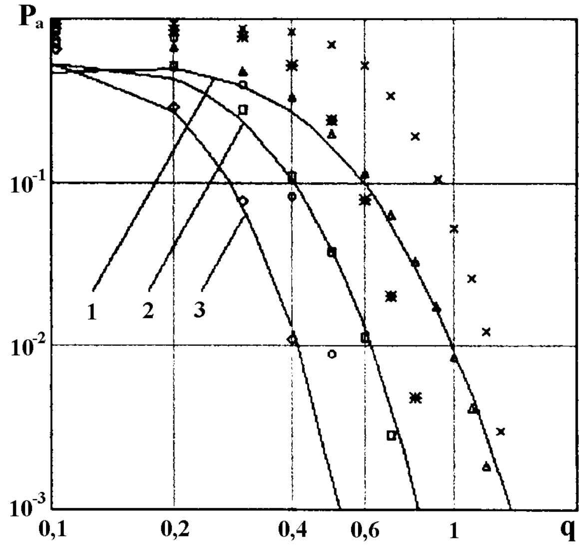


Рис. 4.

импульса известной формы. Для этого в последнем, согласно (14) и [1], необходимо перед согласованным фильтром включить блок, состоящий из фильтра с передаточной функцией $H(\omega)$ и квадратора. Следовательно, наличие случайной субструктуры приводит лишь к незначительному усложнению структуры байесовского алгоритма оценки для квазидетерминированного сигнала.

Теоретический анализ байесовского измерителя выполнить не удаётся. Кроме того, приведённые выше выражения для рассеяния (7), (12) и вероятности аномальной ошибки (9) ОМП $\hat{\lambda}$ являются лишь асимптотически точными с увеличением параметров μ, z, m. При конечных значениях этих параметров определить погрешность найденных формул аналитическими методами пока не представляется возможным. В связи с этим исследование работоспособности байесовского и МП алгоритмов, а также определение границ применимости асимптотических выражений для характеристик ОМП момента появления были выполнены методами статистического моделирования на ЭВМ.

2. РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

При моделировании полагалось, что процесс, описывающий случайную субструктуру импульсного сигнала, имеет прямоугольную спектральную плотность, т. е. $g(x) = I(x)$. Для сокращения затрат машинного времени использовалось представление отклика $y(t)$ узкополосного фильтра с передаточной функцией $H(\omega)$ (4) через его низкочастотные квадратуры [7]. С учётом условия узкополосности (3) процесса $\xi(t)$ это позволило формировать решающую статистику $M(\lambda)$ (4) в виде суммы двух независимых случайных процессов

$$M(\lambda) = [M_1(\lambda) + M_2(\lambda)]/2, \tag{15}$$

$$M_i(\lambda) = \int\limits_{\lambda-\tau/2}^{\lambda+\tau/2} y_i^2(t)\,dt, \quad y_i(t) = \int\limits_{-\infty}^{\infty} x_i(t')h_0(t-t')\,dt',$$

$$x_i(t) = \xi_i(t)I[(t-\lambda_0)/\tau] + n_i(t), \quad i = 1, 2.$$

Здесь $\xi_i(t)$ и $n_i(t)$ — статистически независимые центрированные гауссовские случайные процессы со спектральными плотностями $G_\xi(\omega) = = \gamma I(\omega/\Omega)$ и $G_n(\omega) = N_0$, соответственно. Функция $h_0(t)$ выбирается таким образом, чтобы её спектр удовлетворял условию $|H_0(\omega)|^2 = I(\omega/\Omega)$.

В процессе моделирования на интервале $[\tilde\Lambda_1, \tilde\Lambda_2]$, $\tilde\Lambda_i = \Lambda_i/\tau$, $i = 1, 2$ с шагом Δ формировались отсчёты реализаций случайных процессов $y_i(t)$ (15). Это позволило получить ступенчатую аппроксимацию нормированной решающей статистики $M(\lambda)/N_0$ (4) в виде

$$\tilde M(l) = \frac{1}{2} \sum_{k=k_{\min}}^{k_{\max}} (y_{1k}^2 + y_{2k}^2)\Delta. \tag{16}$$

Здесь $k_{\min} = \mathrm{int}\,\{(l - 1/2)/\Delta\}$, $k_{\max} = \mathrm{int}\,\{(l + 1/2)/\Delta\}$; $l = \lambda/\tau$ — нормированное текущее значение момента появления; $\mathrm{int}\{\cdot\}$ — целая часть числа. При $\Delta = 0{,}05/\mu$ относительная среднеквадратическая погрешность ступенчатой аппроксимации (16) непрерывной реализации функционала (15) не превышала 10%. Отсчёты процессов y_{ik}, $i = 1, 2$, формировались на основе последовательности независимых гауссовских чисел методом скользящего суммирования [4, 8]:

$$y_{ik} = \frac{1}{\sqrt{\Delta}} \sum_{s=k-p}^{k+p} \alpha_{is} H_{ks} + \sum_{s=S_1}^{S_2} \xi_{is} H_{ks},$$
$$\xi_{is} = \frac{1}{\pi}\sqrt{\frac{q}{\Delta}} \sum_{n=0}^{2p} H_{np}\beta_{i\,n+s+1}. \tag{17}$$

Здесь $S_1 = \max(S_{\min}, k{-}p)$; $S_2 = \min(S_{\max}, k{+}p)$; $S_{\min} = \mathrm{int}\{(l_0{-}1/2)/\Delta\}$; $S_{\max} = \mathrm{int}\{(l_0{+}1/2)/\Delta\}$; $l_0 = \lambda_0/\tau$; α_{is}, β_{is} — независимые гауссовские случайные числа с нулевыми математическими ожиданиями и единичными дисперсиями; $H_{ks} = \sin[2\pi\mu\Delta(k - s)]/[\pi(k - s)]$.

Число слагаемых в суммах (17) выбиралось так, чтобы относительное отклонение дисперсии сформированного отсчёта от дисперсии моделируемого процесса не превышало 5%. Это соответствует значению $p = 50$. Формирование независимых гауссовских чисел с параметрами $(0, 1)$ осуществлялось на основе стандартного датчика равномерно распределённых в интервале $[0, 1]$ независимых случайных чисел методом Корниша–Фишера [9].

По полученной на основе формул (16), (17) реализации процесса $M(l)$, согласно (6) и (14), определялись нормированные оценки $\hat l = \hat\lambda/\tau$ и $l_{\text{Б}} = \lambda_{\text{Б}}/\tau$, соответственно:

$$\hat l = \arg\sup \tilde M(l), \quad l \in [\tilde\Lambda_1, \tilde\Lambda_2],$$

$$l_{\text{Б}} = \frac{\sum\limits_{i=0}^{\tilde m} (\tilde\Lambda_1 + i\Delta) W(\tilde\Lambda_1 + i\Delta) \exp[\tilde M(\tilde\Lambda_1 + i\Delta)]}{\sum\limits_{i=0}^{\tilde m} W(\tilde\Lambda_1 + i\Delta) \exp[\tilde M(\tilde\Lambda_1 + i\Delta)]},$$

где $\tilde m = \mathrm{int}\{m/\Delta\}$.

Экспериментальные характеристики нормированных ОМП $\hat l$ и БО $l_{\text{Б}}$ находились следующим образом. Пусть при некотором фиксированном значении l_0 сформировано N оценок $\hat l_1, \hat l_2, \ldots, \hat l_N$ и $l_{\text{Б}1}, l_{\text{Б}2}, \ldots, l_{\text{Б}N}$ для N различных реализаций $M(l)$. Тогда условные рассеяния ОМП и БО определятся соответственно как

$$V(\hat l|l_0) = \frac{1}{N}\sum_{i=1}^{N} (\hat l_i - l_0)^2, \quad V(l_{\text{Б}}|l_0) = \frac{1}{N}\sum_{i=1}^{N} (l_{\text{Б}i} - l_0)^2. \tag{18}$$

При нахождении безусловных характеристик оценок в каждой из N реализаций наблюдаемых данных (1) измеряемый параметр l_0 являлся значением случайной величины, описываемой плотностью вероятности $W(l)$, соответствующей $W(\lambda)$. По сформированным наборам оценок $\hat{l}_i$, $l_{\text{Б}i}$, $i = \overline{1,N}$, аналогично (18) вычислялись безусловные рассеяния ОМП и БО

$$V(\hat{l}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{l}_i - l_{0i})^2, \quad V(l_{\text{Б}}) = \frac{1}{N}\sum_{i=1}^{N}(l_{\text{Б}i} - l_{0i})^2.$$

Здесь l_{0i} — значения случайной величины с плотностью вероятности $W(l)$. Кроме того, находились оценки безусловных вероятностей аномальных ошибок

$$P_{\text{a}} = P[|\hat{l} - l_0| > 1] \approx n_1/N, \quad P_{\text{a}} = P[|l_{\text{Б}} - l_0| > 1] \approx n_2/N.$$

Здесь n_1, n_2 — число зафиксированных аномальных ошибок для ОМП и БО, соответственно.

Некоторые результаты статистического моделирования при $W(l) = = 1/(\tilde{\Lambda}_2 - \tilde{\Lambda}_1)$ (13), $\tilde{\Lambda}_1 = 0$, $\tilde{\Lambda}_2 = m = 20$ представлены на рис. 2–4. На рис. 2 для случая $\lambda_0 = (\Lambda_1 + \Lambda_2)/2$ нанесены экспериментальные значения нормированного условного рассеяния $\tilde{V}_{\text{c}}$ ОМП $\hat{\lambda}$ с учётом аномальных ошибок при $\mu = 50$ (треугольники), $\mu = 100$ (прямоугольники) и $\mu = 200$ (ромбики), а также значения нормированного условного рассеяния БО $\lambda_{\text{Б}}$ момента появления при $\mu = 50$ (крестики), $\mu = 100$ (звёздочки), $\mu = 200$ (кружочки). На рис. 3 треугольниками (при $\mu = 50$), прямоугольниками (при $\mu = 100$) и ромбиками (при $\mu = 200$) показаны экспериментальные значения нормированного безусловного рассеяния $\tilde{V}_{\text{u}}$ ОМП $\hat{\lambda}$, а крестиками (при $\mu = 50$), звёздочками (при $\mu = 100$) и кружочками (при $\mu = 200$) — экспериментальные значения нормированного безусловного рассеяния БО $\lambda_{\text{Б}}$. При этом случайный момент появления λ_0 импульсного сигнала (2) имел априорную плотность вероятности (13). Наконец, на рис. 4 приведены экспериментальные значения безусловной вероятности аномальной ошибки P_{a} ОМП и P_{a} БО. Обозначения рис. 4 такие же, как на рис. 2, 3. Каждое экспериментальное значение на рис. 2–4 получено в результате обработки не менее 10^4 реализаций решающей статистики $\tilde{M}(l)$. При этом с вероятностью 0,9 границы доверительных интервалов отклоняются от экспериментальных значений не более чем на $10 \div 15\%$.

Как следует из рис. 2, 3, теоретические зависимости условного (7) и безусловного (12) рассеяний ОМП $\hat{\lambda}$ с учётом аномальных ошибок удовлетворительно аппроксимируют экспериментальные данные по крайней мере при $\mu \geq 50$, $z \geq 0{,}5$ и $m \geq 20$. С уменьшением q, когда ОСШ $z \leq 5 \div 7$, вероятность P_a аномальных ошибок значительно возрастает и приближается к 1. Это приводит к скачкообразному (по сравнению со случаем надёжной оценки) увеличению рассеяния оценки момента появления. С ростом q, когда $z > 5\ldots7$, рассеяние $\tilde{V}_{c,u}$ сходится к рассеянию $\tilde{V}_0$ и оценка становится надёжной с вероятностью, близкой к 1. При $q > 2\ldots3$ теоретические зависимости $V(\hat{\lambda}|\lambda_0)$ (7), $V(\hat{\lambda})$ (12) отклоняются от экспериментальных значений. Это объясняется тем, что формула (10) для рассеяния $V_0(\hat{\lambda}|\lambda_0)$ надёжной оценки момента появления получена в работе [3] в пренебрежении ошибками оценивания порядка времени корреляции случайной субструктуры импульсного сигнала (2). Следовательно, когда относительное рассеяние ОМП убывает до величины порядка μ^{-2}, погрешность формулы (10) становится значительной.

Сопоставление результатов моделирования байесовского и МП измерителей показывает, что условное рассеяние БО всегда меньше условного рассеяния ОМП и стремится к нулю не только при $q \to \infty$, но и при $q \to 0$. Последнее связано с тем, что истинное значение момента появления случайного импульсного сигнала (2) совпадает в данном случае с его априорным средним λ_{pr}, а БО сходится к величине λ_{pr} с уменьшением q в среднеквадратическом [5].

Безусловное рассеяние БО в пороговой области (при $1 \leq z \leq 5$) практически совпадает с соответствующим рассеянием ОМП $\hat{\lambda}$. Однако при $z \geq 5\ldots7$ рассеяние БО становится меньше рассеяния

ОМП, причём $V(\lambda_Б)/V(\hat{\lambda}) \approx 0{,}8$. Если же ОСШ z весьма мало ($z \leq 1$), то $V(\lambda_Б)/V(\hat{\lambda}) \approx \approx 0{,}5$. Таким образом, точность ОМП $\hat{\lambda}$ в общем случае несколько уступает точности БО $\lambda_Б$. В то же время использование МП измерителя позволяет заметно снизить вероятность аномальной ошибки при $z \geq 1$.

Приведённые результаты получены при поддержке Российского фонда фундаментальных исследований.

ЗАКЛЮЧЕНИЕ

Для определения момента появления импульсного сигнала со случайной субструктурой можно использовать как байесовский, так и МП алгоритмы оценки. Аппаратурная реализация МП измерителя является более простой по сравнению с байесовским. Кроме того, ОМП инвариантна по отношению к априорной плотности вероятности оцениваемого параметра, которая в общем случае может быть неизвестна наблюдателю. В то же время при наличии полной априорной информации байесовский алгоритм обеспечивает предельно достижимую точность оценки.

Таким образом, полученные результаты позволяют в зависимости от объёма имеющейся априорной информации и требований, предъявляемых к простоте аппаратурной реализации и точности получаемой оценки, сделать обоснованный выбор между байесовским и МП измерителями момента появления импульсного сигнала со случайной субструктурой.

ЛИТЕРАТУРА

1. Ванжа А. В., Силаев А. М. //Изв. ВУЗов. Радиофизика, 1995. Т. 38. № 12. С. 1257.
2. Ахманов С. А., Дьяков Ю. Е., Чиркин А. С. Введение в статистическую радиофизику и оптику. — М.: Наука, 1981.
3. Трифонов А. П., Захаров А. В. //Изв. ВУЗов. Радиоэлектроника, 1986. Т. 29. № 4. С. 36.
4. Трифонов А. П., Нечаев Е. П., Парфенов В. И. Обнаружение стохастических сигналов с неизвестными параметрами. — Воронеж: Изд-во ВГУ, 1991.
5. Куликов Е. И., Трифонов А. П. Оценка параметров сигналов на фоне помех. — М.: Сов. радио, 1978.
6. Трифонов А. П., Шинаков Ю. С. Совместное различение сигналов и оценка их параметров на фоне помех. — М.: Радио и связь, 1986.
7. Тихонов В. И. Статистическая радиотехника. — М.: Радио и связь, 1982.
8. Быков В. В. Цифровое моделирование в статистической радиотехнике. — М.: Сов. радио, 1971.
9. Ермаков С. М., Михайлов Г. А. Статистическое моделирование. — М.: Наука, 1982.

Государственный университет,
г. Воронеж, Россия

Поступила в редакцию
30 декабря 1997 г.

OPTIMAL ESTIMATION OF TIME OCCURRENCE OF PULSE SIGNAL WITH RANDOM SUBSTRUCTURE

A. P. Trifonov, O. V. Chernoyarov

The algorithm of the optimal estimation of time occurrence of the pulse signal with random substructure observed on the background of white Gaussian noise has been synthesized. The comparative analysis of Basyes and maximum likelihood meters of the time occurrence has been made. The results of the statistical computer modeling of estimation algorithms have been presented.

КРАТКИЕ СООБЩЕНИЯ И ПИСЬМА В РЕДАКЦИЮ

УДК 551.466.81

НОРМАЛЬНЫЕ КАНОНИЧЕСКИЕ ПЕРЕМЕННЫЕ ДЛЯ БАРОТРОПНЫХ ВОЛН РОССБИ

В. В. Петров

В работе [1] был предложен вариант решения задачи гамильтонова описания баротропных волн Россби в приближении β-плоскости. При этом в процедуре отыскания преобразования к нормальным каноническим переменным была допущена ошибка. Ее устранение и составляет цель настоящего сообщения.

Как и в [1], будем описывать β-волны Россби на языке переменных Клебша [2], представляя азимутальную (u) и меридиональную (ν) проекции вектора скорости жидкости в виде

$$u = \frac{\partial\varphi}{\partial x} + \alpha\frac{\partial\beta}{\partial x} - \sqrt{2\Omega}\alpha\,,$$
$$\nu = \frac{\partial\varphi}{\partial y} + \alpha\frac{\partial\beta}{\partial y} - \sqrt{2\Omega}\beta\sin\lambda\,. \tag{1}$$

Здесь

$$\sin\lambda = \sin\lambda_0 + \frac{\beta_0 y}{2\Omega}\,,$$

$\beta_0 = 2\Omega\cos\lambda_0/a$, Ω —угловая скорость вращения Земли, a —радиус Земли, ось x направлена на восток, ось y — на север. Уравнения динамики жидкости в переменных Клебша (1) приводятся к виду [2]

$$\dot\alpha = -\left(u\frac{\partial}{\partial x} + \nu\frac{\partial}{\partial y}\right)\alpha - \sqrt{2\Omega}\nu\sin\lambda\,,$$
$$\dot\beta = -\left(u\frac{\partial}{\partial x} + \nu\frac{\partial}{\partial y}\right)\beta + \sqrt{2\Omega}u\,, \tag{2}$$

$$\frac{du}{dx} + \frac{d\nu}{dy} = 0. \tag{3}$$

Нетрудно показать, что уравнения (2) имеют гамильтонову структуру и порождаются гамильтонианом

$$H = \int h\,dx\,dy\,, \tag{4}$$

где

$$h = \frac{1}{2}\left(u^2 + \nu^2\right).$$

При этом

$$\dot\alpha = \frac{\partial_ф H}{\partial_ф\beta}\,,\quad \dot\beta = -\frac{\partial_ф H}{\partial_ф\alpha}\,. \tag{5}$$

Здесь
$$\frac{\partial_\text{ф} H}{\partial_\text{ф}\beta}=\frac{\partial h}{\partial\beta}-\frac{\partial}{\partial x}\frac{\partial h}{\partial(\partial\beta/\partial x)}-\frac{\partial}{\partial y}\frac{\partial h}{\partial(\partial\beta/\partial y)}\,,\quad \frac{\partial_\text{ф} H}{\partial_\text{ф}\alpha}=\frac{\partial h}{\partial\alpha}\ \text{—}$$
функциональные производные. Из (2)–(5) следует, что переменные α и β имеют смысл канонически сопряженных координаты и импульса, соответственно, а потенциал φ может трактоваться как некоторый функционал от α и β, определяемый с помощью уравнения (3).

При отыскании преобразования к нормальным каноническим переменным, диагонализирующим квадратичную часть H_0 гамильтониана (4), учтем анизотропию волн Россби, проявляющуюся в том, что они могут распространяться лишь с востока на запад. С этой целью представим преобразование Фурье по координате x для переменных α и β в следующем виде:
$$\alpha=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\left(\alpha_{k_1}\mathrm{e}^{ik_1x}+\alpha^*_{k_1}\mathrm{e}^{-ik_1x}\right)dk_1\,,$$
$$\beta=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\left(\beta_{k_1}\mathrm{e}^{ik_1x}+\beta^*_{k_1}\mathrm{e}^{-ik_1x}\right)dk_1\,, \tag{6}$$
где $\alpha_{k_1}=\beta_{k_1}=0$ при $k_1>0$. Можно показать, что такое преобразование является каноническим и в нем роль новых канонически сопряженных координаты и импульса играют спектральные плотности α_{k_1} и $\beta^*_{k_1}$, соответственно. Переходя в H_0 с помощью (6) и с учетом (3) к новым переменным α_{k_1} и $\beta^*_{k_1}$ и полагая
$$\beta^*_{k_1}=\frac{i}{k_1}\frac{\partial}{\partial y}\left(\frac{\alpha^*_{k_1}}{\sin\lambda}\right),$$
что отвечает каноническому преобразованию
$$\alpha_{k_1}\to\alpha_{k_1},\quad \beta^*_{k_1}\to\alpha^*_{k_1},$$
найдем
$$\tilde{H}_0=-i\beta_0\int\limits_{-\infty}^{\infty}dk_1\,k_1\int dy\,\alpha^*_{k_1}\sin\lambda\int dy'\,G_{k_1}(y,y')\left(\frac{\alpha_{k_1}}{\sin\lambda}\right)\Bigg|_{y=y'}. \tag{7}$$
Здесь $G_{k_1}(y,y')$ — функция Грина, являющаяся решением уравнения
$$\frac{\partial^2 G_{k_1}}{\partial y^2}-k_1^2G_{k_1}=\delta(y-y').$$

Для перехода в (7) к нормальным переменным достаточно произвести еще одно каноническое преобразование вида
$$\alpha_{k_1}(y,t)=\sin\lambda(y)\sqrt{\frac{-2k_1\Omega}{\beta_0}}\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}a_{\vec{k}}(t)\mathrm{e}^{ik_2y}\,dk_2$$
$$(\vec{k}=(k_1,k_2)).$$
При этом квадратичная часть (4) приводится к стандартной форме
$$H_0=\int\limits_{-\infty}^{0}dk_1\int\limits_{-\infty}^{\infty}dk_2\omega_{\vec{k}}a_{\vec{k}}a^*_{\vec{k}},$$

где $\omega_{\vec{k}} = -k_1\beta_0/(k_1^2 + k_2^2)$ — закон дисперсии волн Россби. Суммируя последовательность выше проведенных канонических преобразований, получим для искомого преобразования от исходных гамильтоновых переменных α и β к нормальным каноническим переменным $a_{\vec{k}}$ и $ia^*_{\vec{k}}$ следующие выражения

$$\begin{aligned}
\alpha(x,y,t) &= \frac{\sin\lambda(y)}{2\pi}\sqrt{\frac{2\Omega}{\beta_0}}\cdot \\
&\cdot\int\limits_{-\infty}^{0} dk_1 \int\limits_{-\infty}^{\infty} dk_2 \sqrt{-k_1}\left(a_{\vec{k}}(t)\mathrm{e}^{i(k_1x+k_2y)} + a^*_{\vec{k}}(t)\mathrm{e}^{-i(k_1x+k_2y)}\right), \\
\beta(x,y,t) &= \frac{1}{2\pi}\sqrt{\frac{2\Omega}{\beta_0}}\cdot \\
&\cdot\int\limits_{-\infty}^{0} dk_1 \int\limits_{-\infty}^{\infty} dk_2 \frac{k_2}{\sqrt{-k_1}}\left(a_{\vec{k}}(t)\mathrm{e}^{i(k_1x+k_2y)} + a^*_{\vec{k}}(t)\mathrm{e}^{-i(k_1x+k_2y)}\right).
\end{aligned} \tag{8}$$

Для проверки каноничности (8) достаточно убедиться в инвариантности функциональной скобки Пуассона $[\alpha,\beta]$ относительно этого преобразования. Несложные вычисления показывают, что

$$[\alpha(x_1,y_1,t),\beta(x_2,y_2,t)]_{a_{\vec{k}},a^*_{\vec{k}}} = i\delta(x_1-x_2)\delta(y_1-y_2),$$

т.е., действительно, (8) описывает искомый закон преобразования к нормальным каноническим переменным.

ЛИТЕРАТУРА

1. Куркин А. А., Петров В. В. //Изв. ВУЗов. Радиофизика, 1996. Т. 39. № 9. С. 1067.
2. Селиджер Р. Л., Уитем Г. Б. //Сб. перев. Механика, 1969. № 5. С. 99.

Нижегородский государственный
университет, Россия

Поступила в редакцию
21 мая 1998 г.

NORMAL CANONICAL VARIABLES FOR ROSSBY BAROTROPIC WAVES

V. V. Petrov